

# PROPER NAMES KNOWLEDGE BASE FOR INTELLIGENT MOBILE APPLICATION

GAZIZA YELIBAYEVA<sup>1</sup>, LAURA ORYNBAY<sup>2</sup>, GULMIRA BEKMANOVA<sup>3</sup>, AYAULYM SAIRANBEKOVA<sup>4</sup>

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

E-mail: <sup>1</sup>gaziza362@gmail.com, <sup>2</sup>laura.aktobe.kz@gmail.com, <sup>3</sup>gulmira-r@yandex.kz, <sup>4</sup>sairanbekova98@gmail.com

## ABSTRACT

This article proposes a semantic knowledge base of proper names of the Kazakh language with about 100 semantic features. A semantic knowledge base of proper names of the Kazakh language allows obtaining new scientific results in expanding the functions of the Kazakh language, developing information resources in digital format and scientific-linguistic foundations of the Kazakh language for intelligent information systems with different capabilities. The analysis showed that Kazakh names have a large number of semantic features from the origin of names to their emotional coloring, borrowing from other languages, combining different roots, and so on, which are included as properties of the semantic knowledge base. The results of the work will be used to develop software application “Fascinating Onomastics” that will allow anyone who wants to do the analysis of names by origins, distributions, and variations as well as choose a name according to the given parameters. Moreover, this research contributes to the extraction and classification of proper names in many applications in the field of natural language processing and text analysis. By extracting proper names from the text, we provide a POS tagging information with semantic features. The study is carried out within the framework of the project BR11765535 “Development of Scientific and Linguistic Foundations and IT Resources to Expand the Functions and Improve the culture of the Kazakh Language” and upon completion of the project, the mobile application will be published on the Google Market and the Apple Store.

**Keywords:** *Fascinating Onomastic, Semantic Features, Knowledge Database, Logical Rules, Artificial Intelligence.*

## 1. INTRODUCTION

The purpose of the article is to study the regularities of the language in order to create an application on onomastics, to highlight the semantic features of words for an application on onomastics. For the first time, a semantic knowledge base of proper names of the Kazakh language with about 100 semantic features has been created. Semantic features allow you to make various semantic queries on names in the application. The semantic knowledge base of proper names of the Kazakh language allows obtaining new scientific results in expanding the functions of the Kazakh language, developing information resources in digital format and scientific and linguistic foundations of the Kazakh language for intelligent information systems with different capabilities. The analysis showed that Kazakh names have a large number of semantic features from the origin of names to their emotional coloring, borrowing from other languages,

combinations of different roots, and so on, which are included as properties of the semantic knowledge base. In the future, these properties will allow anyone who wants to analyze names, choose a name according to the given parameters. This makes the “Fascinating Onomastics” app unique in its kind. The study is carried out within the framework of the project BR11765535 “Development of scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language” and upon completion of the project, the mobile application will be published on the Google Market and the Apple Store.

In the Republic of Kazakhstan, there is a need for digital linguistic resources that contribute to the development of the modern Kazakh language, its use on the Internet, the dissemination of digital services, and the development of its culture. Therefore, in modern society, there is an understanding of the need for a comprehensive study and a comprehensive solution to this interdisciplinary problem.

The main idea is to solve the problems posed by the formal and semantic description of the Kazakh language using artificial intelligence and linguistic methods, as well as to develop predicate knowledge bases and implement software of the models and methods obtained. Any language can be extended, developed and turned into a universal tool, being used for various purposes. However, the language of the nation may not be able to fulfill its purpose, its scope may be limited, and it may not be able to express complex thoughts. Therefore, every citizen must speak his/her native language so that the native language does not become unnecessary, the scope of the state language is expanded, it becomes a means of communication in the environment of his/her residence, and is freely used in foreign diplomatic relations and proceedings with other countries, and this requires a lot of support from the state.

In addition to synonymizers that need to be developed when creating software applications in our country, onomastic spaces and digital terminological dictionaries should not be neglected. The modern onomastic space of Kazakhstan is distinguished by innovative, socio-psychological and historical content. The study of personal names in the Kazakh language has its own characteristics. They consist in studying the lexico-semantic and structural forms of personal names, and establishing the reasons for their formation.

The general theoretical problems of the Kazakh onomastics are mainly examined in the works of such researchers of the Kazakh onomastics as T. Zh. Zhanuzakov, Ye. A. Kerimbayev, G. B. Madiyeva. Since 1949, more than 100 doctoral and candidate dissertations on onomastics have been defended in Kazakhstan [1]. However, there are reasons to say that there are no Kazakhstani software applications in onomastics yet.

The mobile application "Fascinating Onomastics" can be used for educational purposes by a wide range of the population, which, in turn, provides the necessary social efficiency and further improves the culture of the Kazakh language, and also has real practical significance in terms of application.

The Internet has not found analogues of special linguistic resources related to Kazakhstani (Kazakh) onomastics. Therefore, the relevance of this study is determined by the fact that it allows to systematically solve the complex problem of developing scientific and linguistic foundations and IT resources of the Kazakh onomastics and raise the status of the Kazakh language to the level of languages of developed countries. The second innovation of the research theme is the fact that for

the first time the structure of the semantic base of Kazakh proper names was determined and the semantic base of Kazakh proper names was developed according to about 100 possible features (properties).

