Linguistic Foundations of Low-Resource Languages for Speech Synthesis on the Example of the Kazakh Language

Gulmira Bekmanova, Banu Yergesh, Altynbek Sharipbay, Assel Omarbekova(✉), and Alma Zakirova

L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan
omarbekova_as@enu.kz

Abstract. The article considers the linguistic foundations of the Kazakh language for the Kazakh speech synthesis system. A formalization of phonological rules is presented for the implementation of an automatic transcription of words into transcriptional notation, which can be applied to speech synthesis and recognition systems.

Keywords: Speech synthesis · Transcriptor · Phonological rules

1 Introduction

This work 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". The aim of the project is to develop scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language as a language of interethnic communication in digital format, which is an urgent and important problem in the strategic direction of development of our state. The solution of this problem will be carried out on the basis of: an analysis of the scientific, methodological and normative foundations of the grammar of the Kazakh language and computational linguistics; research of models and methods of speech synthesis; building a formal description of the grammar of the Kazakh language, artificial intelligence technology for the development of an intelligent synonymizer, an electronic reference book, a mobile application "Fascinating Onomastics", an electronic dictionary of terminology of school textbooks, a Kazakh speech synthesizer and an intellectual system "Akhmettanu". The main approaches to conducting research are the construction of a formal description of the grammar of the Kazakh language, the creation of a database of synonymous words, proper names, and terms for the relevant subject areas (socio-political and public life, grammar of the Kazakh language and school textbooks), taking into account their history, use and possible interpretations, the formation of an audio and text corpus for the synthesis of Kazakh speech, as well as the creation of a knowledge base on Akhmet Baitursynuly's scientific heritage and on all structural tiers of the language in the light of his teachings. The solution to this problem is possible only

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
O. Gervasi et al. (Eds.): ICCSA 2022 Workshops, LNCS 13379, pp. 3–14, 2022.
https://doi.org/10.1007/978-3-031-10545-6_1

with the use of intelligent technologies for computer processing (analysis and synthesis) of oral and written texts (data) in the Kazakh language. This article will consider the linguistic foundations of the synthesis of Kazakh speech based on Akhmet Baitursynuly's works, as well as some ethical aspects of the perception of his scientific and cultural heritage.

2 Related Works

The development of low-resource languages is accompanied by a number of difficulties, which are different in countries depending on their historical and cultural heritage.

There are works [1–4] on natural language processing, where various ethical issues related to the processing of natural languages are solved.

Similar research and development of low-resource languages is carried out to prevent its disappearance and its development, to study the knowledge contained in the original works in these languages in [5–7]. The ethical aspects of the development of low-resource languages are also often touched upon, since often low-resource languages are infringed for various reasons in their development. When it comes to Akhmet Baitursynov, an outstanding Kazakh scientist (1872–1937), philologist, publicist [8], etc., a multifaceted scientist whose life was tragically cut short due to his patriotic and political activities, it is difficult to choose the type of his activity that made the greatest contribution to scientific and social-political life of Kazakhstan [9–12]. However, this work will use the fundamental foundations laid by A. Baitursynov in the linguistics and phonetics of the Kazakh language, which became the basis for developing a transcriptor for the Kazakh speech synthesis system. Particularly surprising is the fact that all his studies were carried out about 100 years ago [13–20], however, they most accurately reflect the modern sound and phonetic structure of the Kazakh language and allow building the highest quality speech synthesizer. Speaking about the ethical aspect of the attitude towards A. Baitursynov's works, it should be noted that the attitude towards his works changed over the years and they were completely unpopular in the Soviet era of the development of the Kazakh language, in view of the unpopularity of the views of the author, who was recognized as an enemy of the then existing government and only newfound independence made possible a fresh look at his writings. The development of artificial intelligence and the creation of systems for recognizing and synthesizing Kazakh speech only emphasized the fundamental nature and scientific validity and value of his scientific heritage.

