Application Of the National Corpus Of The Kazakh Language In Linguistic Research

Pirmanova K.K.¹, Tokmyrzaev D.O.², Pirmanova A.A.³

Abstract

National corpora are essential tools in modern linguistics, offering valuable data for analyzing language structures and their connection to the world. The primary goal of this study is to enhance the perception and analysis of the Kazakh language. It introduces a methodological approach that deepens the understanding of the structure and features of Kazakh. The research utilizes frequency analysis, syntactic parsing, and statistical methods to examine Kazakh language texts. The results demonstrate that the National Corpus provides extensive data for studying linguistic patterns. Frequency analysis identifies the most common words and phrases, while syntactic analysis uncovers patterns and connections between words. The study also highlights the variation in word usage and constructions, showing the language's adaptability to modern communication. The findings are applicable to lexicography, semantics, and syntax studies of the Kazakh language. Additionally, they offer practical benefits in language teaching, lexicographic resource creation, machine translation, and natural language processing. In conclusion, the research provides valuable tools for in-depth analysis of the Kazakh language, with broad applications across linguistic studies and practical fields.

Keywords: Corpus Linguistics; Language Corpus; The National Corpus Of The Kazakh Language; Frequency Dictionary; Frequency Analysis.

Main Points

The essence of the article is to analyze and evaluate the possibilities of applying the National Corpus of the Kazakh Language in the field of linguistic research. The study seeks to gain a deeper understanding of the structure, features, and dynamics of the Kazakh language by utilizing the vast collection of texts available in the corpus.

The research methodology includes frequency analysis and syntactic analysis based on the collected corpus data. Frequency analysis identifies the most frequently used words and expressions, while syntactic analysis reveals syntactic structures and connections between words.

The results of the analysis emphasize the richness and versatility of the National corpus of the Kazakh language. The variability in the use of lexical units reflects the dynamics of the language and its adaptation to the modern communication environment.

This work will not only enrich the methodology of linguistic research but will also be useful in the practice of teaching the Kazakh language and creating lexicographic resources. The study highlights the potential of the corpus for further research, contributing to a deeper analysis of the semantics, structure, and evolution of the Kazakh language.

Introduction

Modern linguistic research strongly depends on access to large volumes of linguistic data, which provide a unique opportunity to analyze language structures and variability. In this context, the National Corpus of Texts become valuable resources, revealing many new ways for researchers to analyze and understand

¹ PhD postdoctoral student of Al-Farabi Kazakh National University, Almaty, Kazakhstan; kunsulu.pirmanova@mail.ru. (Corresponding Author)

² Programmer of Al-Farabi Kazakh National University, Almaty, Kazakhstan; dark.han@mail.ru.

³ PhD student of Al-Farabi Kazakh National University, Almaty, Kazakhstan; pak_mirash@mail.ru.

March 2024

Volume: 3, No: 7, pp. 2806 – 2814 ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)

https://ecohumanism.co.uk/joe/ecohumanism

DOI: https://doi.org/10.62754/joe.v3i7.4418

linguistic phenomena. Within the framework of this article, we pay attention to one of such resources — the National Corpus of the Kazakh language — and its application in linguistic research.

Corpus data are valuable research material, representing a collection of authentic texts in a particular language, covering a variety of genres and topics (Zhubanov, 2017). National Corpus allows systematic analyses of word frequency, syntactic structures, and semantic shades, as well as studying changes in language use over time (Smirnova, 2010).

This study aims to assess the potential of the National Corpus of the Kazakh Language for a deeper analysis of linguistic structures, features, and variation. We aim to reveal the potential of this resource in the context of research related to the Kazakh language and to outline its significance for linguistic science and practical applications.

All this will allow us to consider how the National Corpus of the Kazakh Language becomes an effective tool for a deeper understanding of the structure and peculiarities of this language, as well as for the application of the results in the practice of teaching and research in various fields.

