



**International  
Journal of Society, Culture & Language  
IJSCL**

Journal homepage: [www.ijsc.net](http://www.ijsc.net)  
ISSN 2323-2210 (online)

## **Developing an Online Kazakh-English-Russian Thesaurus of Industry-Specific Terminology**

**Ainur T. Bayekeyeva<sup>1a</sup>, Saule Zh. Tazhibayeva<sup>2a</sup>, Aigul A. Shaheen<sup>3a</sup>, Zhainagul S. Beisenova<sup>4a</sup>, Gulnar Beysenkyzy Mamayeva<sup>5a</sup>**

### **Abstract**

Industry-specific translation is one of the rapidly developing and highly demanded sectors in Kazakhstan. This paper discusses the theoretical and methodological issues of compiling a controlled multilingual vocabulary or thesaurus for the purposes of industry-specific translation. The proposed multilingual thesaurus addresses the gap of existing online dictionaries with the lack of ability to translate directly from English into Kazakh, and provides relevant equivalents for mining and metal terms in three languages: Kazakh, English, and Russian. The study is interdisciplinary and combines methods of linguistic analysis, machine learning, ontological and corpus-based studies that allow an exhaustive analysis, generalization, comparison, systematization, classification, and formalization of the mining industry terminology through a multilingual parallel corpus based on the Zthez international technical platform. The controlled multilingual thesaurus is designed to promote intercultural communication of professionals and translators in the mining industry. Besides it enables the foundation to be built for a Kazakh language corpus, unifying and systematizing Kazakh industry-specific terms in a multilingual way.

© 2022 IJSCL. All rights reserved.

### **ARTICLE HISTORY:**

Received December 2021  
Received in Revised form February 2022  
Accepted February 2022  
Available online March 2022

### **KEYWORDS:**

Controlled vocabulary  
Terminology  
Interoperable thesaurus  
Industry-specific translation  
Multilingual thesaurus

<sup>1</sup> PhD Candidate, Email: [a.baekeeva@rambler.ru](mailto:a.baekeeva@rambler.ru) (Corresponding Author)

Tel: +77-477825011

<sup>2</sup> Professor, Email: [sauletazhibayeva@mail.ru](mailto:sauletazhibayeva@mail.ru)

<sup>3</sup> Associate Professor, Email: [akingan@mail.ru](mailto:akingan@mail.ru)

<sup>4</sup> Associate Professor, Email: [zhaina\\_b@mail.ru](mailto:zhaina_b@mail.ru)

<sup>5</sup> Associate Professor, Email: [gulnar.mamaeva@mail.ru](mailto:gulnar.mamaeva@mail.ru)

<sup>a</sup> L.N.Gumilyov Eurasian National University, Republic of Kazakhstan  
<http://dx.doi.org/10.22034/ijsc.2022.543082.2465>

## 1. Introduction

There has been an increased interest in issues related to intercultural communication within specific professions (Bloschchynskiy et al., 2021). Industry-specific translation (IST) serves as a cross-subject science between theory and practice in translation for industry-specific terminology, especially due to the impact of the ongoing global pandemic on the world's economy and other related industries, including translation services. State bodies and ministries, educational establishments, and schools and universities have turned home to offices, so the need for online resources has significantly increased. At the same time, IST is one of the most difficult and laborious translation services that requires translators' high qualifications and knowledge of a certain subject area. A technical translator needs to be extremely aware of high language acquisition and academic writing skills and have a good sense of both the Source Language (SL) and Target Language (TL); the core requirements rely especially on gaining depth in the knowledge and skills of technical thinking and terminology.

The innovative technologies applied in translation studies implementing the objectives of the State Programme "Digital Kazakhstan" are forcing Kazakhstan's translators, linguists, Turkologists, and industry-specific terminologists to make a new quantum leap. In the digitalization period, it is important to create a new generation of dictionaries that meet modern topical challenges. One of these dictionaries is a controlled multilingual vocabulary or thesaurus on subject areas jointly developed by linguists and experts in technology information corresponding to the subject fields. Industry-specific terms in the Kazakh language are urgently needed to fix, regulate, and identify semantic equivalents in Russian and English. Thesauri in various subject areas are necessary for quick information retrieval with the accurate semantic translation of industry-specific terms not only for translators but also for all interested experts.

The mining industry has received a new development impetus since Kazakhstan gained independence. Many metallurgical plants and complexes started applying modern advanced technologies for exploration, geology, production, and processing. The terminological

base of mining and metal terms has been enriched with new terms related to technological processes, while the original Turkic Kazakh terms have become less used. At the same time, the mining and metallurgical industry, as well as the professions associated with them, occupies a special place in the Atlas of New Professions and Competencies of Kazakhstan. The Terminological Committee of Kazakhstan carries out activities to arrange new terms in all sectors and industries, publishing hardcopy Kazakh-Russian terminological dictionaries. However, industry-specific terms are not always correctly interpreted by linguists and translated into Kazakh or Russian. Moreover, the translation of terms is often not accurate, with a variety of synonyms that make it difficult for the translator to understand and choose only the one semantically correct equivalent.

Therefore, one of the important objectives of this research is to systematize new international terms and document language material related to the Turkic lexical layer in a modern interoperable platform. In this regard, the study of Kazakh mining and metal terms is of particular relevance. Mining terminology was first adopted and developed by Satpayev in 1957. However, since then, the terminological base has not been regularly and systematically updated. Modern Kazakh mining and metal terms have been enriched by international terms that do not reflect their semantic content. Satpayev's approach to compiling Kazakh terms is artificial and leads to a functional limitation of their use in practice. Now, we have to be ready for modern challenges and create terminological dictionaries based on the semantic concepts in various knowledge fields, one of them being the mining industry.

Translating technical texts for industries is now one of the most popular types of translation services in Kazakhstan. This type of translation has a number of features. First of all, it requires not only the knowledge of industry-specific terms applied in a certain subject area, e.g., the mining and metallurgical industry, but also of related areas such as geology, enrichment, beneficiation, refinement, and so forth. Besides, the technical translator must focus on mining and metal terms in a general way and offer clear, exact details in exploration, development, enrichment and beneficiation, equipment and

technology, and other issues of the mining industry.

The role of the Kazakh language in science and technology texts remains in the shade since Russian is the dominant language for IST. In addition, modern controlled interoperable thesauri are usually monolingual, so the lack of English-Kazakh and Kazakh-English dictionaries in subject areas is a serious problem for professionals and translators. Thus, when searching in online dictionaries, the absence of trilingual options, i.e., English-Russian-Kazakh creates difficulties, not only for professionals and translators but for all Automated Information System (henceforth AIS) users in related industries. Many trilingual industry-specific dictionaries are available only in hard copy, and in the current pandemic situation, all AIS users should be given an equal opportunity to access these dictionaries regardless of their location. Therefore, there is an urgent need to create a systematized controlled multilingual and interoperable thesaurus to promote communication and interaction between professionals, translators, and AIS users in industry-specific areas and to offer them a way to freely use special terminology, provide adequate translation in specific subject areas, and find the necessary data related to a particular industry-specific term without having to search several dictionaries for the relevant information and accurate terminology in order to avoid misunderstanding and mistranslation.