3 Kazakh Speech Synthesis System

The Kazakh speech synthesis system is not fundamentally different from any other speech synthesis system. The peculiarity of the language has an influence mainly on the transcriptor, which feeds the transcription of the text to the input of the model that directly synthesizes the sound. Of course, phonological features of the language may influence the choice of a model. However, a number of previous works of the authors have shown that the synthesis of Kazakh speech is implemented quite effectively on existing models, whether they are probabilistic models on Hidden Markov models, models on neural networks, diphonic, phonemic and other models [21–24].

4 Transcriptor

A transcriptor of words into a transcription record is present in all natural language processing systems, such as translators, recognition systems, or speech synthesis. However, depending on the intended purpose of the system, we select one or another transcriptor model. So, for example, a simplified transcription can be used for speech recognition systems, while for speech synthesis systems it is necessary to use the most detailed transcription containing the acoustic characteristics of sounds.

The segmental phonetic transcription of the text, the construction of which completes the work of the linguistic block, is carried out by an automatic transcriptor. To do this, first of all, the standard reading rules that operate in the language are used, that is, equivalents of the type "letter ® phoneme ® sound". The degree of phonetic detail (the number of distinguishable sounds), which is selected in specific synthesis or recognition systems when constructing transcription, can be different. Often the choice is dictated by tradition, followed by phonetic descriptions of the synthesized language and the most representative and commonly used dictionaries.

In a number of languages (for example, English), the relationship between spelling and pronunciation is very complex – many words do not follow standard reading rules. Words whose pronunciation "falls out" of the standard rules have to be processed separately, either by creating narrower, specific rules for them, or by writing the phonetic transcription of the word directly into its dictionary characteristic, i.e., memorizing, and not deriving from reading rules.

Representatives of the linguistic approach emphasize the need for a wider use of phonetic and physiological knowledge in automatic speech recognition and synthesis systems.

According to the American researcher V. Zu [25], one of the most active supporters of this approach, the failures of acoustic phonetic recognition in various systems are due to two reasons:

1) the use of too simplified ideas about the relationship between the speech signal and its phonemic (allophonic) reflection in the language;
2) the use of such methods of primary acoustic processing of the speech signal, which are too far from how it is carried out in the human auditory system.

The phonetic alphabet is the basis for the operation of the speech recognition/synthesis block. The symbols of the phonetic alphabet should unequivocally correspond to those sounds, the distinction of which is essential in the process of recognition. Therefore, minimizing the size of the phonetic alphabet without compromising the quality of recognition should be carried out by identifying in the alphabet only those sounds that are closest in sound from the point of view of a person [26].

Therefore, for the correct construction of the transcriptor and the use of the phonetic alphabet, we first need to translate the record of the word into an intermediate alphabet of 31 letters, which are characteristic of the Kazakh language. At the second stage, this record of the word is translated according to the formalized linguistic (phonological) rules of the Kazakh language [27]. At the third stage, the received words are translated into a transcription record in symbols of international transcription.

It should be noted that the current Cyrillic alphabet of the Kazakh language instead of 31 letters contains 42 letters [28]. Of these, 11 letters were erroneously introduced in 1940 only to ensure that the writing and reading of Russian words in the Kazakh text were carried out in accordance with the norms of the Russian language. These include: Yo, I, Ch, Sch, Ts, ъ, E, Yu, Ya, hard sign, soft sign.

Therefore, for the correct construction of the transcriptor and the use of the phonetic alphabet, we first need to translate the record of the word into an intermediate alphabet of 31 letters. Equivalence of characters is presented in Table 1.

