



DOI: 10.54919/physics/56.2024.91keu6

Language modernization in applied, experimental, and computer linguistics aspects during the translator training in Kazakhstan

Manshuk Yeskindirova*

L.N. Gumilyov Eurasian National University
010008, 2 Satpayev Str., Astana, Republic of Kazakhstan

Saule Zagatova

Eurasian Humanities Institute
010009, 4 M. Zhumabayev Ave., Astana, Republic of Kazakhstan

Saltanat Meiramova

S. Seifullin Kazakh Agrotechnical University
010011, 62 Zhenis Ave., Astana, Republic of Kazakhstan

Aliya Ayazbayeva

Astana IT University
010000, C1 Mangilik El, Astana, Republic of Kazakhstan

Zhanna Kuzar

L.N. Gumilyov Eurasian National University
010008, 2 Satpayev Str., Astana, Republic of Kazakhstan

Abstract

Relevance. The research relevance is predefined by the need to modernize the educational system of Kazakhstan in terms of training translators using numerous technological online resources for the most effective implementation of the language modernization strategy and improvement of the text quality attributes in a globalized world.

Purpose. The research aims to study the fundamental techniques of applied, experimental, and computer linguistics, which are the basis of the professional activity of the translator, the prospects of national languages with the help of online resources (CAT programs, data corpuses, terminological online dictionaries), evaluation of technological perspectives at the present stage of social development.

Methodology. The methodology of this research is based on a theoretical understanding of the linguistic and interdisciplinary studies of authors, the comparative study of methods and techniques in the context of applied, experimental, computer linguistics, introduced in the Kazakh educational process in translator training.

Results. This research analyzes the main technologies used for language modernization, considering the industry specifics of each considered linguistic branch, the analysis of machine text recognition mechanisms, algorithms used to improve oral speech and written texts. Moreover, the principles and prospects of language modernization were analyzed, the optimized methodology of studying the programs based on the artificial intelligence in the training of translators in

Suggested Citation:

Yeskindirova M, Zagatova S, Meiramova S, Ayazbayeva A, Kuzar Zh. Language modernization in applied, experimental, and computer linguistics aspects during the translator training in Kazakhstan. *Sci Herald Uzhhorod Univ Ser Phys.* 2024;(56):916-927. DOI: 10.54919/physics/56.2024.91keu6

*Corresponding author



Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

Kazakhstan was developed. The materials presented in this research can be used to further improve the professional skills of translators, to develop pedagogical models related to language modernization.

Conclusions. This study underscores the imperative of integrating modern technological resources into translator training in Kazakhstan to effectively implement language modernization strategies and enhance text quality attributes in a globalized context. The analysis of applied, experimental, computer linguistics techniques, coupled with the evaluation of online resources, highlights the potential for optimizing translator education, advancing the national languages amidst contemporary technological advancements.

Keywords: automated systems; online resources; CAT programs; linguistic methods; educational process.

Introduction

The research relevance is defined by the fact that the present stage of society's development has a whole range of technologies available that expedite effective language modernization, including at the terminological and lexicographic levels. The translator must meet the global pace of development, learning new technologies and methods to ensure constant professional development in different areas, as linguistics is now an application for many branches of knowledge. At this stage it is not enough to know the language, it is also necessary to know a certain field of science (medical, legal, economic) and apply technical capabilities to improve the qualitative features of the text.

The educational system of Kazakhstan at the present stage is being modernized, and constantly updated under the influence of global challenges and information and communication technologies, which should be based on the principles of consistency, conceptuality, efficiency, and accessibility. According to S. Askarova et al. [1], in the course of this process the relationship between the three-component structure "student-computer-teacher" becomes less tense. Educating a new generation in the context of digitalization is a key task of the modern educational process, which includes the following competencies: creativity, global thinking, creative approach to problem-solving, development of logic, and intelligence.

K. Church and M. Liberman [2] note that computer linguistics has been constantly changing over the past decades, particularly in terms of motives, methods, and applications. With the introduction of computers, linguistic analysis was marked by the introduction of new information theory methods that aligned with ideas prevalent in psychology and philosophy. Then a formal language theory and the idea of AI were introduced, synchronized with the development of cognitive science. This was followed by a revival of empiricism and the appearance of applied statistics.

The main directions of computer linguistics, actualized in the work of A. Zagnitko [3], are linguistic databases, corpus linguistics, computer lexicography, automatic text analysis, quantitative linguistics. Moreover, theoretical and applied aspects of the machine, automated and traditional translation, theoretical and applied discourseology, theoretical and applied aspects of Germanic and Romance linguistics. Additionally, media linguistics, linguopersonology: theory, practice, status. Finally, linguo-computer technology of modern methods of higher education.

Applied linguistics provides integrated learning, providing for the formation of three key competencies:

knowledge of a foreign language, professional training (obtaining specialized knowledge), and knowledge of computer technology used in the current stage of the study of a particular industry (medicine, economics, banking, philology). Interdisciplinary integration of this knowledge area determines the wide use of information technologies and didactic materials on foreign language education in translator training [4].

K. Zhaiykbay and T. Kulgildinova [5] point out that Kazakhstani translators should be able to interpret and generate information in at least two foreign languages after successful training. Interpreter training should not only be based on learning lexical relationships, but also on the proper grammatical structure of sentences with a focus on industry specifics.

A harmonic formulation of thoughts and presentation of information is only possible with the systematization of all acquired knowledge, including computer linguistics and experimental modeling. The research object was an analysis of the application of modern technologies with the help of applied, experimental, and computer-based methods on the example of translator training in Kazakhstan.

The research aims to study the main methods and technologies required for a translator to implement professional activities, considering the key perspectives in the field of applied, experimental, and computer-based linguistics, development of translator training to meet modern requirements for this professional activity, evaluation of the language modernization prospects in the future.

Materials and Methods

The methodological basis of this research is based on the theoretical study of the issues using the analysis and synthesis method. It allows drawing parallels between different methods and ways of implementing language modernization strategy in the context of global technologization and digitalization. It is important to assess the problems of implementing technological processes in the education translators in Kazakhstan.