Description of Materials and Methods

The materials for this study were obtained from the National Corpus of the Kazakh Language (qazcorpus.kz), a vast collection of texts of various genres, styles, and topics representing the contemporary use of the Kazakh language. The corpus includes written and spoken texts, as well as texts from various sources, such as literature, mass media, official documents, etc.

The research methodology is based on a combined approach, which includes two main aspects: frequency analysis and syntactic analysis.

Frequency analysis was performed using specialized software tools. Starting with general statistics on the most common words, we further focused on analyzing the frequency of words in various genres and thematic categories. This allowed us to identify keywords that reflect the peculiarities of the linguistic context.

Syntactic analysis (Chomsky, 1957) was carried out using sentence parsing. We used software tools that can automatically disassemble sentences into components and analyze the connections between words. This analysis helped to identify typical syntactic constructions and structures characteristic of the Kazakh language.

Statistical approaches (Labov, 1966) were also used to process frequency and syntactic analysis data. We applied statistical methods to identify structural features and to analyze variations in the use of words and constructions.

The stages of analysis included:

Compilation of the data corpus: collection and preparation of texts from various sources.

Frequency analysis: determination of the most used words and phrases in various texts.

Syntactic analysis: automatic parsing of sentences to identify syntactic structures.

Syntactic analysis: Automatic parsing of sentences to identify syntactic structures.

Statistical analysis: application of statistical methods to interpret results.

Data quality control was performed by checking the accuracy of the automatic parses and excluding inaccurate data.

Journal of Ecohumanism
March 2024
Volume: 3, No: 7, pp. 2806 – 2814
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i7.4418

Thus, the materials and methods applied in this study allowed us to systematically analyze the linguistic structures and features of the Kazakh language using an extensive corpus of texts.

Literary Review

Linguistic research based on the analysis of textual data has become particularly relevant in modern science. One of the key tools in this area is the use of national corpus of texts. National corpus represents valuable resources (Yule, 2014), which open up many opportunities for researchers to study linguistic phenomena.

Gaston Gross's work "La Fréquence des mots" (Gross, 1952) raises the important question of word frequency in texts. This work formed the basis of many subsequent studies devoted to frequency analysis. The concept of "frequency vocabulary" became widespread in linguistics and has become the key to analyzing lexical variation.

With the development of computational methods and computer technologies, the possibilities of text analysis have expanded significantly. The works of John Singer "The Corpus Revolution in Language Technology" (Singer, 1996) and John Singer and Paul Reiber "Corpus: Pragmatics and Discourse" (Singer & Reiber, 1999) emphasize the role of corpus data in the research of linguistic pragmatics and discourse.

In the context of using text corpus to analyze syntactic structures, it is important to note the contribution of Charles Fillmore's The Case for Case Reopened (Fillmore, 1968). Fillmore introduced a new perspective on syntactic analysis by focusing on semantics and syntactic relations between words.

About the Kazakh language, the works of such linguists as Kaldybai Bektaev's "Mathematical Methods in Linguistics" (2 parts, co-authored 1973-1974) (Bektaev & Piotrovsky, 1973), Askar Zhubanov "Frequency Dictionary of M. O. Auezov's novel "The Way of Abai" (co-authored 1979) (Zhubanov & Bektaev et al., 1979), "Introduction to Computer Linguistics" (Zhubanov, 2007), Eleonora Suleymenova needs a national Corpus of the Kazakh language. (Suleimenova, 2015), Gulmira Madieva Almaty Kazakh language Corpus (2012), computer linguistics (Madiyeva, 2016), Almasbek Belbotayev ("Kumansha-Kazakh frequency dictionary" (co-authored 1978) and etc., Askar Zhubanov's "Corpus linguistics" (Zhubanov & Zhanabekova, 2016), focuses on the linguistic features and trends of the modern Kazakh language, including using corpus data.