Table 1. Symbols of the current alphabet and the intermediate alphabet

Current alphabet	Intermediate alphabet	Transcription	Current alphabet	Intermediate alphabet	Transcription
А	А	[ɑ]	Б	Б	[b]
Ә	Ә	[æ]	В	В	[v]
Е	Е	[e]	Г	Г	[g]
О	О	[ɔ]	Ғ	Ғ	[ɣ]
Ө	Ө	[ɵ]	Д	Д	[d]
Ұ	Ұ	[ʊ, u]	Ж	Ж	[ʒ]
Ү	Ү	[ʏ]	З	З	[z]
Ы	Ы	[ɯ]	Й	Й	[y]
І	І	[ɪ, i]	К	К	[k]
Э	Е	[jɪ]	Қ	Қ	[q]
Я	ЙА	[yɑ]	Л	Л	[l]
Ю	ЙУ	[yw]	М	М	[m]
Ё	ЙО	[yɔ]	Н	Н	[n]
И	ІЙ	[iy]	Ң	Ң	[ŋ]
И	ЫЙ	[ɯj]	П	П	[p]
Ч	Ш	[tʃ]	Р	Р	[r]
Щ	Ш	[ʃ]	С	С	[s]
Ц	С	[tc]	Т	Т	[t]
һ	Х	[h]	У	У	[w]
Ъ	-		Ш	Ш	[ʃ]
Ь	-		Ф	Ф	[f]
			Х	Х	[h]

5 Methods for Analyzing Natural Language Sounds

To study the sounds of natural language, phonetic methods and phonological methods of analysis will be used.

Phonetic methods of analysis are:

– articulation analysis;
– acoustic analysis;
– Perceptual analysis.

Articulatory analysis determines the organs of speech and their movements involved in the formation of the sounds of the language. The articulatory analysis of the tongue establishes the following articulatory features:

– participation of voice and noise in the formation of sounds;
– the place of formation of sounds (both active and passive organs of speech must be indicated);
– the way sounds are produced.

According to the involvement of the voice, the sounds of the language can be divided into vowels and consonants. When pronouncing the vowels of the language, the vocal cords vibrate completely, and there is no contact between the organs of speech and an obstruction in the oral cavity. When pronouncing consonants, there is an incomplete oscillation of the vocal cords (sonorous consonants and noisy consonants) or it is absent (deaf consonants), and in the oral cavity there is contact between the organs of speech and a barrier.

According to the place of formation, the vowel sounds of the language can be divided into: front-lingual (alveolar); pre-lingual; back-lingual (uvular); labial; non-labial.

According to the place of formation, consonant sounds of the language can be divided into: front-lingual dental; front-lingual front-palatal; pre-lingual pre-palatal; back-lingual pre-palatal; back-lingual post-palatal; bilabial; labial-dental. In the first place is the active organ of speech, in the second – the passive organ of speech.

According to the method of formation, the sounds of the language can be subdivided into occlusive and fricative. When pronouncing occlusive consonants in the oral cavity, a complete closure of the organs of speech occurs and the air stream, noisily overcoming the closure, finds a way out. When pronouncing fricative consonants in the oral cavity, a gap is formed between the organs of speech and the air stream seeps noisily through the gaps.

For occlusive consonant sounds, exposure turns on the moment of complete cessation of the flow of the air stream through the oral cavity. Fricative consonants can be subdivided into: explosive; confluent; nasal; flicking; explosive-lateral; vibrant.

The acoustic analysis of the language determines the presence of phonetic patterns in the formation of syllables and words from vowels and consonants. In the Turkic languages, including the Kazakh language, such phonetic regularity is vowel harmony.

Vowel harmony requires that all vowels and consonants in a syllable or a word be pronounced with a single timbre.

Vowel harmony is formed from a combination of four timbres: solid non-labial; soft non-labial; hard labial; soft labial.

Speech recognition uses acoustic analysis based on the physical characteristics of a sound.

Based on the physical nature of sounds, sounds can be classified into nine groups of features:

- vocality – non-vocality

1. vocal sounds: Ә, А, І, Ү, Ұ, Ө, Е, Ы, О, Ң, Н, М, Р, Л, Й, У
2. non-vocal sounds: П, Б, Т, Д, К (Қ), Г (Ғ), С, З, Ш, Ж, Ф, В, Х (һ)

- consonance – non-consonance

1. consonant: П, Б, Т, Д, К (Қ), Г (Ғ), С, З, Ш, Ж, Ф, В, Х(һ), У, Р
2. non-consonant: Ә, А, І, Ү, Ұ, Ө, Е, Ы, О