The research analyzes the theoretical aspects associated with the construction of the educational process with the application of technical competence for linguistic improvement methods and algorithms based on machine and automatic online resources. The comparative and contrastive method, the method of structural analysis, and the method of generalization were also applied.

In the first stage of the research, the theoretical characteristics of the language modernization project implementation were considered: applied linguistics as a field of interdisciplinary knowledge, experimental

linguistics and empirical foundations of the experiment, modeling of linguistic attributes using computer linguistics methods. Moreover, achievements of artificial intelligence, global use of English and the issue of national language formation and development in the context. Additionally, exoglossal modernization, cultural conditionality of language sign development.

Finally, strategies and approaches have been analyzed in terms of current developments in the technological branches of knowledge and perspectives in terms of implementing language policy and language modernization to improve oral and written speech. Text fragments from foreign articles were pre-processed and translated into Russian.

In the second stage of the research, the method of comparative and comparative analysis of the world and Kazakhstan contexts in obtaining a translation education and analysis of the main effective techniques that should be used in the training of translators for further professional implementation in different fields of knowledge was applied. The branches of applied, experimental, and computer linguistics were also analyzed and compared in terms of the prospects of language modernization. Moreover, the mechanisms, approaches, opportunities, and limitations of each of the above areas of linguistics in Kazakhstan were considered.

Structural analysis was also used to address the holistic linguistic field and highlight the key segments of linguistic knowledge. A generalization was used on technological online resources that are most effective in terms of solving translation tasks in the processing of natural language to improve its lexico-semantic, grammatical, and syntactic features, the formation of the model of development of national languages in the course of translation work with texts.

The final stage of the research concludes with the results of the study, summarized in terms of linguistic modernization with the help of online linguistic resources (terminological databases, CAT (computer-assisted translation) programs), search algorithms, and experimental models. The main achievements in terms of the technological aspect of linguistic and translation training at the present stage in Kazakhstan were noted. Key areas requiring special attention in the future were highlighted, and the focus was placed on the priority tasks that must be solved in the process of translation training in the coming years.

Results

Language modernization is the process by which the national language is consolidated and used as a means of communication at different levels and in different branches of knowledge: economic, cultural, political, and social. In this context, it is necessary to discuss the enormous role of language policy.

It provides for a conscious state influence to ensure all the conditions for the functioning and development of language in society in the territory of a particular country. Linguistic security is also related to the concept of language modernization, as it provides for the protection of the national language within the framework of ethnocultural specificity, being in the context of national and information security.

On this basis, not only the role of national languages and their support by mass media is emphasized, but also the vector of learning several languages even in adulthood is set. Knowledge of English is seen as a desire to be part of the European community, to comply with the principles of the global world, and is interpreted as a willingness to integrate into global economic and social processes [6].

Modern training of translators in Kazakhstan is determined primarily by the political, economic, cultural, and social situation. Currently, the language policy of the state is determined by the formation of communicative skills in the direction of three languages (Kazakh as the national language, Russian as the language of interethnic communication, and English as the language of international cooperation).

The implementation of this strategy will allow translators who graduated from Kazakhstani universities to find jobs in the domestic and global labor market and to be flexible and mobile in the socio-professional aspect. Language modernization should focus on the direction of increasing the level of knowledge of the national language with a focus on the development of multilingualism in the translation environment, which is caused by the fact that more than 126 ethnic groups live in Kazakhstan. An important achievement is the creation of experimental information platforms for multilingual education in Kazakh schools and universities [7].

Applied linguistics is devoted to solving practical problems in translation with the use of information technology and is oriented to students seeking to receive training in translation and gain fundamental skills in translation. It also focuses on developing the ability to solve problems in the field of computer processing of texts created in natural language, the use of information search resources, and the formation of media links.

During training students acquire fundamental knowledge of a foreign language by studying the theoretical and practical basics of translation, applied linguistics, programming, structural linguistics, mathematical modeling, and automation. As a rule, students undergo practical training at state enterprises and commercial organizations, where they gain experience in the field of communication, and computer processing of texts, and learn how to properly use search systems. The main issues that are the subject of applied linguistics: teaching a second foreign language through different lingua didactic methods, identifying the extent and characteristics of language performance, bilingualism and individual approach, multilingualism and society, linguistic policy and planning, language assessment and technological applications, translation into a foreign language [8].

Experimental linguistics is a science that deals with the collection of information obtained on an empirical basis, to assess the prospects of creating language patterns, and the regulation of modeling the means of the selected language. At first, in the processing and study of empirical data, the main attention is focused on the techniques of collecting information, and their advantages and disadvantages are determined. Then experimental research of any language ability (for example, the use of a given grammatical form) with the use of advanced developments in the field of verbal working memory is then carried out.

As language use varies with the social context, the results of experiments do not always reproduce at the expected rate. It is impossible to establish complete control of the linguistic translators over language, unlike in the physical sciences, where the parameters of an experiment can remain stable with each iteration, each repetition of a linguistic experiment is a unique social event, so the results obtained each time may vary significantly or insignificantly. Thus, it is important to choose the right experimental methods that will give the most accurate data to derive average values and further use them for linguistic modernization [9].

Computer linguistics investigates how and in what way human language can be automatically processed and interpreted, uses mathematical and logical characteristics of natural language, develops algorithms, and studies statistical phenomena for automatic language processing. Computer-assisted language is used in search engines, and the generation of natural speech is based on the development of computational ability. The perception, understanding, and reproduction of the natural language

process is a key task of computer linguistics. Among the main areas of focus are the following:

1. Corpus linguistics – the subject of study is the development of a conceptual and procedural apparatus for the formation of corpora of texts (text-oriented databases) and their analysis.

2. Automatic text analysis – extraction of grammatical and semantic information from the natural speech stream, built according to a specific algorithm.

3. Machine translation is the process of transforming a text formatted using one natural language into a text formatted through another natural language with the help of computer technology.