## 2. REVIEW OF LITERATURE

### 2.1 Onomastics of Kazakh Names

In modern linguistics, research work on onomastics is of particular interest. Linguistics is constantly developing; accordingly, new directions appear based on the integration of various branches of science. In this regard, in the field of onomastics, in particular in anthroponymy as the science of people's names, new studies appear due to changes in society, which are reflected in anthroponyms that store certain information about political and social events, cultural relations in society [2-3].

According to Abduali Kaidar, the tradition of naming is important in human life: "In ancient times, when humanity began to live in society, there was one of the traditions of civilization that arose due to the need for people to interact, communicate with each other, this is "giving each other names, nicknames. "Without this, people would not be able to live together and in an organized way, learn the secrets of nature and resist its mysterious power" [4].

In Kazakhstani onomastics, especially in research works on anthroponymy, the historical ways of development of anthroponyms are revealed, their semantic, structural, ethnolinguistic features are considered in comparative aspects, their motivational nature is determined.

According to researchers [5], articles about Kazakh names began to appear in the second half of the 19th century. These first articles include "Imena kirgizov [Kyrgyz names]" by A.E. Alektorov [6], as well as "K voprosu o narechii imen u kirgizov [On the issue of the dialect of names among the Kyrgyz]" by A. Divaev [7]. Although both are small in size and scope, they are important as a first step.

At the end of the 19th century, N.F. Katanov made an attempt to lexicograph the proper names of Turkic folklore, which were later reflected in the work "Alfavitnyi ukazatel sobstvennykh imen, vstrechayuschichsya v pervom tome Obrazcov narodnoy literatury tyurkskix plemen, sobrannykh V. V. Radlovym [Alphabetical index of proper names found in the first volume of Samples of folk literature of Turkic tribes collected by V.V. Radlov]" [8]. The Index provides an ethno-linguistic description of personal names, as well as the functions of heroes in folklore, lists traditional epithets containing an additional part of the

semantics of the name. The scrupulous work of Academician N.F. Katanov is an invaluable study in the field of the ethno-cultural space of the Turkic peoples and has not lost its significance to this day.

Among the Kazakh researchers, Sh. Ualikhanov was the first to pay attention and analyze the names of people. In the article “Qazaq shezhiresi [Kazakh genealogy]” [9] he made a scientific analysis of the history, culture, economy, way of life, customs of the Kazakh people, as well as the history of the Khanate and the origin of the Uly zhuz, Orta zhuz and Kishi zhuz tribes, on the basis of which the names of many clans and Khans were given.

The provisions considered by the Kazakh linguist Kh. Zhubanov are extremely important for the formation of onomastics in Kazakhstan. In the article “Iz istorii poryadka slov v kazakhskom yazyke [From the history of the order of words in the Kazakh language]” [10], he paid special attention to the syntactic construction of human names, consisting of combined words, and made a comprehensive analysis. Here he pays attention to the legality of phrases in the modern Kazakh language, which is mainly the determinant before the noun, but notes that this regularity is not observed in all phrases in the ancient Kazakh language: “Personal names Kunsulu, Aisulu, Tansulu, Tanzharyk cannot be syntactically deciphered within the limits of modern norms, that is, as a combination of a definition with a definable, preceding the first to the second, because nouns are no longer used as definitions when an adjective should be defined”.

Onomastists-scientists studying Kazakh anthroponymy consider this field from various angles and comprehensively. Researches in this field were carried out by professor, onomastist and scientist Telkozha Zhanuzakov, who was the first comprehensive researcher of personal names in the Kazakh language, identified the historical development paths of Kazakh anthroponyms and systematized their grammatical and semantic characteristics.

T. Zhanuzakov in his research work “Qazaq esimderining tarihy [History of Kazakh names]” [11] distinguishes human names according to their structural features as follows:

- 1) individual personality;
- 2) derivative personality;
- 3) double composition;
- 4) multi-component or complex personality.

T. Zhanuzakov's work “Qazaq tilindegi zhalky esimder [Personal names in the Kazakh language]” [12] is dedicated to a comprehensive description of personal names in the Kazakh language: the

structure, origin, history of appearance and peculiarities of personal names, processes of transition to common names, grammatical laws and connections with other sciences are studied. T. In this work, Zhanuzakov groups personal names according to their lexical meaning and identifies 30 groups:

- 1) names related to animal husbandry (including names related to camels, horses, sheep, goats, cattle);
- 2) names related to field animals;
- 3) names related to bird names;
- 4) names given with the meaning of growing like a flower or a plant;
- 5) names related to fish names;
- 6) names related to the names of insects;
- 7) names related to the stars in the sky;
- 8) names related to the beginning of the year;
- 9) the names given to be the best, famous hero, prestigious person;
- 10) names of heroes' spouses are given to girls;
- 11) name of a famous saint, prophet;
- 12) names given to have good character;
- 13) names given in the sense that the child should be beautiful, handsome, attractive;
- 14) names given to make the child white or cute;
- 15) names meaning that the child will have a long life;
- 16) names that mean that the child will be full and happy;
- 17) names used to remember the place of birth and place of birth of the child;
- 18) names given in the sense that the child should be smart and reasonable;
- 19) names given in the sense that the flesh of the child should be alive and ripe;
- 20) names given to initiate the occurrence of a certain phenomenon;
- 21) names related to how the child was born when waiting for a child, longing for it;
- 22) names related to different races;
- 23) names used in economic or military professional sense;
- 24) names indicating the social class of people;
- 25) names related to folk names in the international sense;
- 26) names related to relative meaning;
- 27) names meaning the order of birth season;
- 28) names meaning to be rich;
- 29) names with the meaning that the child will be stable and safe from death;
- 30) names given after the October Revolution.