Rapidly growing economic cooperation in Kazakhstan makes it necessary to enrich the terminological base with various new terms denoting technology and equipment. Direct English-Kazakh translation occurs very rarely. For this reason, creating a new generation of dictionaries will enhance opportunities for Kazakhstan's translation services and strengthen the status of this language as a state one.

The purpose of this investigation is to explore the terminology of the mining industry and compile a controlled Kazakh-English-Russian multilingual thesaurus related to this subject area. The analysis of issues related to thesauri is a new interdisciplinary study at the intersection of translation studies, computational linguistics, Turkology, terminology, and information

technology. This interdisciplinary and cross-subject research has the following objectives:

- to study the ways of forming Kazakh mining and metal terms;
- to identify and document the common Turkic layer of vocabulary in Kazakh mining terminology;
- to systematize new international terms that are actively enriching the terminological base of Kazakh industry-specific terms; and
- to identify semantic equivalents in translating special terms into English and Russian.

This paper considers more than 10,000 Kazakh mining and metal terms provided by the Terminological Committee of Kazakhstan. The proposed multilingual thesaurus is shown to provide relevant equivalents of three languages in one data scheme: Kazakh, Russian, and English. The paper demonstrates the use of the controlled multilingual thesaurus in the Digital Library (DL) that enables the user to search for a faithful translation of industry-specific terms. This also lays the groundwork for the future development of the Kazakh-language corpus and for the unification and systematization of industry-specific Kazakh terms in a multilingual way.

## 2. Theoretical Framework

### 2.1. Industry-Specific Translation

Translation of scientific and technical literature differs greatly in translation methods from those of literary works. A literary translator may be perceived as a portrait painter who aspires to portray rather than show a photographic similarity (Bayekeyeva et al., 2020). Scientific and technical translators' quality work depends mainly on the vocabulary (terminology), which tends to be univocal and have ready-made equivalents, and on the use of language style to transfer image accuracy (Franco, 2004). It can be said that fiction translation is associated with the field of art, while technical translation refers to the field of science.

IST starts from developing a special terminological base, which is why a thesaurus in the DL is compiled to solve the industry term systematization. Bayekeyeva et al. (2020) define the DL as "a system for managing structured cataloged collections that provides a

comprehensive search of dissimilar digital resources” (p. 1870). In this sense, the major purpose of compiling a thesaurus is directly related to information support in scientific research and translation activities; it also has to do with the formation and development of its own IST electronic resources through the interoperable management of publications and bibliographies. Bayekeyeva et al. (2020) discuss the theoretical and methodological issues of compiling a multilingual thesaurus for conducting IST in the mining industry. They describe the facts which translators deal with for providing adequate translation when facing industry-specific terms in real texts.

## 2.2. Diachronic Analysis of Traditional Dictionaries of Industry-specific Terms

Practical, empirical preconditions for the systematization of industry terms began in the 1920s, and this is evidenced by the systematization and publication of various industry terms in dictionaries directly related to the development of scientific and technological progress in Europe; for instance, see Fay (1920).

Terminology, which involves the study and compilation of terms belonging to specific fields, has attracted the attention of many Kazakh scholars such as Kaidarov (1993), Baitursynuly (2009), and Aitbayuly (2013). The theoretical and practical aspects of forming industry-specific terms have been investigated by domestic terminologists, among them Abdirasilov (1999), Isakova (2000), Kaliuly (2008), Beisenova (2011, 2014), and Kurmanbayuly (2014), as well as foreign linguists Lotte (1961), Reformatsky (1961), Cabré (1998), Leichik (2009), and Superanskaya (2012). All these researchers in terminology noted the need to arrange industry-specific terms and made attempts to compile traditional dictionaries.

The development of terminology is an independent scientific field that studies the basis of branch translation in linguistics terms and dates back to the 1930s (Cabre, 1998). Due to the rapid development of industrialization, industry-specific terms began to form in the Kazakh and Russian languages. Before the emergence of multilingual dictionaries of Kazakh industry-specific terms, the first dictionaries were compiled in short versions or glossaries by Omarov (1923) in *Physics subject words*, and Karatyshkanov (1927) in *Subject*

*terms*; they were published in Orenburg, Moscow, and Kyzylorda. In this sense, bilingual dictionaries of Russian/Kazakh industry-specific terms began publishing *Russian-Kazakh military words* (1926), *Dictionary of names* (1931), *Terminological dictionary* (1931), and *Terms of the Kazakh language* (1935).

More than 150 terminological dictionaries were published in Kazakhstan from 1940 to 1990. Between 1991 and 2003, more than 100 dictionaries were compiled and published, such as *Explanatory dictionary of the Kazakh language* published 1974–1986, *Phraseological dictionary of the Kazakh language* (1977), *Russian-Kazakh dictionary of terminology* (Ámirbaev et al., 1959), *Dictionary of the Kazakh language* (1999), and the *Dictionary of Turkic languages* (Oztopcu et al., 1996).

In the second half of the 20th century, from 1959 to 1995, other Russian-Kazakh terminological dictionaries were developed relating to various subject fields different from the mining industry, such as geology (Mashanov et al., 1960), mechanism and machinery (Joldasbekov & Ábdirahmanov, 1968; Serikbaev et al., 1974), chemical terms (Birimjanov & Omarov, 1974), agricultural mechanization (Myńbaev et al., 1983), water engineering terms (Ábdirahmanov & Manabaev, 1991), hydrogeology and engineering geology (Sydyqov & Qabiev, 1993), and theoretical mechanics (Joldasbekov & Saǵıtov, 1994).

Furthermore, during this time, there also were Russian-Kazakh dictionaries for the mining industry, specifically on metallurgy (Ábdirahmanov & Býketov, 1959) and physics (Bakaev et al., 1962; Mashanov et al., 1959). In the 2000s, a 31-volume series of 155,000 words of Russian-Kazakh and Kazakh-Russian industry-specific terminological dictionaries were published and approved by the Terminological Committee (Kurmanbaiuly & Sapina, 2004). There were also dictionaries and thesauri related to architecture and construction (Qulmanov, 2009), information systems (Ilyasova, 2010), ship construction (Koyanbayev & Sabirova, 2013), polygraphic terms (Medeubekuly et al., 2013), legal terms (Omashuly, 2013), post terms and phrases (Toksanbay, 2013), veterinary terms (Djanabekov & Djanabekova, 2013), oil and gas terminology (Makhambet, 2015), and mass media and telecommunication terms (Tursyn,

2015). However, in spite of the recent development in the mining industry, no new dictionaries of mining and metal terms have been published for almost two decades.

### 2.3. Kazakh-language Corpus and Thesaurus Compilation

In recent years, the Kazakh language has become an object of major concern to linguists. Several scholars have attempted to make up a corpus of the Kazakh language that would be of great significance in natural language processing tasks, including information retrieval and machine translation. One of the first attempts to compile a Kazakh corpus was made in 2013; it contained more than 135 million words belonging to five different stylistic genres (Makhambetov et al., 2013). Rakhimova and Zhumanov (2017) created a multilingual parallel corpus of general words in the Kazakh, Russian, and English languages using the Bitextor application and demonstrated how dictionaries could be enriched with new words without special linguistic knowledge.