Giuseppe Pizzirillo's study "The Use of Corpus in Linguistic Research" (2009) demonstrates the wide range of possibilities that text corpus provides for the analysis of semantics, vocabulary, syntax, and other linguistic aspects.

Thus, the presented literary review emphasizes the key role of national text corpus in linguistic research. The use of corpus data for frequency and syntactic analyses opens new perspectives on the study of language structures and variation, which is becoming an important direction in modern linguistics.

Results

Frequency analysis

Frequency analysis of the national corpus of the Kazakh language revealed the most commonly used words and phrases in various genres and thematic categories. Examples of keywords and phrases that turned out to be the most frequent are presented below:

Words are markers of time: "кез келген", "азамат жыл", "алғаш рет".

The family's words: "ата", "Эже", "бала".

Numerals: "бір", "екі", "үш".

March 2024 Volume: 3, No: 7, pp. 2806 – 2814

ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)

https://ecohumanism.co.uk/joe/ecohumanism DOI: https://doi.org/10.62754/joe.v3i7.4418

Linking words: "мен", "және", "бірақ".

Pronouns: "менің", "сенің", "олардың".

The frequency analysis of these words shows that the Kazakh language actively uses the mentioned word categories to express time frames, kinship relations, numerical values, and to construct coherent and logical texts. This highlights the importance of temporal markers and pronouns in everyday communication and literary texts.

Syntactic Analysis

Syntactic analysis revealed typical syntactic structures characteristic of the Kazakh language. Examples of such structures are presented below:

Subject-predicate: Мен қазақша оқимын.

Direct addition: Мен кітап оқимын.

Circumstance of Place: Мектепке барамын.

The circumstance of time: Кеше келдім.

These examples demonstrate the structural features of the Kazakh language, such as word order and the use of cases. The analysis shows that typical syntactic structures include subject-predicate patterns and the use of temporal and locative adverbials, which align with the sentence construction norms in the Kazakh language.

Thus, the results of frequency and syntactic analysis allow us to identify the key linguistic features of the Kazakh language, including the use of frequency words, typical syntactic structures, and grammatical constructions. These results contribute to a deeper understanding of the structure and features of the Kazakh language.

Variability analysis

The analysis of the National Corpus of the Kazakh language has also revealed the variability in the use of language elements in different contexts. Examples of variations in the use of words and constructions are presented below:

Variants of phraseological phrases: "Θτε κθπ" vs "τωм κθπ", where "Θτε" and "τωм" represent different dialect variants.

Synonyms: "жеткіншек" vs "жас", representing different shades of meaning.

Grammatical variability: the use of different forms of verbs depending on the person, number, and tense.

This analysis shows that the Kazakh language is rich in variations, reflecting cultural and regional characteristics. For example, the use of different synonyms and idiomatic expressions highlights the richness and dynamism of the language, as well as its ability to adapt to various social and cultural contexts.

Practical Application

The results of the study have practical significance for teaching Kazakh, creating lexicographical resources, and developing teaching methods. Knowledge of the most commonly used words and syntactic structures allows us to more effectively form educational materials and develop courses for different levels of language proficiency.

March 2024

Volume: 3, No: 7, pp. 2806 – 2814 ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)

https://ecohumanism.co.uk/joe/ecohumanism

DOI: https://doi.org/10.62754/joe.v3i7.4418

In addition, the results of the analysis can be used in the field of machine translation and natural language processing. The knowledge of word frequency and typical syntactic structures is important for creating more accurate and qualitative language models.

Examples of variation and dynamics

Variation of lexical variants: " $\kappa\theta\kappa$ " vs "ac π aH", where " $\kappa\theta\kappa$ " is the more archaic version and "ac π aH" is the modern one.

Variations in verb forms: "жазамын" vs "жаздым", where the first form corresponds to the present tense and the second to the past tense.

Variability of appeals: "атам" vs "апам", used depending on dialect and region.

Variations in the expression of time: "таңда" vs "таңертең", where the first expression is an adverb and the second is a participle.