## 2.2 Semantic Database of Proper Names

In works related to linguistics, a significant number of applications use a resource consisting of a list of words provided with morphological, syntactic and semantic information, which is known as an electronic dictionary. Proper names are often ignored when building an electronic dictionary [13]. However, proper names make up a significant part of many texts, and therefore the creation of a semantic base is an important task.

In 1985, J. Miller and his colleagues from the Laboratory of Cognitive Science at Princeton University (USA) developed a model of the human mental lexicon. The resource was named WordNet and eventually became one of the most authoritative and widely used standards for building lexical and semantic databases [14].

The popularity and wide distribution of WordNet are primarily due to its significant content and structural characteristics. The Princeton WordNet and all subsequent variants for other languages are aimed at displaying the composition and structure of the lexical system of the language as a whole, rather than individual thematic areas [15].

For example, the WordNet version in 2003 covers the common vocabulary of modern English – more than 150 thousand words. The dictionary consists of 4 separate files for the main significant parts of speech: nouns, verbs, adjectives and adverbs. The basic structural unit of the Princeton WordNet is a synonym series (synset) that combines words with a similar meaning. It is assumed that each synset represents in the dictionary some lexicalized concept of a given language. For the convenience of using the dictionary by a person, each synset is supplemented with a definition and examples of the use of words in context. Synsets in WordNet are interconnected by such semantic relationships as hyponymy (generic), meronymy (part-whole), lexical inference (causation, presupposition), etc.; among them, hyponymy plays a special role: it allows you to organize synsets into hierarchical structures (trees). The vocabulary of each part of speech is represented as a set of trees (forests). For different parts of speech, generic relations may have additional characteristics and differ in scope, for example, only some groups of adjectives in WordNet are connected by hyponymic relations [16].

The original version of the dictionary does not contain information about the thematic organization of vocabulary, since, according to J. Miller, it represents a different type of vocabulary structuring, different from the semantic one presented in WordNet. The basic principles of WordNet have not undergone significant changes

since its creation, but work continues to improve it, for example, in the version of WordNet 1.7.1, information about the frequency of values has been introduced. The widespread use of WordNet is largely due to the fact that it is distributed freely [17].

One of the foreign works devoted to the description of names in the article “Building The Classical Arabic Named Entity Recognition Corpus (Canercorpus” [18]. In this article, the authors present the classical corpus of recognition of objects with Arabic names as a new corpus of labeled data, which can be useful for solving problems of recognition of objects with Arabic names. And also, based on the Islamic theme, named objects are classified into 20 types, which include objects of a specific area that have not been previously considered, such as Allah, the Prophet, Heaven, Hell and Religion.

Another work related to our topic is discussed in the article “Multilingual Ontology of Proper Names” [19]. In this article, the authors showed that proper names can have a very complex morphological structure in languages with rich morphology, but also that their semantic description does not depend on the language. They also showed that even the extensive morphological electronic dictionaries they used did not provide exhaustive, accurate and comparable descriptions of proper names in different languages. Therefore, they developed a formalism for describing proper names in multilingual environments, implemented in the Prolex project. This formalism allows, as in WordNet, to coordinate the semantic description of proper names in different languages, while specific morphological characteristics are given for each language.

Another work published in 2018 with already familiar authors is the article “Semantic and syntactic patterns of verbose names: an interlanguage study” [20]. In this article, they explore proper names that belong to three main classes: personal names, names of places and organizations, and which can be both one-word nouns and multi-word expressions. To do this, the authors first define common (language-independent) semantic models of proper names, and then provide appropriate syntactic models in English, Bulgarian, French, Greek and Serbian. After that, the authors compare these patterns by grammatical categories of dependent components, definiteness, distribution of clitics, word order and various alternations. The ultimate goal of the authors is to create a universal structure for recognizing named objects.

Kazakh scientists also carried out work on compiling a knowledge base on the Kazakh language [21-24].

But they did not specifically deal with the proper names.

Thus, it is obvious that human names in the Kazakh language have become the subject of some research. But the creation of a knowledge base and the creation of a mobile application based on these studies have not yet become the subject of a separate scientific work [25-26].

### 3. DEVELOPMENT OF THE MOBILE APPLICATION “FASCINATING ONOMASTICS”

One of the most important parts of the lexical fund of any language is proper names. Without them, a person's relationship is impossible. To distinguish between beings in the environment that are identical in nature from each other, it is necessary to come up with special words that perform the function of distinguishing.

Onomastics is a branch of linguistics that studies the corresponding names that serve to individually designate any living or inanimate object or phenomenon. The object of onomastics is the proper names of all categories that make up their own subsystem of the general lexical fund of the language [17, 27]. Proper names are universal in onomastic terms, because without them there would be no language or culture [11].

The research group “Fascinating Onomastics” chose anthroponymy, one of the largest branches of onomastics, as the direction of research. Kazakh anthroponymy is a set of anthroponyms in the Kazakh language, that is, proper names used to name a person in the Kazakh language. Kazakh anthroponymy is distinguished by its diversity and includes both native Kazakh and non-native borrowed names from other languages (mainly from Arabic or Persian). The dominant group of Kazakh names clearly reflects the cultural and historical life of our people between the past and the present [5].

#### 3.1 Justification of the Research Direction

One of the topics that is of interest to any person is to know the history of the origin of his/her name or, if we name it in scientific language, anthroponymy. Anthroponym is the most obvious indicator of human status. All types of anthroponyms are united by common functional features – the ability to use a person when naming and referring to him/her. Among them, only personal names have a high degree of denotational individualization, that is, every person has it.

