Z.M. Bazarbayeva¹ ⓘ **, A.B. Srailova²** ⓘ **, Zh.T.Zhumabayeva¹** ⓘ

¹The Institute of Linguistics named after A.Baitursunuly, Almaty, Kazakhstan
²Kazakh Ablai khan University of International Relations and World Languages, Almaty, Kazakhstan
(E-mail: zeynepmb@mail.ru, maral_albi@mail.ru, janar_tolendi@mail.ru)

Poetic subcorpus: prosodic marking

Abstract. *This article is dedicated to describing the prosodic marking of poetic texts in the Kazakh language. Prosodic marking of a text involves segmenting it into syntagmas, marking syntagmas into rhythmic groups, and marking the intonation type of syntagmas according to specific rules.*
The term 'syntagma' refers to an independent, intonationally meaningful part of a phrase or the entire phrase itself. Establishing the boundaries of syntagmas affects the conveyance of semantic content and the type of intoneme. When dividing the text into syntagmas, it is crucial not to place the boundary of a syntagma where it may disrupt the meaningful perception of speech. Specific rules of syntagmatic segmentation, which is based on punctuation and phonetic analysis of the text, are employed to establish the boundaries of syntagmas in natural speech.
Therefore, the findings of prosodic analysis showed that poetic discourse in Kazakh language, as in other Turkic languages, is based on a syllabic structure, where a stable number of syllables per line is a necessary condition for the existence of rhythm. The prosodic marking of poetic subcorpora indicates that prosodic elements serve as a means of expressing punctuation marks in oral speech.
Key words: *poetic subcorpus, sintagma, prosodic marking, spoken speech, corpus linguistics, annotation, meta-marking.*

Introduction.

Corpus linguistics is one of the most promising directions in contemporary theoretical and applied linguistics, based on the use of language electronic corpora and corpus technologies. A language corpus refers to a collection of texts in a particular language in electronic form, specially processed ("annotated") so that a language researcher can quickly and comprehensively find information of interest about the language in the corpus. Depending on the nature of annotation, it is possible to search for grammatical characteristics of words and sentences in the language (prosodic, morphological, syntactic, semantic parameters), as well as various characteristics of the texts themselves within the corpus: by author, date of text creation, genre, theme, etc.

The richer and more diverse the annotation – and the greater the number of texts in the corpus – then the corpus becomes the more valuable tool for linguistic research. The most valuable and commonly used type of corpus is the so-called "national corpus," which refers to a maximally representative collection of all types of texts available in a given language during a

specific period – including literary and artistic, scientific, and journalistic texts, both written and oral texts.

In world science, the first corpora began to emerge practically simultaneously with the introduction of computer technologies in humanities research. However, the mass growth of corpus studies and the creation of new corpora occurred in the late 1980s and mid-1990s. It was at this time that large national corpora of English, Italian, Finnish, Czech, and several other languages (mainly European) emerged. It is also during this time that the field of theoretical linguistics realizes that a representative corpus is not just a powerful means of searching for examples in texts; it should be regarded as a fundamentally new tool whose application leads to revolutionary results in language research.

Applied linguistic technologies in the modern world are used wherever it is necessary to extract information from a large array of data in natural language. Corpora turn out to be irreplaceable in this expansive area – it is precisely all modern information retrieval systems are created on the basis of corpora – primarily on the Internet – and automatic information extraction, automatic analysis of text structure, automatic translation (using so-called parallel corpora in several languages), automatic speech recognition (using corpora of spoken speech) and solves many other tasks.

The creation, development and use of electronic corpora is one of the most advanced directions in contemporary linguistics; it is within this framework that innovative results are most likely in both theoretical linguistics (acquiring new knowledge about the structure of language) and applied linguistics (developing next-generation technologies for automatic text processing and accelerated modernization of linguistic research methods).