Numerous researchers have contributed to the informatization of natural language processing of the Kazakh language. Thus, Altenbek and Sun (2010) applied N-gram model methods and rule-based methods to extract noun phrases from Kazakh-language texts. Bekbulatov and Kartbayev (2014) focused on the internal morphological structure of Kazakh sentences and statistically demonstrated the impact of morphological preprocessing on the quality of Kazakh-English machine translation. Tukeyev and Karibayeva (2020) constructed a system of Kazakh word endings for effective morphological segmentation. Automated processing tools and algorithms were developed for the KazNLP platform to deal with texts in the Kazakh language (Yessenbayev et al., 2020).

Despite the rapid development of translation theory, mining and metal terminology have not been given due attention. The uniqueness of the Kazakh language in rubrics of online industry-specific terminology is manifested in the fact that the original Kazakh terms of the mining industry constitute a rich lexical layer for providing the names of metals and minerals. New, borrowed international terms normally indicate the names of equipment and technologies. In English and Russian, instead of industry-specific terms, as a rule, abbreviated forms of nomenclature of chemical elements

are frequently used, e.g., Au, Ag, Cu, Fe, Zn, etc. When translating scientific and technical documents, such as industry-specific texts in Kazakh, Russian, and English, written documentation with certain additional features is needed. This study suggests using a controlled vocabulary or thesaurus for translation purposes.

A *thesaurus* is a comprehensive online dictionary of industry-specific terms, compiled in an interoperable system with links, which includes several types of dictionaries such as a glossary, bilingual dictionary, multilingual dictionary, synonymic dictionary, and terminological dictionary. It is one of the major tools in translation theory and practice, particularly in compiling various dictionaries and glossaries for specific fields and purposes, which require the participation of a wide range of professionals in linguistics, information technology, related branches of science and technology, and so forth. The compilation of monolingual thesauri of special terminology has been considered by a number of researchers in various scientific areas. Aleshinskaya and Albatsha (2020) focused on methodological issues and suggested a cognitive model of compiling a monolingual terminological thesaurus in computer science. Heiradi et al. (2020) compiled a corpus of academic words belonging to the pharmacy domain. Niyazbek and Talp (2019) contributed to the construction of a Kazakh terminology recognition platform in information technologies through a conditional random field and manual modification. Aitim et al. (2020) developed a complex methodology for generating a completely automatic thesaurus of Kazakh-language specialized vocabulary that would increase the effectiveness of natural language processing.

A *controlled dictionary* is an interoperable dictionary that meets professional requirements for information retrieval and machine translation. The issues of creating a controlled thesaurus of the Kazakh language have been studied by domestic and foreign scholars in various aspects. Kazakh-English parallel models and constructions have been studied and adopted for machine translation. For example, Kartbayev (2015) attempted to improve random Kazakh-English parallel corpora in terms of phrase tables by reducing the word alignment error rate. In methods of extending machine translation resources, Rakhimova and Zhumanov (2017) demonstrate how transfer

rules can be automatically learned from a parallel corpus. Two parallel Kazakh-English corpora KAZNU and WMT19 were explored in terms of solving the problem of unknown words for neural machine translation (Turganbayeva and Tukeyev, 2020). Akhmed-Zaki et al. (2021) developed several preprocessing tools to generate word forms and handle morphological ambiguity.

The problems of creating bilingual and multilingual thesauri have been investigated by a number of researchers such as Ahmad and Gillam (2005), Mansurova et al. (2017), and Niyazbek and Talp (2019). They focused on designing and developing a media-corpus for the Kazakh language and for recognition systems. Moreover, grammatical, syntactical, and lexicological issues have been studied from the perspectives of the parsing model of Kazakh sentence structure (Sharipbay et al., 2019; Yerimbetova, 2016) and Kazakh morphology (Yelibayeva et al., 2019). Creating a controlled multilingual thesaurus, including Turkic languages, has been considered for ontology in information technology (Fedotov et al., 2016; Sambetbayeva et al., 2016). Thesauri in modern applied sciences such as WordNet, EuroWordNet, or Wiktionary linguistic resources describe the relationship between lexical meanings of natural language as a hierarchical system of synonymic groups, i.e., synsets (Tussupov et al., 2016).

The review of all this scientific literature inherent to the research topic of this paper reveals that a comprehensive study of mining and metal terms in multisystem languages has not been previously conducted, and as the evidence suggests, it is extremely necessary. So far, multilingual online thesauri researches of mining industry-specific terms have been developed neither in Kazakhstan nor abroad. In this respect, it is of the utmost significance to compile a controlled multilingual Kazakh-Russian-English thesaurus for the mining industry that would be of practical value in translation.

### 3. Methodology

#### 3.1. Materials

To ensure industry-specific terms in mining and metal from Kazakh, Russian, and English, it was decided to use the universal Integrated Distributed Information System (IRIS) to

compile the thesaurus developed by the Institute of Computer Technology and Science of the Siberian branch of the Russian Academy of Sciences in 1998 (Fedotov et al., 2016).

#### 3.2. Procedure

This section describes all the steps used to identify, systematize, and compare mining terms for the controlled thesaurus.

The first step is the material selection drawn from a sample of Kazakh terms belonging to the mining industry. More than 10,000 units were selected from bilingual and multilingual Kazakh industry-specific dictionaries (Termincom.kz). The equivalent translation term is based on data from multilingual industry-specific dictionaries (Qudaibergenov, 2009) and English-speaking countries (Fay, 1920; Hornby, 1995; Encyclopedia Britannica, 1998), as well as online resources such as Sozdik (<https://sozdik.kz/>) and Termincom (<http://termincom.kz/terms>).

The second step is parsing the website and compiling and saving information easily (Khairova et al., 2019) in the analysis of digital sources for terminological and industry-specific dictionaries. The most actively used sites for Kazakh translation are Multitran (<https://www.multitran.com/>), Sozdik (<https://sozdik.kz/>), and Termincom (<http://termincom.kz/terms>). This study shows the challenges that Kazakh translators face when translating mining industry terms to Russian and English.

The third step is related to the analysis of the collected data using the Zthes data scheme. This is a family of specifications that enables the representation of thesauri and facilitates interoperability for applications dealing with thesauri (<http://zthes.z3950.org/>). This tool also enables the creation of a thesaurus for the mining industry that unifies and systematizes industry-specific Kazakh terms in conformity with the requirements of modern interoperable thesauri. Thus, the interoperability is shown as an easily integrated way to communicate in the other language that really helps describe and provide Russian and English equivalents, as well as classify the terms into relevant topics using related rubrics and associated relations in a hierarchical order.