- high pitch – low pitch

1. high pitch
2. low pitch

- diffuseness – compactness

1. diffuse
2. compact

- flatness – non-flatness

1. flat: О, Ө, Ү, Ұ, П°, Б°, М°, У, Т°, Д°, Н°, Қ°, Ғ°, Ң°, Й°, С°, З°, Р°, Ш°, Ж°, Л°, П′°, Б′°, М′°, У′, Т′°, Д′°, Н′°, Қ′°, Ғ′°, Ң′°, Й′°, С′°, З′°, Р′°, Ш′°, Ж′°, Л′°
2. non-flat: Ә, А, І, Ы, П, Б, М, У, П′, Б′, М′, У, Т, Д, Н, Т′, Д′, Н′, Қ, Ғ, Ң, Й, К′, Г′ Ң′, Й′, С, З, Р, С′, З′, Р′, Ш, Ж, Л, Ш′, Ж′, Л′, К, Г Ң, Й

- sharpness – non-sharpness

1. sharp: Ә, І, Ү, Ө, П′, Б′, М′, Т′, Д′, Н′, К′, Г′ Ң′, Й′, С′, З′, Р′, Ш′, Ж′, Л′
2. non-sharp: А, Ұ, Ы, О, П, Б, М, Т, Д, Н, К, Г Ң, Й, С, З, Р, Ш, Ж, Л, П°, Б°, М°, У, Т°, Д°,Н°, Қ°, Ғ°, Ң°,Й° ,С°,З°, Р°, Ш°,Ж°, Л°, П′°,Б′°,М′°,У′, Т′°,Д′°,Н′°, Қ′°, Ғ′°, Ң′°,Й′° ,С′°,З′°, Р′°, Ш′°,Ж′°, Л′°

- discontinuity (pausedness) – incessancy

1. non-continuous: П, Б, Т, Д, К (Қ), Г (Ғ)
2. continuous: Ә, А, І, Ү, Ұ, Ө, Е, Ы, О, М, Н, Ң, С, З, Р, Ш, Ж, Л, Й, У

- sharpness – non-sharpness

1. sharp: Р,
2. non-sharp: Ә, А, І, У, Ұ, Ө, Е, Ы, О, П, Б, Т, Д, К (Қ), Г (Ғ), С, З, Ш, Ж, Ф, В, Х (Һ), У

- sonority – voicelessness

1. voiced sounds: Ә, А, І, У, Ұ, Ө, Е, Ы, О, Ң, Н, М, Р, Л, Й, У, Б, Г (Ғ), З, Ж, В
2. unvoiced sounds: П, Т, Д, К (Қ), С, Ш, Ф, Х(Һ)

Phonological methods are:

– synharmonic-articulatory analysis;
– synharmonic-acoustic analysis;
– synharmonic-perceptive analysis.

6 Formalization of Phonological Rules of Sound Combinations in the Kazakh Language

The Kazakh language has 9 vowels and 22 consonants.

Let introduce the following notation for vowels: G is the set of Kazakh vowels, G_1 is the set of hard vowels, G_2 is the set of soft vowels, G_3 is the set of non-labial vowels, G_4 is the set of labial vowels, G_5 is the set of back-lingual vowels, G_6 is a set of front-lingual vowels, G_7 is a set of open vowels, G_8 is a set of closed vowels, G_9 is a median vowel in pairwise distinguishable features of sounds in each group, i.e.

$$
\begin{aligned}
G &= \{А, Ә, О, Ө, Ұ, Ү, Ы, І, Е\};\\
G_1 &= \{А, О, Ұ, Ы\} \subseteq G;\\
G_2 &= \{Ә, Ө, Ү, І\} \subseteq G;\\
G_3 &= \{А, Ә\ , Ы, І\} \subseteq G;\\
G_4 &= \{О, Ө, Ұ, Ү\} \subseteq G\ ;\\
G_5 &= \{А, О, Ұ, Ы\} \subseteq G;\\
G_6 &= \{Ә, Ө, Ү, І\} \subseteq G\ ;\\
G_7 &= \{А, Ә, О, Ө\} \subseteq G;\\
G_8 &= \{Ұ, Ү, Ы, І\} \subseteq G;\\
G_9 &= \{E\} \subseteq G.
\end{aligned}
$$