4. Dialogue systems – simulating the process of communication with the help of a message analysis unit and a response formation unit using computer technology.

5. Expert systems, based on the selection of information from texts, can change the nature of the information received from the speaker, and change the structure or form of the dialogue by recreating special blocks of new communication scenarios.

6. Automated speech recognition and synthesis process – modeling human speech using computer technology into proper speech units [10].

Linguistic modernization in the field of applied linguistics is carried out in the following directions: creation and implementation of new artificial languages to improve the qualitative parameters of the text, discovery of linguistic criteria for the development of search technologies. Moreover, improvement of lexicographic dictionaries and corpus data, formation of the concept of unification of terminology and restoration of a complete statement employing speech segments. The results in

applied linguistics can be achieved by applying the method of classification and the method of modeling, through which the empirical observation of linguistic patterns is performed, and the foundations and conditions of their formation are studied.

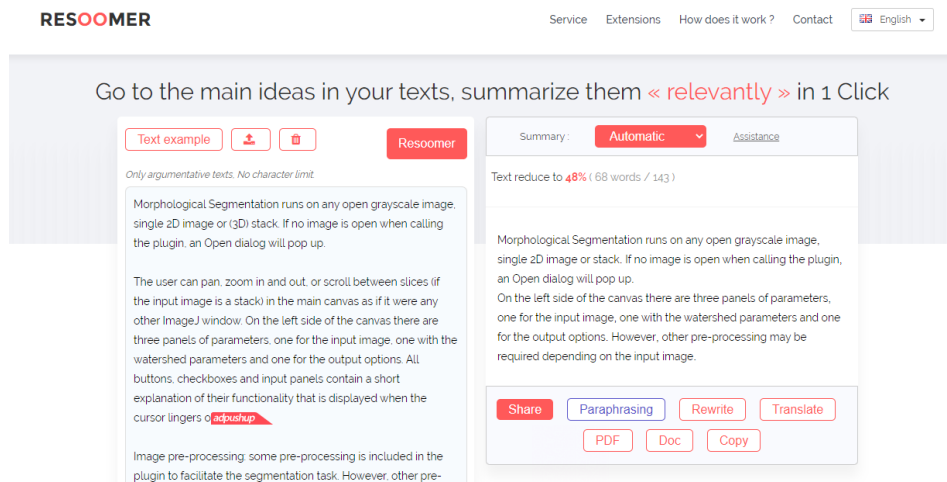
Currently, in the field of experimental, linguistics technologies are actively introduced, such as the automation of the text annotation process, the use of AI for the preservation and development of national languages, the tracking of lexicographic processes occurring in online terminological dictionaries and corpus databases. Additionally, the modeling of interaction patterns between different languages, the establishment of identity through the machine-identification of speech segments. The principle of the experimental method is to track the response to a stimulus at the cognitive level with the use of certain linguistic units. For linguistic modernization, methods of constructing rating scales, reconstruction of lingual clichés into a coherent statement, use of direct lexical meanings, classification, and associative experiment are used.

Computer linguistics conducts language modernization at the level of developing machine translation technologies, in particular online programs and applications, recognizing speech segments in text, automatic processing of written and spoken speech. Moreover, building management systems for automatic control and information retrieval, and creating online thesauruses.

The most accessible advances in computer technology that every translator should be able to use are the following: hyphenation, intended to correctly separate words in natural language texts. Additionally, the detection and correction of grammatical errors, a task of computational linguistics and is done by analyzing adjacent words in a sentence or even in the whole sentence, the correction of stylistic errors, based on the laws of correct use of words and phrases in oral speech or certain literature style [11].

During the translator's practical training, several different methods should be used for linguistic modernization. In the stylistics course, it is recommended to use the text summarization technique, which allows the creation of linguistic sequences in text format that can affect the improvement of the characteristics of the whole text.

For example, used to reduce oral and written statements to make them more concise. For example, students are encouraged to abstract a text using the Resoomer program: first, combine all important information in one file, copy and paste it into the left field, then select the mode: automatic, manual, optimized, or textual analysis. The results in automatic mode are shown: the original text was reduced by 48% (Figure 1).



Resoomer allows translating a text to one of the listed languages (Figure 2), rewriting, and paraphrasing it by selecting one of the available variants. Abstracting results can be copied or saved as Word or PDF files. Students can

use this program to prepare projects, summaries, and abstracts or to make unique texts.

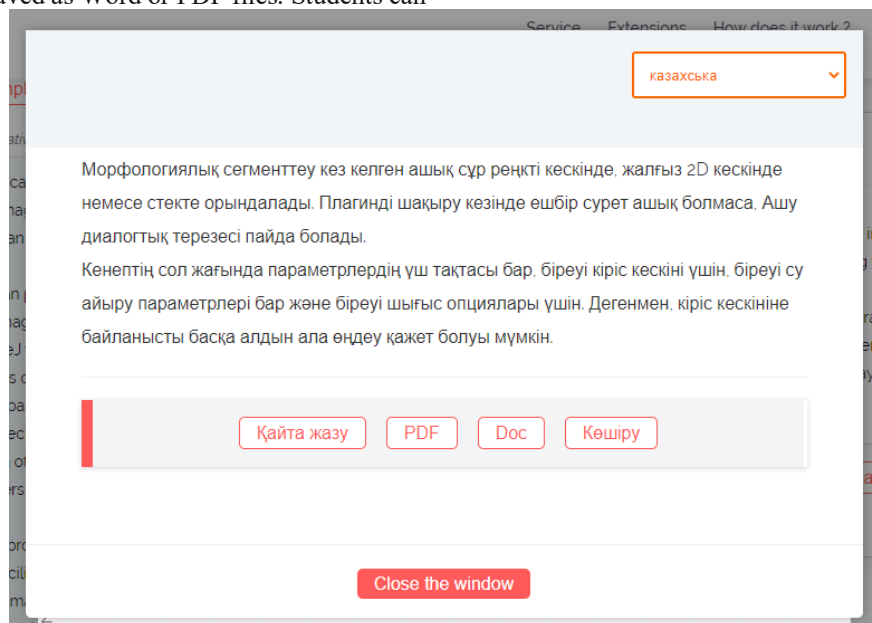


Figure 2. Results of an English-Kazakhstan translation of a text

specialty should necessarily include an introductory lecture on the method of morphological segmentation on the example of the languages studied: the source and target languages.