In this onomastic study, the following methods are used: the descriptive method and the historical

method [28]. The descriptive method is a sequential description of personal names and various onomastic phenomena studied in a particular historical period. This method involves compliance with a number of requirements: a clear understanding of the chosen research theme (anthroponyms, a system of toponyms, etc.), a sequence of description, systematization, grouping or classification, a description of the material in accordance with the research task (qualitative, quantitative). The initial stage of describing the material is its collection, cataloging, reasonable systematization, which allows seeing its various parts (types, variations), general relationships between them, as well as their most important properties. The historical method involves identification of onomastic layers by linguistic affiliation, that is, the origin of the word – language is determined. Furthermore, the method of classification (systematization, grouping of phenomena according to certain features) was taken as the basis of the research. The classification method is a way of dividing a set of systematization (classification) objects into subsets (classification groups).

For conducting research, a research methodology was chosen based on secondary sources. The study of secondary sources is a thoughtful conclusion, a review of the literature and an in-depth profound analysis. Theoretical and experimental research process is to study the examples and semantic structures of Kazakh proper names, as well as to determine the designation of Kazakh proper names and reflect their signs and properties on the basis of semantic knowledge.

### 4. RESULTS

In the course of the semantic analysis of Kazakh proper names, the authors considered the classification of anthroponyms from different points of view, revealing the origin and meaning of the names, in accordance with which the properties of words were determined, and the structures of the semantic base were developed. The semantic base developed in the course of the research work is indicated in Tables 1-11. About 100 possible properties of Kazakh names have been identified in this database [29].

The system of semantic designation of names from different languages of origin and names in the ancient Kazakh language, integrated into the Kazakh language, was also considered.

Among the names in the Kazakh language, there are combined words consisting of several roots, such



names were revealed and supplemented by the authors in full on the basis of developed semantic knowledge. It is possible that each of their roots came from different languages. For example, the name Aimira is formed by two words “ai” – a celestial body in Kazakh and “mir” – in Russian, meaning peace, tranquility. As other examples of names that are a compound word, the following can be indicated: Ainazhan, Akbota, Anargul, Bibigul, Abylgazy, Abdolla, etc. [30]

From the lexical point of view and studying the history of the origin of names, the following properties were identified:

- 1) Names associated with history, nationality;
  - a) Names of clans and tribes (ethno anthroponym): Adai, Naiman, etc.;
  - b) Names associated with a historical event: Zhanis, Sailau, etc.;
  - c) Religious names: Abyz, Azhar, Aisha, etc.
  - d) Names related to traditions: Aidar, Kekilbay, Tulymbek, etc.
- 2) Names related to literary works:
  - a) Names of fairy-tale characters (myths): Aldar, Alpamys, etc.;
  - b) Literary names: Birzhan, Aigerim, Balkadisha, Bayan, etc.;
- 3) Names related to nature:
  - a) Names related to land and water: Altai, Bagdat, Yedil, etc.;
  - b) Names related to nature, natural phenomena: Aidariya, Aiziya, Ainur, Aisaule, etc.;
  - c) Names related to celestial bodies (the Moon, stars): Aibas, Altynai, Bazaraiym, Bader, etc.;
  - d) Names related to the weather: Boran, Dauylbai, etc.
- 4) Names associated with valuable things:
  - a) Names related to precious stones: Almas, Marzhan, etc.;
  - b) Names related to jewelry: Aizere, Aina, Bates, etc.;
  - c) Names related to metal (iron): Bolat, Aikumis, Altyn, etc.;
  - d) Names related to currencies: Dinar, Dinara, etc.
- 5) Names associated with animals and birds:
  - a) Names related to four-legged animal and their cubs: Akbota, Akkozy, Donenbai, Zhanbota, etc.;
  - b) Names related to predatory animals: Arystan, Kaskyrbai, Baibori, etc.;
  - c) Names related to an animal: Bugybai, Akmaral, etc.;
  - d) Names related to a bird: Burkit, Bulbul, Lashyn, etc.;
- 6) Names associated with the name of the plant: Kyzgaldak, Raushan, etc.
- 7) Names associated with fruit names: Alma, Anar, etc.
- 8) Names related to various food names: Alua, Asel, Sheker, Shyrin, etc.
- 9) Names related to house furniture, property: Ainagul, Tabakbai, etc.
- 10) Names related to colors: Akbota, Sary, etc
- 11) Names related to body part, human body: Akkal, Altynshash, Botagoz, etc.
- 12) Names related to the family can be defined as follows:
  - a) Male names given to a family with no boys: Tokhtar, Turar, Otegen, Tulemis, etc.;
  - b) Names related to the old age of parents: Alpys, Zhetpisbai, Kyrykgul;
  - c) Names that are given to a family to have a lot of children: Baizhan, etc.;
  - d) Names that are given to a family to have a lot of boys: Baikoshkar, Koshkarbay, etc.;
  - e) Names for the firstborn child in the family: Abubakir, Auel, etc.;
  - f) Names for children born one after another: Yergesh, Zhalgas, etc.;
  - g) Names given to the youngest child in the family: Kenzhebek, Balsheker, etc.;
  - h) Names given children born with longing: Ansagan, Ansar, etc.;
  - i) Names given to the only child in the family: Birzhan, Balkadisha, etc.;
  - j) Names for twins: Yegizbay, Koszhan, etc.;
  - k) Names given according to the day, month, year of birth of the child: Ashir, Beisenbi, Adita, etc.;
  - l) Names showing family relations: Abira, Zhiyenbai, Tuganbai, etc.