We also introduce the notation for consonants: S – consonants of the Kazakh language, S_1 – unvoiced consonants, S_2 – voiced consonants, S_3 – sonorous consonants, S_4 – labial consonants, S_5 – anterior lingual consonants, S_6 – midlingual consonants,

S_7 – posterior lingual consonants, S_8 – stop consonants, S_9 – round-slit consonants, S_{10} – flat-slit consonants, S_{11} – tremulous consonants, S_{12} – lateral consonants, that is

$$S_1 = \{К, Қ, П, Ф, С, Т, Х, Ш\} \subseteq S;$$
$$S_2 = \{Б, В, Г, Ғ, Д, Ж, З\} \subseteq S;$$
$$S_3 = \{Й, Л, М, Н, Ң, Р, У\} \subseteq S;$$
$$S_4 = \{Б, М, П, У\} \subseteq S;$$
$$S_5 = \{Д, Ж, З, Л, Н, Р, С, Т, Ш\} \subseteq S;$$
$$S_6 = \{В, Г, К, Ф\} \subseteq S;$$
$$S_7 = \{Ғ, Й, Қ, Х\} \subseteq S;$$
$$S_8 = \{Б, Г, Ғ, Д, К, Қ, М, Н, Ң, П, Т\} \subseteq S;$$
$$S_9 = \{З, С, У\} \subseteq S;$$
$$S_{10} = \{Ж, Й, Ш\} \subseteq S;$$
$$S_{11} = \{Р\} \subseteq S;$$
$$S_{12} = \{Л\} \subseteq S;$$

Let A be the alphabet of the Kazakh language, i.e. A = G ∪ S, then the set of all chains (words) of finite length in the alphabet A is denoted as:

$$A^* = A^0 \cup A^1 \cup A^2 \cup A^3 \cup \ldots, \tag{1}$$

where $A^0 = \{\varepsilon\}$ is a set of words of 0 length, $A^1 = A$ is a set of words of 1 length, $A^2 = A \times A$ is a set of words of 2 length, $A^3 = A \times A \times A$ is a set of words of 3length, x is the sign of operations of the direct (Cartesian) product of a set.

The elements of the set A* will be denoted by lowercase (small) Greek letters: α, β, γ,…. In this case, a word of 0 length (the empty word) will always be denoted by ε.

The set of all words of non-zero length in the alphabet A are denoted by $A+$, which is formed by excluding the set of empty words from A^*, i.e.

$$A^+ = A^* \backslash \{\varepsilon\} \tag{2}$$

It is clear that any phonological unit of the Kazakh language is an element of the set A^+ [29–31].

Now we will start the formalization of phonologic rules of the Kazakh language:

1. Connection of two unvoiced consonants. On a connection of two unvoiced consonants in one word or in a case when one word ends to a unvoiced consonant, and another word begins with a unvoiced consonant, both consonants remain unvoiced:

$$\frac{\alpha \subseteq \gamma, \alpha = \alpha_1 x, \alpha_1 \neq \varepsilon, x \in S_{15}, \beta \subseteq \gamma, \beta = y\beta_1, y \in S_3, z \in S_2}{\gamma = (\alpha_1 x)(z\beta_1)} \tag{3}$$

Example: The junction of two unvoiced consonants is possible at the root of the word “akpan”, “akpar”, at the junction of the root and the affix (suffix or ending) “atakpen”, “bakpen”, at the junction of two words “ak sandyk”, “ak sausak”. In all these cases, the transformation of unvoiced consonants does not occur [28].