Parts Of Speech (POS) technology also plays an important role in this context, allowing the assignment of part-of-speech codes to lexical and textual units contained in a corpus for syntactic and semantic analysis. Translation students should learn modern algorithms for part-of-speech tagging as well as online programs: Ninja-PoS, FastTag, Topia TermExtractor, Stanford Log-linear, Stagger, petraTAG Part-of-speech tagger, Free CLAWS WWW tagger (Figure 3).

Since Kazakh is an agglutinative language, affixes have a morphological function besides the word-formation function, determining the tense, the verb mode, or the case of a noun. A morphological course in the "Translation"

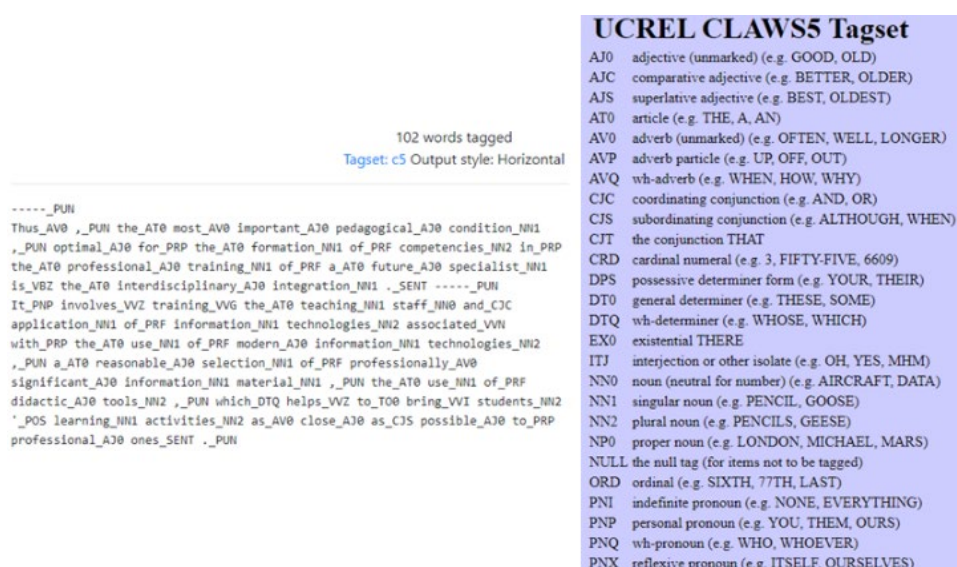


Figure 3. Results of the part-of-speech tagging in the Free CLAWS WWW tagger

Computational lexicography is one of the main achievements of modern computational linguistics, which includes research design, construction, and use of electronic dictionaries in natural language processing, recording relevant linguistic information from the database, and editing lexicographic titles and textual information. This technology allows the analysis of old data corpuses and terminological systems of different national languages and then carries out the modernization of the terminological system by adding new units, building synonymous terminological series, and deleting obsolete lexemes and examples of their use.

Translation training should use Word Sense Disambiguation technology, which allows the automatic selection of a term from an online resource. The lexicographic course should familiarize with the WordNet English Semantic Dictionary, in which lexemes are divided into synonymic groups (synsets), and brief information on the general definition and semantic relations between these words are provided. Symbol "S" means to show synset (semantic) relations, and "W" means to show word (lexical) relations. In practical lessons, this program can be used to assess the frequency of use of a particular meaning in different contexts (Figure 4).



Figure 4. Demonstration of WordNet results for "pen"

Machine-readable technologies are used for automatic lexical analysis: searching for words, word sequences, and parts of speech in a text set, and using data manipulation to

enrich the language. Natural-language understanding (NLU) technology is built not only on the knowledge of language grammar but also on background knowledge,

involves the transformation of input data in natural language, and is organized around three levels of word representation: syntactic, lexical, and logical. In translation training, such programs should be presented in lecture format for familiarization: STUDENT, ELIZA, ATN, SHRDLU. Natural Language Processing (NLP) technology allows to read and understand languages using machine technology, allows the use of user interfaces in natural language, and obtains knowledge directly from written sources.

Improving the efficiency of oral and written language processing allows the use of new statistics and experimentation with phonemes, morphemes, lexemes, and syntaxes, modeling speech patterns that are most

understandable and accessible to each native speaker. Translators should familiarize themselves with NLP technologies using online platforms. For example, Watson Discovery can be used to translate corporate documents (analysis of tables, PDFs, templates, hidden and unstructured data). Watson Natural Language Understanding can be used in class to analyze web pages and social media to identify concepts, keywords, categories, semantics and emotions. Watson Assistant can improve customer service while reducing costs. IBM Watson Annotator for Clinical Data enables textual analysis and concept extraction in healthcare and life sciences (Figure 5).