#### 3.2. Compiling an Industry-Specific Thesaurus

The research is mainly based on linguistic analysis methods that enable the creation of an

exhaustive analysis, generalization, comparison, systematization, classification, and formalization of mining industry terms. This method was used to (a) systematize the mining and metal terms of the Kazakh language, (b) identify features in forming Kazakh mining terminology, and (c) perform a comparative analysis with similar terms in English and Russian. Formalizing and locating are implicit in the translation process when adapting the borrowed terms to Kazakh. To create a controlled thesaurus, supportive methods of machine translation were used, such as automatic rubricating, object, and relation, creating ontologies of the subject area. To guarantee reliable results, other supportive methods such as statistics and documentation were applied. In this way, there is an integration of methods that combine interdisciplinary and cross-subject research relations.

Furthermore, to analyze and systematize mining industry terms, a multilingual parallel corpus for Kazakh-Russian-English is used. The terms are taken from Kazakh and translated into Russian and English, respectively. A parallel corpus is particularly important in studying language, and translation features since it provides precision and accuracy in the equivalent analyzed lexicon through machine translation (Khairova et al., 2019).

Some previous research has approached the ontological and computational study of bilingual and multilingual thesauri for the mining industry and other related fields, focusing on particular linguistic aspects. Ontological studies establish a clear and decisive relationship between concepts and categories of the studied subject area. For the mining industry field, ontology is a set of concepts and statements about their classification, relations between them, and the hierarchy on the “general-part” and “part-whole” relations (Bayekeyeva et al., 2020). During this research, it was extremely necessary to consider the conceptual methodology aspects in a systemic way and manifest them as a unity of “whole-part”, “external-internal”, “static-dynamic”, “cause-effect”, and “communication process”, among others.

Computational studies also offer a great contribution to the development of the present

research, especially with the use of its tools that provide proven accuracy in the results and also facilitate the organization and development of the research process. Accordingly, Kuandykova et al. (2014) constructed an English-Kazakh parallel corpus at the School of Mechanics and Mathematics and created a statistical machine translation system to present problems and solutions to legal texts. In the same way, Myrzakhmetov and Makazhanov (2016) present an initial experiment on Russian to Kazakh phrase-based statistical machine translation (SMT) through a parallel corpus based on morphological processing techniques. They state that “machine translation from Russian to Kazakh poses certain challenges of both linguistic and technical nature” (p. 153).

These researches help the present paper as a baseline to support the study of a controlled vocabulary in DL: terminology, translation, and thesaurus for Kazakh-Russian-English languages in the mining industry field.

## 4. Results

### 4.1. Digital Sources for Terminological and Industry-Specific Dictionaries

Due to the rapid development of computer technology, professionals in the field of IST have used a variety of electronic dictionaries, including ABBYY Lingvo, DICT, FreeDict, Free On-line Dictionary of Computing, WordNet, Multilex, etc. Among the electronic dictionaries most actively used by Kazakh translators are Multitran (<https://www.multitran.com/>), Sozdik (<https://sozdik.kz/>), and Termincom (<http://termincom.kz/termins>). However, when searching for the necessary terms, Kazakh translators come across multifaceted issues connected with using trilingual contexts.

A serious problem is the lack of direct Kazakh-English online dictionaries and a Kazakh-language corpus for industry-specific contexts. Existing electronic databases such as Sozdik, Multitran, and Termincom do not always meet the requirements for a quick search of information. Translators particularly use Kazakh-Russian dictionaries for Kazakh-Russian bilingual texts (see figures 1 and 2).

The screenshot displays the Sozdik online dictionary interface. At the top, the logo 'sozdik' is visible, along with navigation links for 'Сөздік' (Dictionary), 'Аудармашы' (Translator), and 'Конвертер' (Converter). Below the navigation bar, there are icons for 'Сөздік', 'Тарих' (History), 'Таңдамалы' (Favorites), 'Сөз қосу' (Add word), and 'Нұсқама' (Help).

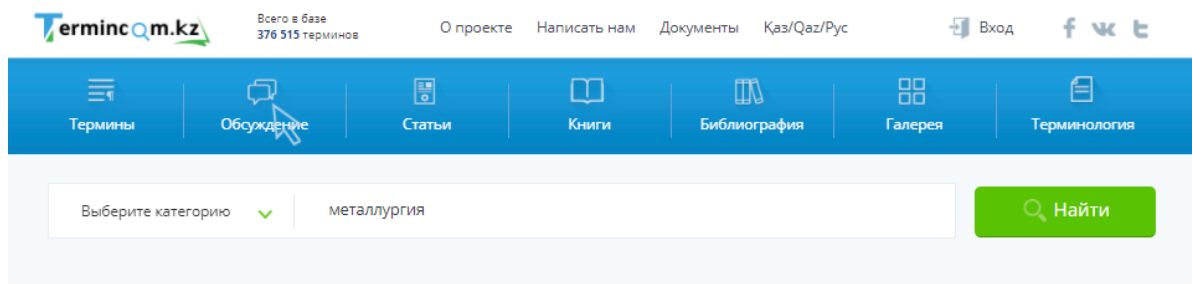
The main heading is 'Қазақша-орысша сөздік: металлургия' (Kazakh-Russian dictionary: metallurgy). Below this is a virtual keyboard with Kazakh characters. The search interface includes a 'Бағыт' (Direction) dropdown set to 'Қазақша → Орысша' (Kazakh → Russian), a search button 'Іздеу', and a search scope dropdown 'тек сөздік мақалаларынан және мысалдар...'. Below the search bar, it indicates 'Қазақша-орысша сөздік 56865 мақала, 424149 мысал.' (Kazakh-Russian dictionary 56865 articles, 424149 examples).

The search term 'Сөз/тіркес \* металлургия' is entered in the search box. A blue 'Аудару' (Translate) button is present. Below the search box, examples are shown: 'Мысалы: [хош](#), [488653597](#) немесе [с\\*\\*д\\*к](#)'. At the bottom right, there are icons for a heart, a document, and a printer. The word 'металлургия' is displayed in bold, with a link to 'металлургия' below it.