Synonyms with different stylistic shades: " κ apa" vs " κ θ p", where " κ apa" is more formal and " κ θ p" is more colloquial.

Variability in the use of pronouns: "мен" vs "мені", where the differences are related to the use in different speech situations.

These examples emphasize the diversity in the use of lexical units, grammatical forms, and constructions in the Kazakh language, which indicates the richness and dynamics of this language.

Variations in syntactic constructions: Қазақстанның тәуелсіздік жылдары" vs "Тәуелсіздік жылдарындағы Қазақстан" (Years of Kazakhstan's Independence / Kazakhstan in the years of independence), where the order of words in the sentence is changed to emphasize different aspects.

Semantic variations: "денсаулық" vs "саулық", where the first word is used in a more medical context and the second in everyday speech.

Variation in first and last names: "Дархан" vs "Дарқан", where differences may be related to regional peculiarities.

Variation in color naming: "Қызыл" vs "Қырмызы", where the first word is more commonly used and the second is less common.

Synonyms for emotion: "шат" vs "шаттық", where the first word can be used in more everyday situations and the second for deeper emotions.

Variations in forms of politeness: "ci3" vs "ci3re", where the second form expresses a more formal address.

These examples emphasize the richness of the Kazakh language and the diversity of its use in different situations and contexts. Variation in linguistic elements and constructions is a reflection of linguistic dynamics and cultural influences.

Variation in the denotation of time: "таңертең" vs "таң", where the first word is used for the earlier time and the second word for the day after midnight.

Variations in the use of prepositions: "бойы" vs "бойынша", which can be used in different contexts to indicate direction or mode of action.

March 2024

Volume: 3, No: 7, pp. 2806 – 2814 ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)

https://ecohumanism.co.uk/joe/ecohumanism DOI: https://doi.org/10.62754/joe.v3i7.4418

Grammatical alternatives: "жатамыз" vs "жатыптық", where the first form expresses an action in the present tense and the second in the past tense.

Variability in the formation of verb forms: "тарту" vs "аша тарту", where the difference is due to the prefix changing the meaning of the verb.

Lexical variations in colloquial speech: "жақсы" vs "мақұл", where the first word is used more often in colloquial speech, and the second – in more formal situations.

Variations in the expression of cause: "себебі" vs "θйткені", where each expression can be used to indicate causality, but in different styles.

Variation in the formation of compound words: "OTE" vs "aca", which can be part of different word combinations.

Variations in the forms of politeness: "ci3дepre" vs "ceндepre", where the first form is more formal and the second is closer to a commonplace address.

These examples continue to emphasize the diversity in the use of lexical and grammatical units in Kazakh and demonstrate how this language is affected by context and societal changes.

Variations in the denotation of the time of year: "күз" vs "алтын күз", where the second expression emphasizes a specific time of year.

Variation in the use of the phrase " θ тінемін ": "сізден θ тінемін" vs "сіздерден θ тінемін", where the second form expresses a more formal address.

Semantic alternatives for expressing surprise: "мәәә" vs "oxoñ", where each word conveys a different nuanced reaction.

Lexical alternatives for expressing gratitude: "paĶмeт" vs "κθπ paĶмeт", where the second expression conveys deep gratitude.

Variability in the formation of passive constructions: " θ Ay" vs "ажалы жету", where the second form has a more abstract meaning.

Variations in the formation of adverbial participle: "imin" vs "imy", where different forms can be used in various contexts.

Grammatical variation in plural formation: "қала" vs "қалалар", where the first form corresponds to the singular and the second to the plural.

Variations in expressing comparison: "coa бала" vs "aна бала", where different definitions can be used in many contexts.

Synonyms for expressing the state are: "жату" vs "тұру", where the differences are related to the duration of the action.

Variation in the use of interrogative particles: "ба" vs "бе", "па" vs "пе" which can be used in interrogative sentences.