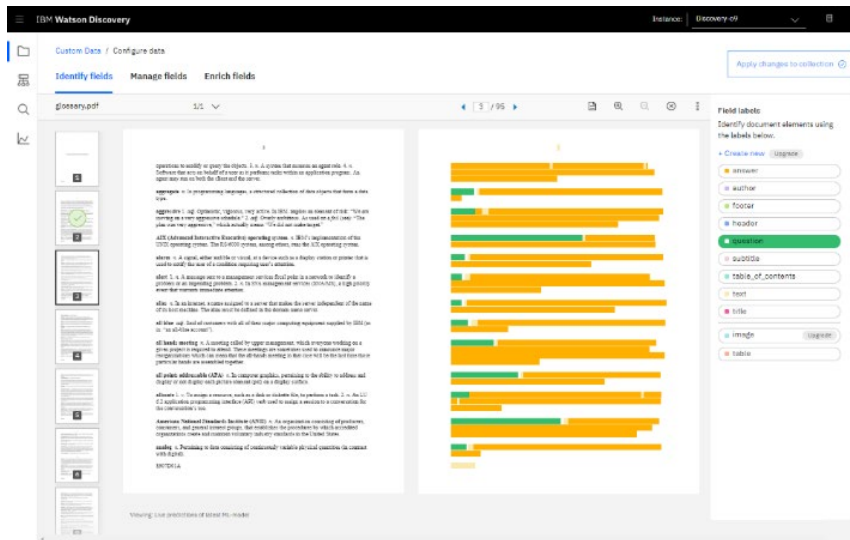


Figure 5. Transformation of a text in Watson Discovery

Audio and video search aims to extract information from audio and video sources not only by metadata but also by content search, including automatic speech recognition, and image similarity matching. The development of multilingual multimedia content is based on the blending of language with other modes of communication and information media in digital texts (e.g., speech with

gestures and facial expressions, text with images and sounds).

Optical Character Recognition (OCR) assists in identifying the appropriate or related language from an image of printed text. Translators should be trained to use basic OCR software that translates text from uneditable to editable formats: ABBYY Fine Reader (Figure 6), Brainware, COCR2, and CuneiForm.

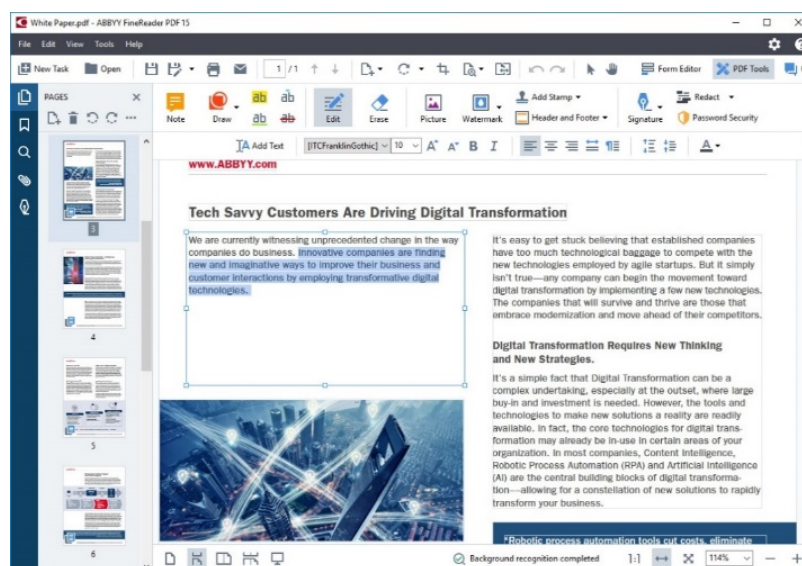


Figure 6. ABBYY Fine Reader interface

The development of Natural Language Interface to Data Bases (NLI-DB) technology is necessary for storing and managing large amounts of data (in railroad reservation, library science, banking, management, and information). This technology is required for use by translators, in particular in the training of applied linguists. The main objective of NLI-DB is to eliminate a barrier between a user and special computer systems, using natural language to enter queries. The NLI (Natural Language Interface) technique, based on experimental questioning, involves developing a natural language interface based on queries in a database, searching for information from texts, its main task is to provide answers to questions posed in natural language.

Automatic translation tools at this stage cannot achieve semantic-syntactic adequacy, but in terms of genre-thematic focus, there is a positive trend. For example, Google Translate intelligently selects appropriate verbal and phraseological equivalents, but there are problems with the construction of grammatical structures, so the improvement of grammatical patterns is the primary task of online translators. In addition, automatic systems operate based on neural networks, which allow the translation of phrases and syntactic constructions considering the context [12-14].

Translators should master Computer-Assisted Translation (CAT) software, which can memorize text fragments and automatically substitute matching segments, thus saving translation time and making it better in terms of terminology. CAT tools also help in the translation of author techniques, idioms, and phrases. Among the most popular programs to be studied in translation courses are the following:

1. Trados Studio is a program for professional translation and documents, presentations, and HTML (HyperText Markup Language) files, suitable for large projects.
2. SmartCAT is a program with a user-friendly and easy-to-use interface, usually used for small projects.
3. MemoQ is often used by programmers and technical specialists for translation purposes.
4. Wordfast allows the configuration of the functions by yourself and has an intuitive interface.
5. Déjà Vu is a program that allows working with 32 formats, providing a convenient way to work with glossaries.
6. Transit NXT Professional is a website and application translation tool that monitors contexts and allows the stylistic adaptation of the text.

Translator training should include a lecture course and practical sessions on machine translation using 3-5 of the above programs, the teachers should focus on the key advantages and disadvantages, capabilities and limitations of these programs, and the basic algorithms of machine translation implementation. After the course, Trados Studio, SmartCAT, and others can be used to complete homework assignments.

Today's linguistic modernization is related to application clustering, sentiment analysis, chatbots, knowledge extraction, the development of natural language interfaces, and content filtering. This process must be carried out at all linguistic levels, from phonetic to textual,

and must necessarily be supported by the language policy of the state. Modern translators in Kazakhstan should be trained in machine translation as these are the resources that will further facilitate professional activities, text template formation in the target language, and post-editing.

Discussion

Applied linguistics is the use of linguistic theories, methods, and results to study issues and solve problems related to language, present in other fields of scientific knowledge. This linguistic branch involves acquiring knowledge in other interdisciplinary fields along with language skills. These include native language mastery, computer linguistics, forensic linguistics, logopedics, neurolinguistics, second language mastery. Moreover, clinical linguistics, educational linguistics, lexicography, translation and stylistics, sociolinguistics and psycholinguistics, language and economy, language and culture. Finally, political discourse, discourse analysis, language, and law, corporal linguistics [8; 15].