**Figure 1**

*Web Page of Kazakh-Russian Sozdik Online Dictionary*





**РЕЗУЛЬТАТЫ ПОИСКА**

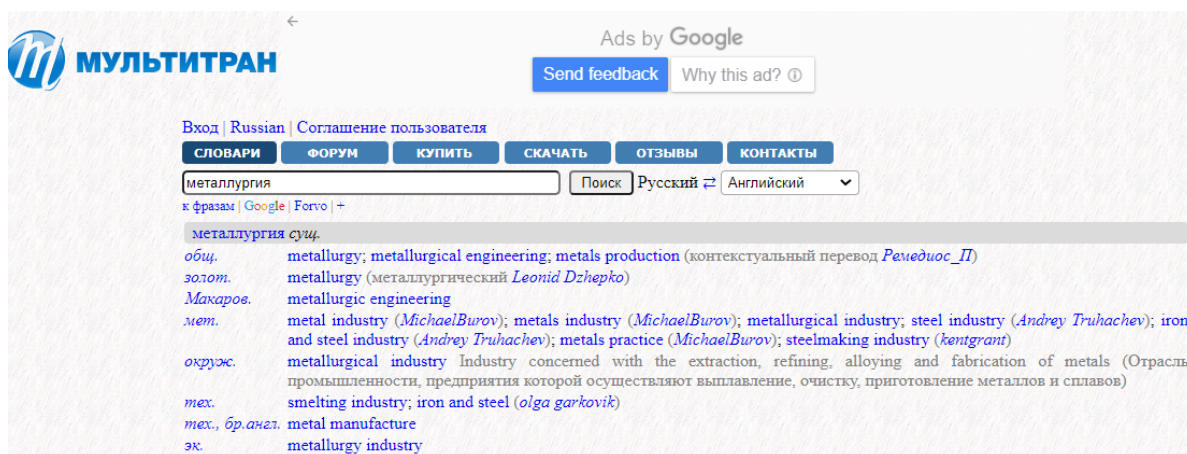
№	Название на казахском	Название на русском	Категория
1	тау-кен-металлургиялық кешен (индустрия)	горно-металлургический комплекс	2015 жылы бекітілген термин
2	Қорытпа (металлургия)	Сплав	1971-1981 жж. бекітілген термин
3	Жасыту (металлургия)	Отжиг	1971-1981 жж. бекітілген термин
4	Қорыту цехының шебері (металлургия)	Мастер плавильного цеха	1971-1981 жж. бекітілген термин
5	тау-кен-металлургиялық кешен	горно-металлургический комплекс	Термины 2012-2015
6	Электрметаллургия	Электрометаллургия	Физика и астрономия
7	Металлургиялық эффект	Эффект металлургический	Энергетика
8	Металлургия	Металлургия	Энергетика
9	Металлургиялық сынықтар	Лом металлургический	Машиностроение
10	Металлургия	Металлургия	Машиностроение
11	Пирометаллургия	Пирометаллургия	Машиностроение
12	Металлургия	Металлургия	Транспорт и средства связи
13	Металлургиялық сынықтар	Лом металлургический	Транспорт и средства связи
14	Металлургиялық химия	Металлургическая химия	Химия
15	Алтынды бөліп алудың гидрометаллургиялық тәсілі	Гидрометаллургический способ выделения золота	Химия

**Figure 2**

Web Page of Kazakh-Russian Terminological Dictionary by Terminological Committee

Translators have to use Russian-English dictionaries for Russian-English bilingual texts (see Figure 3). So, in order to translate the texts

from Kazakh into English, translators utilize two different dictionaries because of the absence of trilingual dictionaries.



**Figure 3**

Web Page of Russian-English Online Dictionary

In order to face the Kazakh translators' reality and to give an opportune answer to the problems mentioned above, this study presents a clever and convenient solution that consists of creating a new generation of multilingual Kazakh-Russian-English dictionaries. In this regard, it is necessary to immediately study, systematize, and give a scientific description of industry-specific terms in the Kazakh language.

The Zthez international technical platform, tested by the Institute of Computational Technologies SB RAS, has been chosen for the compilation of a thesaurus of industry-specific terms in DL (<http://db4.sbras.ru/elbib/data/admin/adm.phtml>). The members of the research group represent in Kazakhstan the traditions of the linguistic science school and the IT technology and science school of the Siberian branch of the Russian Academy of Science (see Figure 4).



**Figure 4**

*Web Page of the Thesaurus of Industry-specific Terms in Digital Library*

Developing a thesaurus contributes to the creation of an online database of industry-specific terms of the Kazakh language, with the most accurate semantic equivalents, synonyms, associate ranges, and other empirical material for the further enrichment of the Kazakh language corpus with texts, metadata, and translations of materials into English and Russian. The main advantage of this multilingual thesaurus is in its interoperable characteristics, so this thesaurus can be easily integrated into international databases such as the UNESCO, NISO, and LOGOS thesauri, compiled in accordance with the international standards (see National Information Standards Organization, 2003, 2005; IFLA, 2009).

#### **4.2. Controlled Multilingual Thesaurus of Mining Terms**

Industry-specific terms should provide a clear and accurate reference to real objects and phenomena, establishing an unequivocal

understanding of the transmitted information by professionals. Therefore, a special requirement is imposed on these types of words. While considering the terms, it is necessary to bear in mind that, first of all, the term has the exact meaning in each field of study, i.e., it has a strictly defined meaning, which can be revealed by a logical definition, establishing the place of the designated concept by the term in the conceptual systems of a certain industry, science, or technology.

Gradually, the content of technical knowledge of the source language (SL) begins to penetrate the signs of the target language (TL) and enrich its vocabulary. In any language, the word or phrase is already inseparable from its meaning, and here the content of technical knowledge becomes an element of the language in a certain field, in this case, the mining industry. The technical knowledge which has found its expression in a word, in a term of SL, passes into a qualitatively new stage, joining the

semantic structure of TL, becoming a component of the terminological system of the given language.

Based on the foregoing, it can be concluded that industry-specific terms evolved over a long time and have now formed a strict terminological system. Within this system, certain forms of term building have been created, the degree of productivity of which makes it possible to determine the dynamics of growth of terms and further means of their development. In connection with the progress of industry, science, and technology, where new concepts are constantly being opened, it is necessary to use more and more new terms in the English, Kazakh, and Russian languages. Term building takes place both at the expense of the resources of the English language itself and by borrowing from other languages. This process allows language development and makes the terminology international, which allows scientists and professionals to understand each other better.

The ongoing research focuses on compiling an interoperable, controlled multilingual thesaurus for translation and terminology, and makes a contribution to the Kazakh language corpus of industry-specific terms. As a result of this study, major rubrics have been created for about 10,000 Kazakh terms related to the mining and metal industry. The terms in the proposed thesaurus are listed as a result of the analysis of more than 5,000 documents developed in Kazakhstan's mining projects (Tau Ken Samruk, its subsidiaries Tau-Ken Altyn, Tau-Ken Temir, Shalkiya Zinc, Silicon Mining, Masalsky GOK, Altyntau Kokshetau, Alaigy, Spassk, etc.). All the terms of "Mining and Metallurgy" are grouped into the following rubrics: "Mining and Metals", "Minerals", "Exploration", "Production", "Equipment and Technology".

The proposed multilingual thesaurus provides relevant equivalents for three languages in one data scheme: Kazakh, Russian, and English. As a major translator's tool, the controlled multilingual thesaurus is a collection of information (corpus, arch) encompassing with maximum fullness the concept, definitions, and terms of a special field of knowledge or industry, with examples in real contexts, i.e., their use in texts. It is a particular type of special vocabulary in which semantic relations

(synonyms, antonyms, paronyms, hyponyms, hyperonyms, and so forth) between lexical units are indicated.

The controlled multilingual thesaurus in the DL designed for IST comprises at least three elements:

- a list of structured and systematized terms, placed on the DL platform;
- the relations stated between the terms, indicated by their hierarchical relative position (e.g., *metal* – broader term; *iron concentrate* – narrower term, *iron content and Fe-content* – synonym, etc.);
- a set of rules on how to use the thesaurus in the DL.

Thus, in order to form the interoperable Information Retrieval Thesaurus for IST, it is strictly recommended to provide the following basic criteria of compiling a controlled vocabulary:

- a set of thesaurus descriptors should be sufficient to describe an arbitrary domain document, particularly for the subject area of industries or study fields; and
- the number of descriptors should not be too large; they should contain the necessary information for the subject area only.