The use of words associated with various desires and goals in order to name children is also common [10]. Wishing parents to have children with best qualities: to be a brave child (Aibar, Aitemir, Aibyn, Kaisar, Zhiger, etc.), to be intelligent and smart (Akylbek, Abutalip, Agila, Beddana, etc.), to be beautiful (Azhar, Aibike, Ademi, etc.), to live a long life (Amantur, Abylkaiym, Aisha, Amangul, Zhanuzak, etc.), to be respectful (Ashim, Ardak, Aziza, Galiya, etc), rich (Baiakhmet, Aida, Alisha, etc.), to have many friends (Baidos, Anisa, etc.), to be a boss (Bizhan, Degdar, Amira, etc.), to be honest and fair (Agnessa, Arna, Adila, etc.), to be happy

(Bagymsha, Baktygul, Bakytbek, etc.), to be generous (Akzhibek, Anna, Arshat, etc.), to be kind (Meyirim, Amina, Meyirgul, etc.), to be hard-working (Amila, Badiga, etc.), to be a child of the people (Yelaman, Yeldar, Yelzhas, etc.), to be artistic (Bulbul, Onerbai, etc.).

The results obtained can be generalized using the structure of this semantic database to describe the semantic structures of concepts and examples in other areas of proper names (toponymics, zoonymics, cosmonymics, ethnonymics, phytonymics, etc.). The developed database of semantic knowledge makes it possible to unify the designation of concepts of other areas of proper names.

During this period, all the tasks set were fully solved. The database of semantic knowledge developed in the section of anthroponymy of Kazakh proper names includes all the semantic properties of Kazakh male and female names, but may require additions and refinement if new names appear.

In the course of the results obtained at this stage of the research, the structure of the semantic base of Kazakh proper names was described. This base is classified according to the following properties of names: names related to the origin (language), history, nationality, literary works, nature, valuable things, animals and birds, plant names, fruit names, names related to various foods, home furnishings, property, color names, body part, human body, family, etc.

Based on the study of examples and structures of Kazakh proper names, determining the designation of all concepts of anthroponyms and reflecting their properties in the semantic base, determining the relationship between Kazakh names and names in other languages and choosing a way to designate them in the semantic base, the following results were obtained:

- there was developed the structure of the designation of proper names in the Kazakh language in the semantic base with about 100 properties;
- there was developed semantic knowledge base of proper names in the Kazakh language.

The performed research has a high scientific and technical level and corresponds to the best achievements in the field of the Kazakh onomastics, since a sample of such a large base of semantic knowledge has not been found on the Internet. For example, in the Kazakhstan mobile applications “Kazaksha Yesimder” (Kazakh Names) [31], “Kazaksha Yesimder Magynasy” (Meaning of Kazakh Names) [32], “Kazakhskiye Imena” (Kazakh Names) [33] there are only two or three categories (gender, letters), while the foreign

program “Namsor” [34] classifies personal names by gender, country of origin or ethnic belonging, while the “List of Names Dataset” [35] database classifies each name only by gender and letters.

#### 4. CONCLUSION

A database of proper names, examples and a prototype of the user interface in the Kazakh language were created. The article highlights the semantic features of words for applications in onomastics. For the first time, a semantic knowledge base of proper names of the Kazakh language has been created, containing about 100 semantic features. Semantic features allow you to make various semantic queries on names in an application.

The results obtained can be used separately and the application will be posted on Google Play for free use and all sorts of semantic analysis of Kazakh names. Also, the obtained knowledge base and algorithms for processing the semantics of proper names can be used within any linguistic corpus for any analysis of text containing proper names.

#### REFERENCES:

- [1] G.B. Madiyeva, “Onomastika Kazakhstana: analiticheskiy obzor (podkhody issledovaniya onimov) [Onomastics of Kazakhstan: an analytical review (onym research approaches)]”, Kazak universiteti (Almaty), 2018. 80 p.
- [2] Utley, F. L. (1963). The linguistic component of onomastics. *Names*, 11(3), 145-176.
- [3] Algeo, J., & Algeo, K. (2000). Onomastics as an interdisciplinary study. *Names*, 48(3-4), 265-274.
- [4] A. Kaidar, “Kazaktar ana tili aleminde (ethnolinguistics sozdik) [Kazakhs in the world of mother tongue (ethnolinguistic dictionary)]”, Sardar (Almaty), 2013, 728 p.
- [5] T.Zh. Zhanuzak, “Kazak onomastikasy – Kazakhskaya onomastika. [Kazakh onomastics] I vol.”, IC-Servis LLP (Astana), 2006, 400 p.
- [6] A.E. Alektorov, “Imena kirgizov [Kyrgyz names]”, *Orenburgskii listok*, 1868, pp. 4-6.
- [7] A.A. Divaev, “K voprosu o narechii imen u kirgizov [On the issue of the dialect of names among the Kyrgyz]”, *Turkestanskije vedomosti*, 1916.
- [8] N.F. Katanov, “Alfavitnyi ukazatel sobstvennykh imen, vstrechayuschichsya v pervom tome Obrazcov narodnoy literatury