Linguists seek to grasp language through a variety of methodologies, ranging from a rationalistic self-reflection on the nature of general linguistic knowledge to an empirical investigation of specific uses of language. Empirical research is often comparative in nature, which involves looking for relationships between dependent linguistic variables and an almost infinite set of independent linguistic and extra-linguistic variables, including independent variables related to time, language, setting, manner, task, and speaker [16-18]. Experimental linguistics, according to researcher J. Grieve [9], also collects data by manipulating context and exposing subjects to certain stimuli to minimize social influences on the process of language production or perception. This can be accomplished by assigning participants to groups before the experiment, each of which will be exposed to different factors to maximize the elimination of unknown characteristics in language practice [19; 20].

The empirical approach to linguistics involves describing natural changes in language, for example, depending on dialect, case, and speaker. It helps in the identification of real differences between populations, but it is impossible to establish cause-effect relationships using this method of observation [9; 21].

The main objective of computational linguistics, following N. Pushik [10], is to model the processes of basic linguistics, analysis, and synthesis of written and spoken language, improve technical tools, and improve operational capabilities for language modernization. Natural language processing addresses several practical problems: machine translation, text summarization, speech stream recognition and generation, information retrieval, and text content analysis.

Advances in artificial intelligence are directly related to computational linguistics and computer science, as language modeling is computer-assisted and allows the automation of language mechanisms with a focus on recreating the subtleties, dialects, emotional background, and context of each national instrument of cognitive experience. Native speakers today have close contact with English through the following branches: business, politics, the Internet, mass media, and popular culture. In the era of

globalization, economic interactions between countries are almost always conducted in English, as are global politics and diplomacy [22-24]. Endangered languages must be supported by government planning. E. Prifti [25] identifies the main reasons for borrowing new words:

- close contact between the two national languages, particularly in multilingual groups;
- the mixing of elements from different languages;
- the dominance of some languages over others (in the cultural, economic, political, and religious spheres);
- the feeling of necessity in explaining certain material with an instructional purpose;
- the prestige associated with the use of words from another language.

Considering the issues of language modernization and politics, the linguist S.N. Sridhar [26] defines macro-strategy – the selection of language as the main source of lexical development. But also focuses on exoglossal modernization, affecting the structure of language in terms of its productive morphological processes, stratification of phonological and morphological system, expansion of syntactic and rhetorical properties.

S. Ibrahim and S.M. Awan [27] use the example of the Potohari language to develop a language modernization strategy that allows language preservation only in connection with culture (preservation of values, norms, traditions, and ways of life). The authors are committed to the concept of preserving culture through the native language instead of replacing it with other languages, and consider the role of cultural integration, which affects education, communication, and social forms of interaction. A study by O. Nnamani [28] indicated how important effective reading skills are for learning grammar, interviewing and questionnaires were used to determine the level of reading interest in students.

The study by K. Nowakowski et al. [29] presented findings on the improvement of POST-AL, a computer processing tool for the endangered Ainu language. Improvements were made to algorithms for text transcription, word segmentation, and part-of-speech labeling, and the vocabulary base was expanded by combining corpus data from two dictionaries. The researchers believe that the use of modern tools such as SVMTool and Stanford Log-Linear Tagger to create a corpus based on the principle of morphological segmentation of speech is promising.

Statistical language models will allow the development of further language technologies for speech recognition and machine translation, speech synthesizer development, and the creation of online dictionaries in mobile applications. The work of P. Heinrich [30] describes lexical modernization in four states (Japan, China, Korea, and Vietnam) in which transcultural integration was actively used. Japan was the first country to undergo linguistic modernization in East Asia, in the course of which terminological gaps were filled.

The spread of Sino-Japanese modern terminology through the characters of Chinese writing proves that writing has a direct influence on language modernization, being a "transcultural and transnational resource". T.B. Roettger [31]) suggests the concept of pre-registration, which can be used to increase the reproducibility of

experimental linguistic research. Pre-registration is a time-stamped document in which researchers enter information about the planning of data collection and analysis, helps draw the line between exploratory and confirmatory research, transparently tracks analytic flexibility, and counteracts publication bias.

Linguists D. Hershcovich et al. [32], explore how cultural knowledge, preferences, and values can influence NLP practices. The authors agree with the view that modern computational science has inherited colonizing practices that are realized as "homogenization of perspectives" and "algorithmic monoculture," but express their thoughts on the modernization of national language models. However, the representation of a particular language in the cultural plane is not enough, it is also necessary to represent cultures outside Western, educated, wealthy and democratic countries, showing its features and characteristics due to cultural, socio-economic, and political influences [33-35].

E. Gerfanova [36] considers the development of special teaching materials to implement the strategy of professional trilingual education, the introduction of innovative methods, in particular e-learning, as well as training of qualified teaching staff who could conduct classes in the three languages as a priority objective for foreign language learning in Kazakhstan. It is possible to agree with the researcher's opinion that the effectiveness of the national language modernization depends on the revision of the status of the Kazakh language as the only state language, which will give it a basis for further development.

As such, ensuring language modernization and the quality of the educational process, in particular in the training of translators, must be complex. Modern research is based on the study of factors contributing to the improvement of language parameters, and the development of new methods and approaches to the implementation of language policy to preserve national languages and ensure their further development in the context of globalization. The priority in the field of language modernization is the development of new pedagogical strategies related to the formation of linguistic, technological, and subject competence to further professional implementation.

Conclusions

The technological approach in the training of Kazakhstani translators is aimed at setting and achieving educational goals in terms of creating a cycle of operations to be performed algorithmically. With the help of methods and tools of applied, experimental, and computer linguistics translators can predict the results with a high degree of accuracy, systematize scientific bases, corpus data. Additionally, accumulated practical experience, improve the quality of machine translation in terms of terminology, lexicography, grammar, and syntax design. Moreover, participate in joint development of templates to effectively translate the source text by the target language based on statistical and experimental data.

The research analyzes the main ways and methods used in applied, experimental, and computer linguistics for language modernization, processing of speech samples, text recognition, development of new machine-readable mechanisms, and implementation of new algorithms and

patterns to improve both oral and written speech. The key principles and contributions of different linguistic and applied branches to language modernization at the present stage of development of society in Kazakhstan in the context of globalization have been considered.