The scientific novelty of this research lies in the fact that the annotation of poetic subcorpus at the prosodic level has not been previously conducted in Kazakh linguistics, as Kazakh prosody, as a direction within domestic linguistics, is exclusively investigated in the phonetics department of the Institute of Linguistics named after A.Baitursynov, this underscores the uniqueness of this research work.

The aim of this work is to describe the prosodic marking of a poetic subcorpus for the development and creation of the national corpus of the Kazakh language. To achieve this aim, it is necessary to identify the connection between prosodic elements and punctuation marks. In our work, punctuation is understood as the written representation of the semantic, intonational, and syntactic capabilities of oral speech. The prosodic elements play a primary role in understanding and comprehending an oral text, reflecting the corresponding punctuation marks. At the same time, there is no basis to assert that the oral form of speech has direct correspondences in writing, indicating the absence of a one-to-one correspondence between intonation and punctuation, but suggesting the existence of an organic connection between intonational and punctuation methods of speech segmentation.

The object of the study is the poem 'Batyr-Bayan' by Magzhan Zhumabaev.

The subject of the study is the prosodic characteristics of a poetic text to create a phonetic subcorpus of the Kazakh language

Materials and methods of research.

The methodological basis for this study was the theoretical concepts of Russian and foreign scientists on the problems of studying national corpora of various languages, as well as the works of Kazakh linguists on intonation and prosody of spoken speech [1, 2, 3, 4]. The material for the study was the poem "Batyr-Bayan" by Magzhan Zhumabaev. The methods of experimental phonetics: auditory, prosodic, phonological, and auditory analysis methods were used in this research [5, p.54-55]. As a result of auditory analysis, intonational transcription is used with conditional markings of prosodic elements, which must meet the following requirements: all notations should be semantically relevant, and the number of symbols should be minimal to avoid cluttering the reader's attention. A special program (PRAAT) was used in describing the poetic subcorpus, compiling according to a prosodic marking algorithm, which consisted of dividing

50 № 1(146)/2024

Л.Н. Гумилев атындағы ЕҰУ Хабаршысы. Филология сериясы
ISSN: 2616-678X, eISSN: 2663-1288

the voiced text into phrases, syntagmas, and rhythmic groups ('bunak' – 'бунақ'). A pause for separating phrases is indicated by three vertical lines (///). A pause for separating syntagmas is indicated by two vertical lines (//). A break in the melodic contour for separating rhythmic groups is represented by a single vertical line (/). The number of verses, lines, and syllables per line were determined for a comprehensive description of the poetic subcorpus. Punctuation marks served as indicators for segmenting expressions into syntagmas. Two intonemes were chosen to describe the melody, they are – the intoneme of completion (↓) and the intoneme of non-completion (↑), which can convey punctuation marks such as comma, full stop, semicolon, colon, dash, question mark, and exclamation mark.

Literature review.

To carry out prosodic annotation, it is necessary to understand the basic characteristics of Kazakh poetic texts. The features and regularities of versification depend on the properties and structure of each language. In Russian and other stress-based languages, significant importance is given to stress. V.V. Radlov notes that in Turkic languages, the function of stress is performed by synharmonism: "The task of prominently accentuated stress in Indo-European languages, in Turkic languages falls on the harmony of vowels, while in Indo-European languages, word stresses are a morphological moment, in Ural-Altaic languages, it is only a means of euphony, that is, it serves only to give alternation to a monotonously sounding series of syllables, pleasing to the ear, and more sharply denote the connection of agglutinating words." [6, p.22-24].

Kazakh versification by its nature belongs to the syllabic system of versification. Regarding this, Z. Akhmetov writes, who specifically dealt with versification issues: "In Kazakh poetry, the measure of a line is the buyn (буын-syllable), that means, it is measured by the number of syllables in the line. Such a system of syllabic verse based on the number of syllables is used not only in Kazakh poetry but also in the languages of all Turkic-speaking people such as Kyrgyz, Uzbek, Tatar, Turkmen, Azerbaijan, Bashkir, and also in the languages of people like French, Czech, Italian and others." [7, p.205].