- tyurkskix plemen, sobrannyx V. V. Radlovym [Alphabetical index of proper names found in the first volume of Samples of folk literature of Turkic tribes collected by V.V. Radlov]", Imp. Akad. Nauk, 1888.
- [9] Sh. Ualikhanov, "Tangdamaly [Custom]", Almaty, 1985, 557 p.
- [10] Kh.Kh. Zhubanov, "Iz istorii poryadka slov v kazakhskom yazyke [From the history of the order of words in the Kazakh language]", Almaty, 1999, 186 p.
- [11] T. Zh. Zhanuzakov, "Kazak yesimderdin tarikhy [History of Kazakh names (Linguistic and historical ethnographic analysis)]", Gylm (Almaty), 1971, 218 p.
- [12] T.Zh. Zhanuzakov, "Kazak tilindegi jalky yesimder [Proper names in the Kazakh language]", KHOZU SM KazSSR (Almaty), 1965, 145 p.
- [13] Rojas García J. et al. Semantic Representation of Context for Description of Named Rivers in a Terminological Knowledge Base, 2022.
- [14] Miller G., Beckwith R., Fellbaum C., Gross D., Miller K.J., "Introduction to WordNet: An On-line Lexical Database.", International Journal of Lexicography, Vol.3, No.4, 1990, pp. 235-244.
- [15] Azarova I.V., Mitrofanova O.A., Sinopalnikova A.A., "Kompyuternyi tezaurus russkogo yazyka tipa WordNet [Computer thesaurus of the Russian language type WordNet]", *Computational linguistics and intelligent technologies. Proceedings of the International Conference Dialogue*, 2003, pp. 43-50.
- [16] Ferschke, O., "Semantic Relations in WordNet and the BNC", 2009.
- [17] Ye.A. Kerimbayev, "Kazakhskaya onomastika v etnokul'turnom, nominativnom funktsional'nom aspektakh [Kazakh onomastics in ethno-cultural, nominative functional aspects.]", Almaty, 1995, 248 p.
- [18] Salah, R. E. and Zakaria, L. Q. B., "Building the classical arabic named entity recognition corpus (caner corpus)", *Journal of Theoretical and Applied Information Technology*, 96, 2018.
- [19] Cvetana Krstev, Dusko Vitas, Denis Maurel, Mickael Tran, "Multilingual Ontology of Proper Names", *2nd Language & Technology Conference (LTC'05)*, 2005, pp.116-119.
- [20] Krstev, Cvetana, and Denis Maurel, "A note on the semantic and morphological properties of proper names in the Prolex project", *Lingvisticae Investigationes 30.1*, 2007, pp. 115-133.
- [21] Gaziza Yelibayeva, Altynbek Sharipbay, Gulmira Bekmanova, Assel Omarbekova, "Ontology-Based Extraction of Kazakh Language Word Combinations in Natural Language," DATA'21, Processing in International Conference on Data Science, E-learning and Information Systems 2021, Association for Computing Machinery, New York, NY, USA, pp. 58-59. <https://doi.org/10.1145/3460620.3460631>
- [22] G.Yelibayeva, A. Mukanova, A.Sharipbay, A. Zulkhazhav, B. Yergesh and G.Bekmanova, "Metalanguage and Knowledgebase for Kazakh Morphology," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11619 LNCS, pp. 693-706, 2019, DOI: 10.1007/978-3-030-24289-3\_51
- [23] L. Zhetkenbai, A.A. Sharipbay, G.K. Yelibayeva, A.S. Mukanova, B.Zh. Yergesh, "Ontological model of nouns of Kazakh and Turkish languages," (Kazak zhane turik tilderinin zat yesiminin ontologiyalyk modeli [in Kazakh]) KazUTZU Khabarshysy, vol 3. 2019, pp. 439-445.
- [24] G. K. Yelibayeva, A. S. Mukanova, B. Sh. Razakhova, B. Zh. Yergesh, and L. Zhetkenbay, "Linguistic tagging and formal modeling of Kazakh noun adjoinment joining phrases," Messenger of Almaty University of power engineering and telecommunications, vol. 4, no 47, pp. 230-236, 2019.
- [25] Koeva S., Krstev C., Vitas D., Kyriacopoulou T., Martineau C., & Dimitrova T., "Semantic and syntactic patterns of multiword names: A cross-language study", *Multiword expressions*, 2018, 31 p.
- [26] G.B. Madiyeva, V. Suprun, "Teoriya i praktika onomastiki [Theory and practice of onomastics]", Kazakh university (Almaty), 2003, 15p.
- [27] Madiyeva, G. B., & Tilesbekkyzy, A. A. (2020). On some current trends in kazakh anthroponymy (based on materials from almaty). *Voprosy Onomastiki*, 17(2), 312-322. doi:10.15826/vopr\_onom.2020.17.2.030
- [28] Syzranova, G. Yu. (2013). *Onomastics*. Togliatti: TSU Publishing House, 248.
- [29] T.Zh. Zhanuzak, "Yesimder syry – Tainy imyen [Secrets of names.]", Almaty, 2004, 208 p.
- [30] Orynbay L., Sairanbekova A., Yelibayeva G., Bekmanova G., "Qazaq esimderining semantikalyyq bazasynyng qurylymyn anyqtau