2. Connection of two consonants unvoiced and sonorous.

2.1. During the connection of two consonants, unvoiced and sonorous, in one word or in one case when one word ends with a unvoiced front tongue consonant (с, т, ш), and another begins with a sonorous consonant, the unvoiced consonant remains unvoiced, and soa norous consonant remains sonorous:

$$\frac{\alpha \subseteq \gamma, \alpha = \alpha_1 x, \alpha_1 \neq \varepsilon, x \in S_1 \cap S_5, \beta \subseteq \gamma, \beta = y\beta_1, y \in S_3}{\gamma = (\alpha_1 x)(y\beta_1)} \tag{4}$$

where $S_1 \cap S_5 = \{\text{С, Т, Ш}\}$.

Example: The junction of two unvoiced and sonorous consonants is possible at the root of the word "asmar", at the junction of the root and the postfix (suffix or ending) "betmonshak", at the connection of two words "zhas mal". In all these cases, the transformation of consonants does not occur [28].

2.2. During the connection of two consonants, unvoiced and sonorous, in one word or in one case when one word ends with an unvoiced non-front tongue consonant (к, к, п, ф,), and another begins with a sonorous consonant sound, the unvoiced consonant is replaced with its matching sonorous consonant:

$$\frac{\alpha \subseteq \gamma, \alpha = \alpha_1 x, \alpha_1 \neq \varepsilon, x \in S_{15}, \beta \subseteq \gamma, \beta = y\beta_1, y \in S_3, \text{z} \in S_2}{\gamma = (\alpha_1 x)(z\beta_1)} \tag{5}$$

where $S_{15} = S_1 - (S_1 \cap S_5)$

Example: The connection of two unvoiced and sonorous consonants is possible at the junction of two roots Bekmagambet (pronounced Begmagambet), Aknur (pronounced Agnur) in compound words and at the junction of two words ak mandai (pronounced agмandai), kok nayza (pronounced kognayza). In all these cases, there is a transformation of consonants [28, 31].

3. During the connection of two consonants, voiced and unvoiced, in one word or at the connection of two words, the voiced consonant remains voiced, and the unvoiced consonant is replaced with its matching voiced consonant

or

voiced consonant is replaced with its matching unvoiced consonant and unvoiced consonant remains unvoiced

$$\frac{\alpha \subseteq \gamma, \alpha = \alpha_1 x, \alpha_1 \neq \varepsilon, x \in S_2, \beta \subseteq \gamma, \beta = y\beta_1, y \in S_1, \text{z} \in S_2}{\gamma = (\alpha_1 x)(z\beta_1)} \tag{6}$$

$$\frac{\alpha \subseteq \gamma, \alpha = \alpha_1 x, \alpha_1 \neq \varepsilon, x \in S_2, z \in S_1, \beta \subseteq \gamma, \beta = y\beta_1, y \in S_1}{\gamma = (\alpha_1 z)(y\beta_1)} \tag{7}$$

Example: The junction of two voiced and unvoiced consonants is possible at the junction of a root and a postfix and at the junction of two roots sozsiz (pronounced sozsiz), kymyzkor (pronounced kymyzkor) and at the junction of two words tez pisti (pronounced tes pisti), zhuz pyshak (pronounced zhus pyshak). However, it is possible to pronounce it as follows: kymyz kor – kymyz gor, tez pisti – tez bisti, zhuz pyshaq - zhuz byshak [28, 31].

4. During the connection of two consonants, unvoiced and voiced, in one word or at the connection of two words, the unvoiced consonants is replaced with its matching voiced consonant, and the voiced consonant remains voiced
   or
   unvoiced consonant remains unvoiced, and voiced consonant is replaced with its matching unvoiced consonant.

$$\frac{\alpha \subseteq \gamma, \alpha = \alpha_1 x, \alpha_1 \neq \varepsilon, x \in S_1, z \in S_2, \beta \subseteq \gamma, \beta = y\beta_1, y \in S_2}{\gamma = (\alpha_1 z)(y\beta_1)} \tag{8}$$

$$\frac{\alpha \subseteq \gamma, \alpha = \alpha_1 x, \alpha_1 \neq \varepsilon, x \in S_1, \beta \subseteq \gamma, \beta = y\beta_1, y \in S_2, \ z \in S_1}{\gamma = (\alpha_1 x)(z\beta_1)} \tag{9}$$

   Example: The connection of two unvoiced and voiced consonants is possible at the root of the word akbar (pronounced akpar), does not occur at the junction of the root and the postfix, at the junction of two words ak gul (pronounced akkul). In all these cases, the voiced consonant turns into an unvoiced consonant. However, you can pronounce these words in a different way: akbar (pronounced agbar), ak gul (pronounced aggul [28, 31, 32].