Most professional guides, dictionaries, and glossaries designed for IST are normally systematized in alphabetical order and usually have multiple synonyms for industry-specific terms. In contrast, the controlled multilingual thesaurus makes it possible to reveal the meaning of a special term with the help of descriptions and other concept group correlations and to relate broader and narrower terms.

## 5. Concluding Remarks

This information technology era offers great opportunities to develop, enhance, and solve communication issues between frontiers. This technology is applied in various fields of human knowledge, particularly in those related to linguistic communication, and technical translation in the industry area is one of them. Machine translation systems used by translators of any nationality and culture expand, open, and feed the communication process for all social sectors. In this regard, this paper considers the translation issues faced by Kazakh translators when dealing with the special terminology of

mining industry fields to be translated into Russian and English. According to the study results, it is concluded that these issues are mainly related to the lack of a multilingual thesaurus that allow a suitable translation – simultaneously – in the three languages, keeping the same semantic structure, meaning, and function, to maintain a clear and truthful translation process.

To make life easier in the Kazakh translation business, it is, therefore, necessary to develop an industry-based thesaurus or multilingual thesaurus for a specific field of study. The mining industry is a rapidly developing industry whose terminology needs careful systematization to allow quick translation from English into Kazakh and vice versa. The creation of a controlled-vocabulary Kazakh-English-Russian thesaurus answers the main question of this research: Kazakh translation issues not having a multilingual tool to provide accurately translated equivalents in Russian and English.

The thesaurus based on the Zthes data scheme is a uniquely created, interoperable thesaurus that provides equivalents for industry-specific terms in Kazakh, English, and Russian, respectively. Unlike other multilingual dictionaries, the controlled multilingual thesaurus enables the revelation of the meaning of a term with the help of descriptions, correlations with other concepts and their groups, and related broader and narrower terms.

The novelty of this study is based on the prompt and proper solution offered for translation issues that not only helps the Kazakh scholarly community but also others. Results have shown the benefits of creating and using a multilingual thesaurus to give adequate answers to specific problems when applying certain search criteria. These benefits are to (a) establish semantic relations in lexical units to deal properly with the vocabulary in the required languages, (b) structure and systematize the terms to relate them in the three languages in an interoperable relation, and (c) hierarchize the term positions.

Therefore, the study of industry-specific terms and multilingual thesauri is a promising area of modern linguistics that unites translation, terminology, and computational linguistics in an interdisciplinary approach. Therefore, this study may help to improve, develop, and create further research in multilingual parallel

corpora, computational linguistics, and morphological and ontological studies as well.

The controlled thesaurus of industry-specific terms described in the present study can serve as a model for compiling similar multilingual dictionaries for other sectors and industries in Kazakhstan and other countries. Furthermore, it can be used to enrich the knowledge base for intelligence systems, as well as build the foundation for a Kazakh language corpus, unifying and systematizing Kazakh industry-specific terms in a multilingual way. Due to its interoperability, the presented multilingual thesaurus can easily be integrated into international databases such as UNESCO, NISO, and LOGOS.

### Acknowledgments

This work was written within the framework of the joint projects for a PhD thesis and a multilingual digital thesaurus at L. N. Gumilyov Eurasian National University and Novosibirsk State University (the Governmental grant for 2017–20) with the support of Goethe University Frankfurt.

### References

- Ábdirahmanov, A., & Býketov, E. (1959). *Metallýrgua ónerkásibi termunderiniń qysqasha oryssh-qazaqsha sózdigi* [Short Russian-Kazakh dictionary of terms of metallurgical industry]. Almaty: Rayan.
- Ábdirahmanov, Á., & Manabaev, B. (1991). *Sý tehnikasy termunderiniń oryssh-qazaqsha sózdigi* [Russian-Kazakh dictionary of water engineering terms]. Almaty: Nauka.
- Abdirasilov, E. (1999). *Lingvistikalyq termunderdiń semantikalyq qatynastary* [Semantic relations of linguistic terms]. Almaty: Nauka.
- Ahmad, K., & Gillam, L. (2005). Automatic ontology extraction from unstructured texts. In R. Meersman, & Z. Tari (Eds.), *On the move to meaningful Internet systems 2005: CoopIS, DOA, and ODBASE* (pp. 1330-1346). Berlin, Heidelberg: Springer.
- Aitbaiuly, O. (2013). *Qazaq til biliminiń terminologuasy máseleleri* [The terminological issues of Kazakh linguistics]. Almaty: Abzal-Ai.

- Aitim, A. K., Satybaldiyeva, R. Z., & Wojcik, W. (2020). The construction of the Kazakh language thesauri in automatic word processing system. In R. Uskenbayeva (Ed.), *Proceedings of ICEMIS'20: 6th ACM International Conference on Engineering & MIS 2020* (pp. 1-4). New York: ACM Press.
- Akhmed-Zaki, D., Mansurova, M., Madiyeva, G., Kadyrbek, N., & Kyrgyzbayeva, M. (2021). Development of the information system for the Kazakh language preprocessing. *Cogent Engineering*, 8(1), 1896418.
- Aleshinskaya, E., & Albatsha, A. (2020). A cognitive model to enhance professional competence in computer science. *Procedia – Computer Science*, 169, 326–329.
- Altenbek, G., & Sun, R. (2010). Kazakh noun phrase extraction based on N-gram and rules. In M. D. G. Zhou, H. Qi, & M. Zhamg (Eds.), *Proceedings of IALP 2010: 2010 International Conference on Asian Language Processing* (pp. 305–308). Los Alamitos, CA: IEEE Computer Society.
- Ámirbaev, M., Bektaev, Q., Bókeihanov, R., Jáýtikov, O., Orazbaev, B., & Sátbaev, M. (1959). *Oryssha-qazaqsha terminologua sózdigi* [Russian-Kazakh dictionary of terminology]. Almaty: Publishing House of the Academy of Sciences of the Kazakh SSR.
- National Information Standards Organization (2003). *Information retrieval (Z39.50): Application service definition and protocol specification* (ANSI/NISO Z39.50-2003). Bethesda, MD: NISO Press.
- National Information Standards Organization (2005). *Guidelines for the construction, format and management of monolingual controlled vocabularies* (ANSI/NISO Z39.19-2005). Bethesda, MD: NISO Press.
- Baitursynuly, A. (2009). *Til – qural*. [Language as a tool]. Almaty: Sardar.
- Bakaev, M., Nuǵymanov, Q., Seidyaliyev, Z., Ybyraev, Sh., & Ulyqbekov, O. (1962). *Oryssha-qazaqsha terminologua sózdigi. Taý-ken isteri* [Russian-Kazakh dictionary of terminology. Mining]. Almaty: Publishing House of the Academy of Sciences of the Kazakh SSR.
- Bayekeyeva, A., Tazhibayeva, S., Shaheen, A., Beisenova, Z., & Mamayeva, G. (2020). Multilingual thesaurus of industry-specific terms as major aids for translators. *Opción*, 36(27), 1864-1900.
- Beisenova, Z. (2011). *Otraslevaya terminologiya: Sistemnost, tipologiya, funktsionirovaniye* [Industry-specific terminology: System, typology, functions]. Astana: CBO and MI.
- Beisenova, Z. (2014). *Intercultural approach of professional lexicon in veterinary medicine*. London: Aitmatov Academy.
- Bekbulatov, E., & Kartbayev, A. (2014). A study of certain morphological structures of Kazakh and their impact on the machine translation quality. In A. Adamov (Ed.), *Proceedings of Application of Information and Communication Technologies: 8th IEEE International Conference* (pp. 1-5). Los Alamitos, CA: IEEE Computer Society.
- Birimjanov, B., & Omarov, S. (1974). *Himua terminderiniń oryssha-qazaqsha sózdigi* [Russian-Kazakh dictionary of chemical terms]. Almaty: KazGosNII.
- Bloschchynskiy, I., Borakovskyy, L., Prihodko, G., Novikova, T., Moroz, N., & Kalyniuk, N. (2021). The comparative analysis of the English and German term-formation in the legislative documents (based on the Schengen Border Code). *International Journal of Society, Culture & Language*, 9(3), 73-81.
- Cabrè, M. (1998). *Terminology: Theory, methods and applications*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Djanabekov, K., & Djanabekova, G. (2013). *Oryssha-qazaqsha veterinarlyq sózdik* [Russian-Kazakh veterinary dictionary]. Almaty: Sózdik Slovar.
- Encyclopedia Britannica* (15<sup>th</sup> ed.). (1998). Chicago: Encyclopedia Britannica.
- Fay, A. (1920). *A glossary of the mining and mineral industry*. Washington: Government Printing Office.
- Fedotov, A., Tussupov, J., Sambetbayeva, M., Fedotova, O., Sagnayeva, S., Bapanov, A., & Tazhibayeva, S. (2016). Classification model and morphological analysis in multilingual scientific and educational information systems.