- zholdary [Methods Of Determining The Structure Of The Semantic Database Of Kazakh Names]”, Bulatov A.Zh. (Nur-Sultan), 2022, pp. 303-309.
- [31] UBEK GAME. 2019. Kazakhsha Yesimder (Version 1.5) [Mobile App]. Google play. URL: <https://play.google.com/store/apps/details?id=com.ubekgame.kazbookname>
- [32] Harmonics. 2021. Kazakhsha Yesimder Magynasy (Version 1) [Mobile App]. Google play. URL: <https://play.google.com/store/apps/details?id=com.esim.qazaqshamagina>
- [33] Vladimir Demchenko. 2017. Kazakhskie imena (Version 2.0.46) [Mobile App]. App Store. URL: <https://apps.apple.com/kz/app/%D0%BA%D0%B0%D0%B7%D0%B0%D1%85%D1%81%D0%BA%D0%B8%D0%B5-%D0%B8%D0%BC%D0%B5%D0%BD%D0%B0/id907802750>
- [34] Namsor. A name checking technology. 2022. URL: <https://namsor.app/>
- [35] Back4App Inc. List of Names Dataset. 2022. URL: <https://www.back4app.com/database/back4app/list-of-names-dataset>

Table 1: 1-10 Marks (Properties)

Names	male's name	female's name	meaning	Affectio- nately spoken	expo- sed to chan- ge	Ma- in roo- t	Multip- le-root (comp- ound word)	Anci- ent nam- es	Clan- tribe names (ethnoanthr- oponym)	Histo- rical name
	1	2	3	4	5	6	7	8	9	10
Abyl ai	1		Name from Arabic- Abyl, Abil — grandfather, father. Abdilma mbet is a shortene d form of the name.					1		1
Abyl gazy	1		In Arabic, Gazi is a warrior, a soldier, that is, the grandfather, the father of a warrior, of the army.				1			
Agila		1	Smart, reasonab le		1	Agi- lya				
Agila		1	Derived from Arabic and Persian. Agilya is smart, reasonab le.					1		

Table 2: Multiple Root Attributes (Properties)

Names	Multiple-root (compound word)	1-root	language	meaning	part of speech	2-roots	language	meaning	part of speech
Aibibi	1	ai	Kazakh	celestial body	Noun	bibi	Tajik	woman, princess	Noun
Aibike	1	ai	Kazakh	celestial body	Noun	bike	Arabic	woman	Noun
Aibubi	1	bibi	Tajik	woman, princess	Noun	ai	Kazakh	celestial body	Noun
Aibyn									
Aigul	1	ai	Kazakh	celestial body	Noun	gul	Tajik	rose	Noun
Aigansha	1	ai	Kazakh	celestial body	Noun	khansha	Kazakh	khan's wife	Noun
Aigyzy	1	ai	Kazakh	celestial body	Noun	kzy	Arabic	silk	Noun
Aidana	1	ai	Kazakh	celestial body	Noun	dana	Tajik	expert, genius, scientist	Adj.; Noun
Aidariya	1	ai	Kazakh	celestial body	Noun	dariya	Iranian	sea, big river	Noun
Aizhamal	1	ai	Kazakh	celestial body	Noun	zhamal	Arabic	beautiful, graceful	Adj.
Aizhan	1	ai	Kazakh	celestial body	Noun	zhan	Iranian	life, the energy of life in the living, the food of life	Noun
Aijaniya	1	ai	Kazakh	celestial body	Noun	zhan	Iranian	life, the energy of life in the living, the food of life	Noun
Aizada	1	ai	Kazakh	celestial body	Noun	zada	Iranian	infant	Noun
Aizere	1	ai	Kazakh	celestial body	Noun	zer	Iranian	gold	Noun
Aiziba	1	ai	Kazakh	celestial body	Noun	ziba	Iranian	beautiful	Noun
Aiziya	1	ai	Kazakh	celestial body	Noun	ziyna	Arabic	light, beam, glow	Noun
Aikumis	1	ai	Kazakh	celestial body	Noun	kumis	Kazakh	precious metal	Noun

Table 3: 11-20 Marks (Properties)

Names	New names	Names associated with a historical event	Ancient Kazakh language	Ancient Turkish language	Ancient Germanic language	Ancient Hebrew	Borrowed from Western Europe	English	Latin	Spanish
	11	12	13	14	15	16	17	18	19	20
Agata								1		
Agnessa									1	
Ada						1				
Beibitgul		1								
Berta					1					
Bibisara						1				
Viktoriya		1							1	
Gulzhenis		1								
Gulmariya						1				
Damira	1									
Darkhan				1						
Yedige				1						
Yedil				1						
Zhanoraz			1							
Zhenisgul		1								
Zarema	1	1								
Ibrakhim						1				
Isa					1					
Marat							1			
Mariya						1				
Elvira										1
Emma					1					

Table 4: 21-30 Marks (Properties)

Names	Germ a	Scandinavi an	Celti c	Gree k	Irania n	From Arab ic	India n	Alt ai	Kalma k	Kyrg yz
	21	22	23	24	25	26	27	28	29	30
Ada	1									
Azat					1					
Aida						1				
Aibike						1				
Aigyz						1				
Aidariya					1					
Aizada					1					
Ainasha						1				
Arsen				1						
Artur			1							
Adina					1					
Asemka n					1					
Baibarys										1
Batu									1	
Bizhiken								1		
Vasilisa				1						
Venera										
Vera				1						
Dora				1						
Yelena				1						
Yelizave ta		1								
Inga		1								
Indira							1			
Klara	1									



Table 5:31-40 Marks (Properties)