The formal phonological rules of the Kazakh language obtained in this way will make it possible to build a transcriptor that automatically builds a transcription of a word, which is an important element in the development of a Kazakh speech synthesis system.

7 Conclusion

This paper presents the results of the formalization of phonological rules and their application in the implementation of a transcriptor for a speech synthesis system. The transcriptor was based on the works of the Kazakh scientist, reformer of national writing, the founder of Kazakh linguistics A. Baitursynov. Formalization, digital processing of the works of A. Baitursynov contributes to the improvement of the culture of the Kazakh language, as a language of interethnic communication in digital format, will allow obtaining new scientific results in terms of the scientific and linguistic foundations of the Kazakh language for various intelligent information systems that process natural languages, will allow transferring knowledge to the future generation in implemented applications, and will also have a direct impact on the development of the ICT industry in Kazakhstan and the quality of digital services in the Kazakh language.

8 Further Work

This work 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”. In total, 7 big tasks are planned in the project, 2 of which are devoted to the synthesis of the Kazakh speech, and 1 – to the creation of an intellectual knowledge system on A. Baitursynov’s heritage. After the

creation of the transcriptor, it is planned to create a speech synthesis system based on the integrated End-to-End model. At the same time, it is also symbolic that the modern phonetics of the Kazakh language, used as a theoretical basis for the synthesis of the Kazakh speech, is based on the works of this outstanding scientist.

Funding. This research is funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (grant no. BR11765535).

References

1. Baia, A.E., Biondi, G., Franzoni, V., Milani, A., Poggioni, V.: Lie to me: shield your emotions from prying software. Sensors **22**(3), 967 (2022). https://doi.org/10.3390/s22030967
2. Milani, A., Franzoni, V., Biondi, G.: Parsing tools for Italian phraseological units. In: Gervasi, O., et al. (eds.) ICCSA 2021. LNCS, vol. 12955, pp. 427–435. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87007-2_30
3. Atif, M., Franzoni, V., Milani, A.: Emojis pictogram classification for semantic recognition of emotional context. In: Mahmud, M., Kaiser, M.S., Vassanelli, S., Dai, Q., Zhong, N. (eds.) BI 2021. LNCS (LNAI), vol. 12960, pp. 146–156. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86993-9_14
4. Franzoni, V., Milani, A., Mengoni, P., Piccinato, F.: Artificial intelligence visual metaphors in e-learning interfaces for learning analytics. Appl. Sci. **10**(20), 7195 (2020). https://doi.org/10.3390/app10207195
5. Magueresse, A., Carles, V., Heetderks, E.: Low-resource languages. a review of past work and future challenges. ArXiv, abs/2006.07264 (2020)
6. Sebastian, R.: The 4 biggest open problems in NLP (2019)
7. Emily, B.: High Resource Languages vs Low Resource Languages. The Gradient (2019)
8. Dossanova, N., Abdimanuly, O., Kapagan, E.: The influence of Akhmet Baitursynov's literary publicistic heritage on the spiritual development of Kazakh people at the beginning of The XXth century. J. Hist. Cult. Art Res. **9**, 286 (2020). https://doi.org/10.7596/taksad.v9i4.2669
9. Allworth, Ed.: Central Asia: A Century of Russian Rule. 1st Edn. University Presses of California, Columbia and Princeton (1967)
10. Abdimanuly, O.: Akhmet Baytursynuly: zertteu-essay. Arda, Almaty, 293 b. (2007)
11. Dakenov, M.: Baitursynov's role in the development of the sociology of culture and education in Kazakhstan. Bulletin Abay Kazakh National Pedagogical University (2012)
12. Aibarshyn, A.: The Man Who Led His Nation to Enlightenment: Kazakhstan Marks 150th Anniversary of Prominent Scholar Akhmet Baitursynov. The Astana times (2022)
13. Baytursinov, A.: New primer with drawings (8–17 thousand). Qizilorda (1926). (in Kazakh)
14. Baytursinov, A.: Problems of Kazakh linguistics. Abzal-Ai, Almaty (2013). (in Kazakh)
15. Baytursinov, A.: Til – qural. Sardar, Almaty (2009). (in Kazakh)
16. Baytursinov, A.: Til tagilimi. Pavlodar (2008). (in Kazakh)
17. Baytursinov, A.: Forty examples: poems. Zhazuschy, Almaty (2001). (in Kazakh)
18. Baytursinov, A.: Alphabet: new manual. Rauan, Almaty (1998). (in Kazakh)
19. Baytursinov, A.: New rules. Zhety zhargy, Almaty (1996). (in Kazakh)
20. Baytursinov, A.: A five-volume collection of works. Volume 1: Poems, literary and scientific research. Alash, Almaty (2003). (in Kazakh)
21. Yessenbayev, Z., Karabalayeva, M., Sharipbayev, A.: Formant analysis and mathematical model of Kazakh vowels. In: Proceedings - 2012 14th International Conference on Modelling and Simulation, UKSim 2012, article№ 6205485, pp. 427–431 (2012). https://doi.org/10.1109/UKSim.2012.66