Attention focused on the training of translators, in particular the study of methods and techniques to align text fragments at different linguistic levels, build semantic and grammatical chains, optimize terminological system and corpus data in different branches of knowledge (the method of understanding the natural language NLU, NLP natural language processing methodology).

Moreover, the method of optimized application of these methods in practical training of a translator was developed. The main CAT programs allowing the implementation of the language modernization strategy are considered: Trados Studio, SmartCAT, MemoQ, and Transit NXT Professional.

Future prospective directions are the creation of new linguistic templates and algorithms to improve the speech

flow and improve the quality attributes of natural language translation, optimization of terminological resources using corpus databases. Additionally, creation of interfaces and chatbots in natural language, modeling types of interlanguage interaction, and development of artificial languages. The priority vector of language modernization in the training of translators in Kazakhstan is the study of automatic programs to optimize lexical, stylistic, and terminological online resources to unify and standardize translation into target languages.

Acknowledgements

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19679666).

Conflict of Interest

None.

References

- [1] Askarova S, Tursunbayeva K, Kulekenova Zh. Development of information and telecommunication technologies in the education system of Kazakhstan. *Sci Col "Interconf"*. 2021;44-50.
- [2] Church K, Liberman M. The Future of Computational Linguistics: On Beyond Alchemy. *Front Art Intel*. 2021;4:625341.
- [3] Zagnitko A. Typology of text modifications: application aspect. *Ling Comp Res*. 2022;15:79-83.
- [4] Dubinka SA, Dubinka-Hushcha LA. Communicative Competence as the Priority in Professional Language Training. *Adv Soc Sci Ed Hum Res*. 2021;537:1-8.
- [5] Zhaiykbay K, Kulgildinova T. Conceptual basis for the formation of foreign language pragma professional communicative competence of future IT specialists. *Bull Kazakh Nat Wom Ped Univ*. 2020;4(84):80-87.
- [6] Tarasyuk YuM. The concept of linguistic security and language policy of Turkey in the context of the modernization process. *Bull ONU Mechnikova*. 2016;21(4):70-76.
- [7] Yeskeldiyeva B, Tazhibayeva S. Multilingualism in Modern Kazakhstan: New Challenges. *As Soc Sci*. 2015;11(6):54-64.
- [8] Betti MJ. Applied Linguistics. 2021. <https://goo.su/JuiJ>.
- [9] Grieve J. Observation, experimentation, and replication in linguistics. *Linguistics*. 2021;59(5):1343-1356.
- [10] Pushik N. Computer Linguistics and Artificial Intelligence. *Young Sci*. 2021;2(90):151-155.
- [11] Ahmad H. Applications of Computational Linguistics. *Comp Ling*. 2022;1-8.
- [12] Gudmanyana A, Sitko A, Struk I. Pragmatic adequacy of the machine translation. In: Social communications in the innovative educational space: theoretical and applied aspects. Kyiv: National Aviation University; 2020.
- [13] Moldabayeva G, Suleimenova R, Buktukov N, Mergenov M. Improvement of oil field development using enhanced oil recovery methods. *Nauk Visn Nats Hirnych Univer*. 2021;6:23-28.
- [14] Moldabayeva GZ, Suleimenova RT, Bimagambetov KB, Logvinenko A, Tuzelbayeva SR. Experimental studies of chemical and technological characteristics of cross-linked polymer systems applied in flow-diversion technologies. *News Nat Acad Sci Republ Kazakhstan, Ser Geol Techn Sci*. 2021;4(448):50-58.
- [15] Akberdin AA, Kim AS, Tolymbekova LB, Sultangazyev RB. Development of Technology for Producing Complex Boron-Containing Aluminum-Silicon Ferroalloy. *Metall*. 2023;67(5-6):749-756.
- [16] Panfilova A, Mohylnytska A, Gamayunova V, Fedorchuk M, Drobitko A, Tyshchenko S. Modeling the impact of weather and climatic conditions and nutrition variants on the yield of spring barley varieties (*Hordeum vulgare* L.). *Agr Res*. 2020;18(Special Issue 2):1388-1403.
- [17] Panfilova A, Korkhova M, Gamayunova V, Fedorchuk M, Drobitko A, Nikonchuk N, Kovalenko O. Formation of photosynthetic and grain yield of spring barley (*Hordeum vulgare* L.) depend on varietal characteristics and plant growth regulators. *Agr Res*. 2019;17(2):608-620.
- [18] Bagasharova ZT, Abdelmaksoud AS, Abdugaliyeva GY, Sabirova LB, Moldabayeva GZ. Recovery of water aquifers after the impact of in-situ leaching of Uranium. *Int Multidisc Sci GeoConf Surv Geol Min Ecol Manag, SGEM*. 2015;1(4):19-26.
- [19] Kropachev R, Imanov M, Borisevich Y, Dhomane I. Information technologies and the future of education in the Republic of Kazakhstan. *Sci J Astana IT Univ*. 2020;30-38.