These examples continue to illustrate the many variations and dynamics in the use of linguistic elements in the Kazakh language, which reflects its living nature and ability to adapt to a variety of communication situations.

Journal of Ecohumanism March 2024 Volume: 3, No: 7, pp. 2806 – 2814

ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online) https://ecohumanism.co.uk/joe/ecohumanism DOI: https://doi.org/10.62754/joe.v3i7.4418

Discussion

The results of this study, based on the analysis of the National Corpus of the Kazakh Language, have important implications for linguistic science and practical application. In the following, we consider several aspects that confirm the significance and contribution of this study.

Disclosure of linguistic dynamics: The variability revealed as a result of frequency and syntactic analysis reflects the living dynamics of the language. The study of the diversity in the use of words and constructions allows us to understand how language adapts under the influence of cultural, social, and communicative factors.

Linguistic richness and cultural heritage: Examples of variability and diversity demonstrate the richness of lexical and grammatical resources of the Kazakh language. This not only reflects the deep roots of this language but also emphasizes its ability to express a wide range of meanings and shades.

Adapting to modernity: Analysing frequency and syntactic structures in different genres of texts allows us to understand how language adapts to modern reality. Identifying keywords and constructions relevant to specific domains (e.g. technical, scientific, or social) allows a better understanding of how language reflects current trends and technological changes.

Potential for education and communication: The results of the study can be beneficial to the development of educational materials, and methods of teaching the Kazakh language, as well as for improving automatic translation and natural language processing. Knowledge about word frequencies and typical syntactic structures allows for more efficient formation of curricular and the creation of tools for learning and communicating in the language.

The results of this study, based on the analysis of the National Corpus of the Kazakh Language, offer interesting perspectives for understanding and interpreting language dynamics, lexical richness, and cultural features. In this section, we will elaborate on the identified linguistic aspects by providing case studies.

Variation and Cultural Adaptation: analyzing examples of word usage variation, we can observe how linguistic elements are adapted to different spheres and contexts. For example, the synonyms "жеткіншек" and "жас" for "young" reflect different nuances and cultural associations. Such variations emphasize the ability of language to express subtle shades of meaning associated with specific situations.

Preservation of linguistic heritage: Examples of variation in the use of appeals ("атам" vs "апам") and lexical alternatives ("қызыл" vs "қырмызы") emphasize the richness of the lexicon and traditional forms of communication. These variations reflect the historical and cultural heritage of the language and at the same time show how these traditions are combined with modern needs.

Language enrichment and expressiveness: Examples of variation in expressions of emotion ("шат" vs "шаттық") and grammatical alternatives ("жатамыз" vs "жатыптық") emphasize the expressiveness of the Kazakh language. These variations allow speakers to convey their thoughts, feelings, and attitudes more accurately.

Adaptation to modernity and technology: Examples of variation in the use of words and constructions in different spheres ("бойы" vs "бойынша") reflect how language adapts to modern realities. This is particularly relevant in the context of technological development and Internet communication, where language often adapts to new ways of expressing thoughts.

Variability in the formation of the season: Examples of variability in the designation of the season ("күз" vs "алтын күз") show how a language can adapt to the seasons and natural changes. This highlights the importance of the seasonal cycle in culture and everyday life.

Journal of Ecohumanism
March 2024
Volume: 3, No: 7, pp. 2806 – 2814
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i7.4418

Preservation of cultural traditions through linguistic variations: Examples of variability in the expression of gratitude ("paκmet" vs "κθπ paκmet") and appeals ("aтam" vs "aπam") indicate how the language preserves cultural customs and the level of politeness. This helps to maintain the unique features of national identity.

Dynamics in modern communication: Examples of variability in syntactic structures and expressions of time ("таңертең" vs "таң") reflect changes in the language under the influence of modern communication needs. The variations show how a language can adapt to more concise and dynamic forms of communication.