- Journal of Theoretical and Applied Information Technology*, 86(1), 96–111.
- Franco, J. (2004). The study of technical and scientific translation: An examination of its historical development. *The Journal of Specialized Translation*, 1, 29-49.
- Heidari, F., Jalilifar, A., & Salimi, A. (2020). Developing a corpus-based word list in pharmacy research articles: A focus on academic culture. *International Journal of Society, Culture & Language*, 8(1), 1-15.
- Hornby, A. (1995). *Oxford advanced learner's dictionary*. Oxford: Oxford University Press.
- IFLA. (2009). *Guidelines for multilingual thesauri* (IFLA professional reports: 115). Retrieved from <http://www.ifap.ru/library/book411.pdf>
- Ilyasova, G. (2010). *Oryssha-qazaqsha terminologualyq anyqtamalyq sózdik (Aqparattyq júe salasy)* [Russian-Kazakh terminological dictionary. Information systems]. Almaty: Memlekettik Tildi Damytý Institutýty.
- Isakova, S. (2000). *Qazaq tilindegi psihologua termunderi* [Terms of psychology in the Kazakh language]. Almaty: Sózdik Slovar.
- Joldasbekov, Ó., & Ábdirahmanov A., (1968). *Oryssha-qazaqsha mehanizm men mashinalardyń terminologualyq sózdigi* [Russian-Kazakh terminological dictionary of mechanisms and machines]. Almaty: Publishing House of the Academy of Sciences of the Kazakh SSR.
- Joldasbekov, Ó., & Saǵıtov M. (1994). *Teorualyq mehanika termunderiniń oryssha-qazaqsha sózdigi* [Russian-Kazakh dictionary of terms of theoretical mechanics]. Almaty: Nauka.
- Kaidarov, A. (1993). *Qazaq terminologuasyna jańasha kózqaras* [A new view on Kazakh terminology]. Almaty: Nauka.
- Kaliuly, B. (2008). *Qazaq termintanymynyń ózekti máseleleri* [Topical issues of Kazakh terminology]. Almaty: Sózdik Slovar.
- Kartbayev, A. (2015). *Learning word alignment models for Kazakh-English machine translation*. In V.-N. Huynh, M. Inuiguchi, & T. Demoeux (Eds.), *Proceedings of Integrated Uncertainty in Knowledge Modelling and Decision Making: 4th International Symposium, IUKM 2015* (pp. 326-335). Cham: Springer.
- Karatyshkanov, N. (1927). *Pán sózderi* [Subject terms]. Almaty: Er-Daulet.
- Khairova, N., Kolesnyk, A., Mamyrbayev, O., & Mukhsina, K. (2019). The aligned Kazakh-Russian parallel corpus focused on the crime theme. In V. Lytvyn, N. Sharonova, T. Hamon, O. Cherednichenko, N. Grabar, A. Kowalska-Styczen, & V. Vysotska (Eds.), *Proceedings of Computational Linguistics and Intelligent Systems: 3rd International Conference, COLINS-2019* (pp. 116-125). Aachen: CEUR Workshop Proceedings.
- Koyanbayev, A., & Sabirova, N. (2013). *Keme qurylysynyń sózdigi. Oryssha-aǵylshynsha-qazaqsha* [Dictionary of ship construction. Russian-English-Kazakh]. Almaty: Sózdik Slovar.
- Kuandykova, A., Kartbayev A., & Kaldybekov, T. (2014). English-Kazakh parallel corpus for statistical machine translation. *International Journal on Natural Language Computing*, 3(3), 65-72.
- Kurmanbaiuly, S. (2014). *Qazaq terminologuasý: Zertteýler, oqýlyq, sózdik, bibliografua* [Kazakh terminology: Research, manual, dictionary, bibliography]. Almaty: Sardar.
- Kurmanbaiuly, S., & Sapina, S. (2004). *Oryssha-qazaqsha, Qazaqsha-Oryssha termunder men ataýlar sózdigi. Memterminkom bekitken termunder men ataýlar sózdigi* [Russian-Kazakh, Kazakh-Russian dictionary of terms. Dictionary of terms approved by Terminological Committee]. Almaty: Sózdik Slovar.
- Leichik, V. (2009). *Terminovedenie: Predmet, metody, struktura* [Terminology: Subject, methods, structure]. Moscow: Librokom.
- Lotte, J. (1961). *Osnovy ostroeniya nauchno-tehnicheskoy terminologii* [Fundamentals of building scientific and technical terminology]. Moscow: AS USSR.
- Makhambet, A. (2015). *Munai men gaz olynsha Qazaqsha-oryssha aǵylshynsha sózdik* [Kazakh-Russian-English dictionary for oil and gas]. Almaty: Asyl Sóz.
- Makhambetov, O., Makazhanov, A., Yessenbayev, Z., Matkarimov, B., Sabyrgaliyev, I., & Sharafudinov, A. (2013). Assembling the Kazakh