22. Yessenbayev, Zh., Saparkhojayev, N., Tibeyev, T.: Implementation of the intelligent voice system for Kazakh. J. Phys.: Conf. Ser **495**(1), статья № 012043 (2014). https://doi.org/10.1088/1742-6596/495/1/012043
23. Sharipbayev, A.A., Bekmanova, G.T., Shelepov, V.U.: Formalization of Phonologic Rules of the Kazakh Language for System Automatic Speech Recognition. http://dspace.enu.kz/handle/data/1013. Accessed 12 May 2022
24. Bekmanova, G., Yergesh, B., Sharipbay, A., Mukanova, A.: Emotional speech recognition method based on word transcription. Sensors (Basel, Switzerland) **22**(5), 1937 (2022). https://doi.org/10.3390/s22051937
25. Zu, V.V.: Linguistic approach to automatic recognition of speech signals. In: Proceedings of the Institute of Electrical and Radio Electronics Engineers (TIIER). Speech Communication with Machines, no. 73, pp. 51–54 (1985)
26. Sharipbayev, A., Torekhanov, S.: Qazaq tili dauisti dibistarinin matematikalik modeli. L.N.Gumilyov atindagi Euraziya ulttik universiteti. «Khabarshi» № 3–4, 2002. 14. Qazaq grammatikasi, Astana (2002)
27. Bekmanova, G., Nicenko, A., Sharipbayev, A., Shelepov, V.: O nekotoryh voprosah svyazannyh s raspoznavaniem kazakhskoy rechi. Bulletin ENU, Astana (2009)
28. Kazakh Grammar. Phonetics, Word Formation, Morphology, Syntax; Astana-Poligraphy: Astana, Kazakhstan (2002). (in Kazakh)
29. Babin, D., Mazurenko, I., Holodenko, A.: On the Prospects of Creating a System of Automatic Recognition of Fused Russian-Speech. Mir, Moscow (1999)
30. Bekmanova, G.: Transcript of Kazakh words for speech recognition. Bulletin NAN RK, Almaty (2009)
31. Sharipbayev, A., Bekmanova, G.: Mathematical methods of recognizing combinations of sounds of the Kazakh language. In: Proceedings of the International Scientific-Practical Conference "Innovative Technologies: Integration of Science, Education and Business" KazNTU. K. Satpayeva. Almaty (2008)
32. Bekmanova, G., Nicenko, A., Sharipbayev, A., Shelepov, V.: Algorithms for recognizing Kazakh words as a whole. Structural classification of words of the Kazakh language. Bulletin ENU, Astana (2010)