Potential for education and intercultural communication: Examples of variability and synonymy can be used to study cultural nuances and immersion in the language environment. It also highlights the importance of language education to understand the diversity of cultures and cross international borders.

The results of this study, based on the analysis of the National Corpus of the Kazakh language, reveal intriguing perspectives for understanding and interpreting linguistic dynamics, lexical richness, and cultural characteristics. The variability and abundance of linguistic resources confirm the vibrant nature of the Kazakh language and its ability to adapt to various communicative situations.

Conclusions

The study of the National corpus of the Kazakh language allows a deeper study of the linguistic features and structures of this language. The analysis of the frequency of words, syntactic structures, and variability allows us not only to better understand linguistic phenomena but also to use the acquired knowledge in practical applications such as learning and machine translation. Thus, this study reveals the richness and potential of the National Corpus of the Kazakh language as a valuable tool for linguistic research and practical application.

The study of the National Corpus of the Kazakh language opens up new horizons for learning this rich and dynamic language. By analyzing the frequency, syntax, and variability, it was possible to identify the key features of the Kazakh language in different contexts. This study emphasizes the importance of preserving and studying the linguistic heritage and adapting it to modern needs and technological challenges.

A study of the National Corpus of the Kazakh language confirms its richness, dynamism, and ability to adapt to a changing world. The examples of variation and diversity presented in this work provide a deeper understanding of linguistic features and their role in social and cultural life. This study invites further study and preservation of the linguistic heritage, and its creative use in a contemporary context.

The research conducted based on the analysis of the National Corpus of the Kazakh language is a valuable resource for the study of language dynamics, the richness of the lexicon, and cultural characteristics. The examples of variability given in the work reveal the versatility of the Kazakh language and its ability to adapt to changing conditions.

Further research in this area can not only deepen our understanding of language processes but also contribute to the preservation and enhancement of the rich cultural heritage. Examples of the variability and dynamics of the Kazakh language analyzed in the framework of the study emphasize its ability to adapt, expressiveness, and preservation of cultural values.

This study provides an interesting perspective on the evolution of language and its importance to society. The results obtained can be used to further study the dynamics of language and the development of educational and communicative practices.

The research conducted based on the analysis of the National corpus of the Kazakh language is an important contribution to the understanding of linguistic dynamics and the richness of the language. Through analyses of frequency, syntax, and variation, key features of the Kazakh language in different

Journal of Ecohumanism
March 2024
Volume: 3, No: 7, pp. 2806 – 2814
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i7.4418

contexts and genres of texts were identified. The results of the study allowed us to see how the language adapts to contemporary challenges, preserves cultural traditions, and is enriched through a variety of lexical and grammatical variations. The examples of variability given in the work emphasized the expressiveness and flexibility of the Kazakh language, and its ability to accurately convey subtle meanings and shades. This study also marked an important step in understanding the impact of modern technology on language change. The analysis of syntax and lexical variability in various fields, including education and intercultural communication, emphasized the importance of language learning and adaptation for successful communication in the modern world. Knowledge of how a language adapts and develops is essential for education, cultural exchange, and international relations. The study of language dynamics enriches our understanding of the diversity of cultural expressions and contributes to a better understanding of different linguistic and cultural communities.

Acknowledgments

«This research has been/was/is funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP15473441)».

References

Bektaev K. B., Piotrovsky R. G. Mathematical Methods in Linguistics: (Textbook) / Ministry of Higher and Secondary Special Education of the Kazakh SSR. Kazakh State University named after S.M. Kirov. Chimkent Pedagogical Institute. – Alma-Ata: [n.p.], 1973. – 281 p. [Bektaev K. B., Piotrovskij R. G. Matematicheskie metody v jazykoznanii (Mathematical methods in linguistics): (Ucheb. posobie) / M-vo vyssh. i sred. spec. obrazovanija KazSSR. Kaz. gos. un-t im. S. M. Kirova. Chimkent. ped. in-t. – Alma-Ata: [g. i.], 1973. – 281 s. [in Rus]

Chomsky N. Syntactic Structures. Mouton de Gruyter. – 1957. – 68 p.