- [20] Deryaev AR. Features of forecasting abnormally high reservoir pressures when drilling wells in the areas of Southwestern Turkmenistan. *SOCAR Proceed.* 2023;1:7-12.
- [21] Deryaev AR. Analysis of the opening of zones with abnormally high reservoir pressures in the oil and gas fields of the Western part of Turkmenistan. *SOCAR Proceed.* 2023;1:20-25.
- [22] Doroshkevich AS, Zakharova AS, Oksengendler BL, Lyubchyk AI, Lyubchyk SI, Lyubchyk SB, Tatarinova AA, Kirillov AK, Vasilenko TA, Gorban OO et al. The Rectifying Contact of Hydrated Different Size YSZ Nanoparticles for Advanced Electronics. *Nanomater.* 2022;12(24):4493.
- [23] Khardazi S, Zaitouni H, Neqali A, Lyubchyk S, Mezzane D, Amjoud M, Choukri E, Kutnjak Z. Enhanced thermal stability of dielectric and energy storage properties in 0.4BCZT-0.6BTSn lead-free ceramics elaborated by sol-gel method. *J Phys Chem Sol.* 2023;177:111302.
- [24] Ahmadov F, Ahmadov G, Akbarov R, Aktag A, Budak E, Doganci E, Gurer U, Holik M, Kahraman A, Karaçali H, Lyubchyk S, Lyubchyk A, Mammadli A, Mamedov F, Nuruyev S, Pridal P, Sadigov A, Sadygov Z, Urban O, Yilmaz E, Yilmaz O, Zich J. Investigation of parameters of new MAPD-3NM silicon photomultipliers. *J Instrument.* 2022;17(1):C01001.
- [25] Prifti E. Language Modernization vs. Linguistic Protectionism. 2009. <https://files.eric.ed.gov/fulltext/ED507223.pdf>.
- [26] Sridhar SN. Language Modernization: Structural and Sociolinguistic Aspects. 2008. https://www.academia.edu/3736009/Language_Modernization_Structural_and_Sociolinguistic_Aspects.
- [27] Ibrahim S, Awan SM. Modernization, Cultural Integration and Language Modification: A Study of the Potohari Language in Pakistan. *Glob Reg Rev.* 2020;5(1):128-135.
- [28] Nnamani O. Research in English and Applied Linguistics. *J Teach Eng Com Skills.* 2022;10(1):60-69.
- [29] Nowakowski K, Ptaszynski M, Masui F, Momouchi Y. Improving Basic Natural Language Processing Tools for the Ainu Language. *Information.* 2019;10(11):329.
- [30] Heinrich P. Language Modernization in the Chinese Character Cultural Sphere China, Japan, Korea and Vietnam. In: *The Cambridge Handbook of Language Standardization*. Cambridge: Cambridge University Press; 2021.
- [31] Roettger TB. Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics.* 2021;80-83.
- [32] Hershcovich D, Frank S, Heather L, Lhoneux M. Challenges and Strategies in Cross-Cultural NLP. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Dublin: Association for Computational Linguistics; 2022.
- [33] Metaksa G, Moldabaeva G, Alisheva Z. Obtaining preset properties in the hydrogenation process by controlling the state of phase boundary. *E3S Web Conf.* 2018;56:03028.
- [34] Khalbekova HU. Amino acid composition of species of the genus *Atriplex* L. growing in the Aral Sea (Uzbekistan). *E3S Web Conf.* 2023;411:02063.
- [35] Nikitina EV, Ortikov E, Beshko NY, Khalbekova KU. Molecular authentication of some rare Iris (Iridaceae) species from Uzbekistan. *Plant Sci Today.* 2023;10(3):444-454.
- [36] Gerfanova E. Foreign Language Education of Kazakhstan: Current Trends and Future Perspectives. *People Int J Soc Sci.* 2018;4(3):735-745.

Модернізація мови в аспектах прикладної, експериментальної та комп'ютерної лінгвістики під час підготовки перекладачів у Казахстані

Маншук Єскіндірова

Євразійський національний університет ім. Л. Н. Гумільова
010008, вул. Сатпаєва, 2, Астана, Республіка Казахстан

Сауле Загатова

Євразійський гуманітарний інститут
010009, просп. М. Жумабаєва, 4, Астана, Республіка Казахстан

Салтанат Мейрамова

Казахський агротехнічний університет ім. С. Сейфулліна
010011, просп. Женис, 62, Астана, Республіка Казахстан

Алія Аязбаєва

Університет Астана ІТ
010000, вул. Мангілік Ел, С1, Астана, Республіка Казахстан

Жанна Кузар

Євразійський національний університет ім. Л. Н. Гумільова
010008, вул. Сатпаєва, 2, Астана, Республіка Казахстан

Анотація

Актуальність. Актуальність дослідження передбачена потребою у модернізації освітньої системи Казахстану щодо навчання перекладачів за допомогою численних технологічних онлайн-ресурсів для найефективнішої реалізації стратегії мовної модернізації та покращення характеристик якості тексту в глобалізованому світі.

Мета. Метою цього дослідження є вивчення фундаментальних методик прикладної, експериментальної та комп'ютерної лінгвістики, які є основою професійної діяльності перекладача, перспектив національних мов за допомогою онлайн-ресурсів (програми САТ, корпуси даних, термінологічні онлайн-словники), оцінка технологічних перспектив на сучасному етапі соціального розвитку.

Методологія. Методологія цього дослідження базується на теоретичному розумінні лінгвістичних та міждисциплінарних досліджень авторів, порівняльному вивченні методів та технік в контексті прикладної, експериментальної, комп'ютерної лінгвістики, введених в освітній процес Казахстану у навчанні перекладачів.

Результати. У цьому дослідженні проаналізовані основні технології, що використовуються для модернізації мови, з урахуванням промислової специфіки кожного розглянутого лінгвістичного напрямку, а також проведений аналіз механізмів визнання машинного тексту, алгоритмів, які використовуються для покращення усної мови та письмових текстів. Більше того, були проаналізовані принципи та перспективи мовної модернізації, розроблена оптимізована методологія вивчення програм на основі штучного інтелекту у навчанні перекладачів в Казахстані. Матеріали, представлені у цьому дослідженні, можуть бути використані для подальшого вдосконалення професійних навичок перекладачів, розробки педагогічних моделей, пов'язаних з модернізацією мови.

Висновки. Це дослідження підкреслює необхідність інтеграції сучасних технологічних ресурсів у навчання перекладачів в Казахстані для ефективної реалізації стратегій мовної модернізації та покращення характеристик якості текстів у глобалізованому контексті. Аналіз прикладних, експериментальних та комп'ютерних лінгвістичних технік, спільно з оцінкою онлайн-ресурсів, підкреслює потенціал для оптимізації освіти перекладачів, просуває національні мови серед сучасних технологічних досягнень.

Ключові слова: автоматизовані системи; онлайн-ресурси; САТ програми; лінгвістичні методи; освітній процес.