Names	Mongolian	Tibetan	MANC HU	Turkish	Turkmen	Azerbaijani	Tajik	Tatar	Georgian	Kazakh
	31	32	33	34	35	36	37	38	39	40
Abira							1			
Adliya							1			
Azhar										1
Azharkhan										1
Aza							1			
Aimira										1
Akylbek				1						
Akylzhan				1						
Alpysbek				1						1
Altai	1									
Babas						1				
Baibarys					1			1		1
Bayan	1									
Bayansulu	1									1
Beisebai			1							
Bekdana				1			1			
Beksul				1						1
Dila									1	
Zemfira									1	

Table 6: 41-50 Marks (Properties)

Names	Russian	Ukrainian	In relation to God	Religious names	In relation to the ritual	For having no children	For having no boys	Depending on how old parents are	For many children in the family	Wishing more boys in the family
	41	42	43	44	45	46	47	48	49	50
Aimar	1									
Aisha				1						
Alpys								1		
Alpysbai								1		
Adila				1						
Alisha				1						
Baizhan									1	
Baikorkar										1
Bates	1									
Yermek								1		
Zarema	1									
Isa			1							
Kekil					1					
Kekilbai					1					
Oksana		1								
Sveta	1									
Toibala					1					
Toktar							1			
Toktarbai							1			
Turgan						1				
Tursynbai						1				

Table 7: 51-60 Marks (Properties)

Names	Given to the first born in the family	Given to the following born child	Given to the youngest in the family	Given to a child born with longing	Given to a long awaited boy	given to the only child	Given to twins	Related to the day, month, year of birth of the child	Fairy-tale characters (myth)	Literary names
	51	52	53	54	55	56	57	58	59	60
Aldar									1	
Alpamyas									1	
Ansagan				1						
Ansar				1						
Abubakir	1									
Aigerim										1
Auyel	1									
Ashir								1		
Bagdat										
Baibarys								1		
Balbala			1							
Balkadisha						1				1
Balsheker			1							
Yegizbai							1			
Yegizbek							1			
Yenlik										1
Yergesh		1								
Yerezhap								1		
Yerzada					1					
Yerziba					1					
Zhalgas		1								

Table 8: 61-70 Marks (Properties))

Names	Relat ed to land and water name s	Relat ed to the natur e	Relat ed to celest ial bodie s (Moo n, Star)	Realat ed to the weath er	Relat ed to precio us stones	Related to decorati ons	Relat ed to metal (iron) name s	Relate d to curre ncy	Relat ed to four- legge d anim al and their cubs	Relate d to predat ory animal s
	61	62	63	64	65	66	67	68	69	70
Aibike			1							
Aidana			1							
Aidari ya		1	1							
Aina						1				
Ainag ul						1				
Aisaul e		1	1							
Akbot a									1	
Akzhi bek						1				
Akkoz y									1	
Aktokt y									1	
Altai	1									
Altyn					1		1			
Altuna i			1		1		1			
Altyng ul					1		1			
Araish a		1								
Arysta n										1
Boran				1						
Boran bai				1						
Dinara								1		
Dilda					1		1	1		
Yedil	1									

Table 9: 71-80 Marks (Properties)

Names	In the name of the animal	in the name of the bird	in the name of the plant	in the name of the fruit	in the name of the dish	Related to family relations	Related to house furniture, property	Related to the color	Related to a body part, human body	Wishing a child to be a hero (brave, courageous, a warrior)
	71	72	73	74	75	76	77	78	79	80
Abira						1				
Azhargul			1							
Aibar										1
Aibat										1
Aibyn										1
Aina							1			
Ainagul			1				1			
Akbala								1		
Akmaral	1							1		
Aktamak								1	1	
Alkagul			1							
Alma				1						
Almagul			1	1						
Altynshash									1	
Alua					1					
Anar				1						
Anargul			1	1						
Asel					1					
Burkit		1								
Bugybai	1									
Bulbul		1								
Zylika						1				



Table 10: 81-90 Marks (Properties)

Names	Wishing a child to be smart	Wishing a child to be beautiful	Wishing a child to live long	Wishing a child to be dignified, respected	Wishing a child to be rich	Wishing a child to have many friends	Wishing a child to be a boss (head)	Wishing a child to be fair and honest	Wishing a child to be happy	Wishing a child to be generous and noble
	81	82	83	84	85	86	87	88	89	90
Abai	1									
Abzal				1						
Abyz	1									
Azhar		1								
Azhargul		1								
Azharkhan		1			1					
Aiadil								1		
Aibarsha					1					
Aibas										1
Aidana	1	1								
Akylbek	1									
Akylzhan	1									
Aman			1							
Amanbala			1							
Amangul			1							
Anarkhan				1						
Anisa						1		1		
Asemkhan		1		1			1			
Tamila						1				
Tanzila				1			1			
Tattibala		1								
Tendik								1		
Togzhan	1				1				1	

Table 11: 91-97 Marks (Properties)

Names	Wishing a child to be kind	Wishing a child to be hardworking	Wishing a child to be the child of his country and people	Wishing a child to whimsical	Wishing a child to artistic	Related to profession	Related to fabric
	91	92	93	94	95	96	97
Aibarsha							1
Aiymzhan	1						
Akbatas							1
Akzhibek							1
Aziza	1						
Amila		1					
Balym				1			
Batsayi	1						1
Badiga		1					
Bulbul					1		
Gulzhanys			1				
Gulzhikhan			1				
Gulnaz				1			
Dariga					1		
Dilnaz	1			1			
Duriya							1
Yelaman			1				
Yeldar			1				
Yerke				1			
Yerkebai				1			
Zhazira	1		1				
Kulasiya	1					1	
Mulipa						1	
Mushtari						1	