Fillmore Ch. The Case for Case Reopened. Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1-88. 1968.

Gross G. "La Fréquence des mots" (The Frequency of Words). Fayard. 1952. – 245 p. [in Fra]

 $Labov\ W.\ The\ Social\ Stratification\ of\ English\ in\ New\ York\ City.\ Center\ for\ Applied\ Linguistics.-1966.-655\ p.$

Madiyeva G.B. Computer Linguistics. A Textbook. – Almaty: Qazaq University, 2016. – 240 p. [Madieva G.B. Kömpüterlik lingvistika (Computer linguistics). Oqu qūraly. – Almaty: Qazaq universitetı, 2016. – 240 b. [in Kaz]

Singer J. The Corpus Revolution in Language Technology. Perseus Books Group. - 1996. - 115 p.

Singer J., Reiber P. Corpora: Pragmatics and Discourse. Perseus Books Group. - 1999. - 514 p.

Smirnova T.N. Linguistic Analysis of the Frequency of Words in the Corpus of Texts of Different Genres. Vestnik of Moscow University. Series 9: Philology, (6), 2010. – pp. 23–32. [Smirnova T.N. Linguisticheskij analiz chastotnosti slov v korpusah tekstov raznyh zhanrov (Linguistic analysis of the frequency of words in the corpus of texts of different genres). Vestnik Moskovskogo universiteta. Serija 9: Filologija, (6), 2010. – S. 23–32. [in Rus]

Suleimenova E.D. Does the Kazakh Language Need a National Corpus? Journal "Tiltanym" of the Institute of Linguistics named after A. Baitursynov, 2015. – №2. – pp. 42–45. [Suleimenova E.D. Qazaq tılınıñ ülttyq korpusy kerek pe? (Needs a national corpus of the Kazakh language?) A.Baitūrsynūly atyndağy Tıl bılımı institutynyñ «Tıltanym» jurnaly, 2015. – №2. – 42–45 bb. [in Kaz]

Yule G. The Study of Language. Cambridge University Press. - 2014. - 340 p.

Zhubanov A.K. Introduction to Computer Linguistics: A Manual. – Almaty: Qazaq University, 2007. – 204 p. [Jūbanov A.Q. Kömpüterlık lingvistikağa kırıspe (Introduction to computer linguistics: a manual): Oqu qūraly. – Almaty: Qazaq universiteti, 2007. – 204 bet. [in Kaz]

Zhubanov A.K. Corpus Linguistics. Almaty: "Qazaq tılı" Publishing House, 2017. – 336 p. [Jūbanov A.Q. Korpustyq lingvistika (Corpus linguistics). Almaty: «Qazaq tılı» baspasy, 2017. – 336 b. [in Kaz]

Zhubanov A.K., Bektaev K.B., Myrzabekov S., Belbotaev A.B. Frequency Dictionary of the Novel "The Way of Abai" by M.Auezov. – Almaty: Ğylym, 1979. – 336 p. [Jūbanov A.Q., Bektaev Q.B., Myrzabekov S., Belbotaev A.B. M.Äuezovtiñ «Abai joly» romanynyñ jillik sözdigi (Frequency Dictionary of the novel "The Way of Abai" by M.Auezov). – Almaty: Ğylym, 1979. – 336 b. [in Kaz]

Zhubanov A.K., Zhanabekova A.A. Corpus Linguistics. A Textbook. – Almaty: Qazaq tılı Publishing House, 2016. – 336 p. [Jūbanov A.Q., Jañabekova A.Ä. Korpustyq lingvistika (Corpus Linguistics) / Oqu qūraly. – Almaty: Qazaq tılı baspasy, 2016. – 336-b. [in Kaz]