- language corpus. In D. Yarowsky, T. Balwin, A. Korhonen, K. Livescu, & S. Bethard (Eds.), *Proceedings of EMNLP 2013: 2013 Conference on Empirical Methods in Natural Language Processing*, (pp. 1022–1031). Seattle, WA: Association for Computational Linguistics.
- Mansurova, M., Madiyeva, G., Aubakirov, S., Yermekov, Z., & Alimzhanov, Y. (2017). Design and development of media-corpus of the Kazakh language. In N. Nguyen, G. Papadopoulos, P. Jędrzejowicz, B. Trawiński, & G. Vossen (Eds.), *Proceedings of Computational Collective Intelligence: Lecture Notes in Computer Science*, (Vol. 10449, pp. 509–518). Cham: Springer.
- Mashanov, A., Ábdirahmanov, A., & Sadyqov, J. (1960). *Geologua termunderiniñ oryssh-qazaqsha sózdigi* [Russian-Kazakh dictionary of geological terms]. Almaty: Publishing House of the Academy of Sciences of the Kazakh SSR.
- Mashanov, A. Mýsin, A., & Aitashev, Ğ. (1959). *Metallýrgua ónerkásibi, tay-ken isi jáne fizika termunderiniñ oryssh-qazaqsha sózdigi* [Russian-Kazakh dictionary of terms of metallurgy, mining and physics]. Almaty: Publishing House of the Academy of Sciences of the Kazakh SSR.
- Medeubekuly, S., Omarova, B., & Abdiyeva, R. (2013). *Baspa jjáne poligrafua termunderiniñ túsindirme sózdigi* [Dictionary of polygraphs terms]. Almaty: Sózdik Slovar.
- Myńbaev, K., Shynybaev, M., & Baimirov, M. (1983). *Aýyl sharýashylygyn mehanikalandyrydyñ oryssh-qazaqsha sózdigi* [Russian-Kazakh dictionary of agricultural mechanization]. Almaty: Nauka.
- Niyazbek, M., & Talp, K. (2019). The Kazakh terminology recognition system in the IT field. *IOP Conference Series: Earth and Environmental Science*, 371(5), 052028.
- Omarov, E. (1923). *Fizika pán sózderi. Fizika* [Physics subject words. Physics]. Astana: IC-Servis.
- Omarshuly, D. (2013). *Zańnama úgymdarynyñ tezayrýsy* [Thesaurus of legal terms]. Almaty: Sózdik Slovar.
- Oztopcu, K., Abuov, Zh., Kambarov, N., & Azemoun, Y. (1996). *Dictionary of the Turkic languages*. London and New York: Routledge.
- Qudaibergenov, R. (2009). *Tehnikalyq terminder sózdigi* [Dictionary of technical terms]. Almaty: Taimas.
- Qulmanov, S. (2009). *Oryssha-qazaqsha terminologualyq anyqtamalyq sózdik (Sáýlet jáne qurylys salasy)* [Russian-Kazakh terminological dictionary. Architecture and construction]. Almaty: Memlekettik Tildi Damytý Ortalygy.
- Rakhimova, D., & Zhumanov, Z. (2017). Complex technology of machine translation resources extension for the Kazakh language. *Advanced Topics in Intelligent Information and Database Systems, Studies in Computational Intelligence*, 710, 297–307.
- Reformatsky, A. (1961). *Chto takoe termin i terminologiya. Voprosy terminologii* [What is a term and terminology. Questions of terminology]. Moscow: Nauka.
- Sambetbayeva, M. A., Fedotov, A. M., Tusupov, J. A., Sagnayeva, S. K., Bapanov, A. A., Nurgulzhanova, A. N., & Yerimbetova, A. S. (2016). Using the thesaurus to develop it inquiry systems. *Journal of Theoretical and Applied Information Technology*, 86(1), 44-61.
- Serikbaev, D., Joldasbekov, O., Tajibaev, S., & Abdrahmanov, A. (1974). *Rýssko-kazahsku terminologichesku slovar po mashinostroeniyú* [Russian-Kazakh terminological dictionary on machine building]. Almaty: Publishing House of the Academy of Sciences of the Kazakh SSR.
- Sharipbay, A., Razakhova, B., Mukanova, A., Yergesh, B., & Yelibayeva, G. (2019). Syntax parsing model of Kazakh simple sentences. In I. Hoballah (Ed.), *Proceedings of DATA'19: Second International Conference on Data Science, E-Learning and Information Systems* (pp. 1–5). New York: Association for Computer Machinery.
- Superanskaya, A., Podolskaya, N., & Vassilyeva, N. (2012). *Obshchaya terminologiya: Voprosy teorii* [General terminology: Questions of theory]. Moscow: Librokom.

- Sydyqov, J., & Qabiev, F. (1993). *Gidrogeologua men injenerlik geologua jónindegi oryssh-qazaqsha terminologualyq sózдіk* [Russian-Kazakh terminological dictionary of hydrogeology and engineering geology]. Almaty: Nauka.
- Toksanbay, S. (2013). *Poshta termunderiniń jáne poshta qyzmetteri salasynda jii goldanylatyn sóz tirkesteriniń qazaqsha oryssh sózдіgi* [Russian-Kazakh dictionary of post terms and phrases]. Almaty: Oı Sana.
- Tukeyev, U., & Karibayeva, A. (2020). Inferring the complete set of Kazakh endings as a language resource. In M. Hernes, K. Wojtkiewicz, & E. Szczerbicki (Eds.), *Proceedings of Advances in Computational Collective Intelligence: Communications in Computer and Information Science* (Vol. 1287, pp. 741–751). Cham: Springer.
- Turbanbayeva, A., & Tukeyev, U. (2020). The solution of the problem of unknown words under neural machine translation of the Kazakh language. *Journal of Information and Telecommunication*, 5(2), 214–225.
- Tursyn, Q. (2015). *Mass media jáne telekommýnikatsua termunderiniń qazaqsha -oryssh-aǵylshynsha anyqtamalyq sózдіgi* [Kazakh-Russian-English dictionary of mass media and telecommunication terms]. Almaty: Qazaq Entsiklopediıasy.
- Tussupov, J., Sambetbayeva, M., Fedotov, A., Sagnayeva, S., Bapanov, A., Nurgulzhanova, A., & Yerimbetova, A. (2016). Using the thesaurus to develop IT inquiry systems. *Journal of Theoretical and Applied Information Technology*, 86(1), 44–61.
- Yelibayeva, G., Mukanova, A., Sharipbay, A., Zulkhazhav, A., Yergesh, B., & Bekmanova, G. (2019). Metalanguage and knowledgebase for Kazakh morphology. In S. Misra et al. (Eds.), *Proceedings of Computational Science and Its Applications: ICCSA 2019, Lecture Notes in Computer Science*, (Vol. 11619, pp. 693–706). Cham: Springer.
- Yerimbetova, A. (2016). *Baulanystar grammatikasyn zertteý, má tinder relevanttylyǵyn jáne taqyrybyn anyqtay* [Study of grammar of connections, determination of relevance and theme of texts]. Astana: Delta Press.
- Yessenbayev, Z., Kozhirbayev, Z., & Makazhanov, A. (2020). KazNLP: A pipeline for automated processing of texts written in Kazakh language. In A. Karpov, & R. Potapova (Eds.), *Proceedings of Speech and Computer: 22nd International Conference, SPECOM 2020* (pp. 657–666). Berlin, Heidelberg: Springer-Verlag.