The peculiarities of Kazakh versification, which differ in essence from Russian and other European languages, are rooted in the prosodic system with its synharmonism and unexpressed word stress. In the Kazakh language, as well as in French, words and syllables form a single phonetic word with a unified rhythmic stress. L.V. Shcherba notes, who comprehensively studied the phonetic system of the French language: "A serious difference between French and Russian, German, English stress is that the difference between stressed and unstressed syllables in these latter languages is striking, but in French it is hardly noticeable: all syllables seem to be approximately equally stressed to the inattentive observer. "In French, we cannot talk about word stress in any way but must talk only about phrasal stress, which we will subsequently refer to as rhythmic stress" [8, p. 169]. This statement of the scientist is entirely and fully suitable for the Kazakh language. In Kazakh, like in other Turkic languages, the prosodic dominant is synharmonism, not word stress. Therefore, in Kazakh versification, a certain number of syllables in a line is a necessary condition for the existence of rhythm.

A. Baitursynov also drew attention to the regularities that organize the internal structure of verses. "The poetic structure is called 'aishyq' (айшық). Each segment of 'aishyq' is called 'shumak' (шумақ). In each 'shumak,' there are several 'tarmaq,' and 'tarmaq' (тармақ) means the path of each poetic line. Within 'tarmaq,' there are several 'bunaq,'" [9, p. 192]. According to A. Baitursynov, "tarmaq" refers to a line, "shumak" corresponds to a stanza of a song, and "bunaq" is something other than a rhythmic group. "Tarmaq" represents a semantic and intonational whole, corresponding to the universal linguistic unit – a syntagma.

Research results and discussion. Magzhan Zhumabayev's poem "Batyr-Bayan" was chosen as a poetic subcorpus for analysis. The following data are indicated in the meta-marking.

ВЕСТНИК Евразийского национального университета имени Л.Н. Гумилева. Серия Филология
BULLETIN of L.N. Gumilyov Eurasian National University. PHILOLOGY Series

№ 1(146)/2024

51

Author	Magzhan Zhumabayev
Gender	man
Year of birth	1893
Place of birth	North Kazakhstan region
Read by	Baigali Yessenaliev
Title of the text	Batyr Bayan
First publication of the text	1921
Source of text	Magzhan: poems / M. Zhumabayev. Nursultan: Foliant, 2019.
Genre of the text	poem
Form of the text	Written Oral
Graphics	Arab
Style	Literary fiction
Title of the text	"Otan" (Motherland)
Type of distribution	Book
Number of stanzas	89 stanzas
Number of lines	10 lines
Number of rhythmic groups	3 rhythmic groups
The number of syllables	11 syllables
Types of rhyme	1) AABA rhyme scheme (қара өлең ұйқасы)
Age of the audience	Youth literature (18-34 years old) Neutral
Subcorpus	Poetic texts
Number of used words	4500
Source of the text	Internet resource
Marker	Z.Bazarbayeva
Time of entry into the corpus	22.10.2023y.

A syntagma sometimes corresponds to a rhythmic group, but most often in large poetic forms, it consists of several "bunaqs" (rhythmic groups). The relative independence of the line "tarmaq" is achieved through its intonational pattern, where it corresponds to specific intonemes that combine not only phonetic-phonological properties but also semantic-syntactic ones. Realizing in various intoneme variations, the line carries a certain semantic meaning depending on the context and situation: completion, non-completion, question, urging, etc. The independence of a line and its boundaries, is determined not only by changes in the direction of movement of the main tone, the melodic curve, but also by a prosodic pause and a break in the melodic contour at the junction of lines.

The intonational nature of poetic speech finds its expression in poetic rhythm. Rhythm as a component of intonation gives it a certain coloring. The basic unit of poetic rhythm is the line (tarmaq). Kazakh verse is defined as syllabic, with the same number of syllables in a poetic line. Maintaining a certain number of syllables in each poetic line is a necessary condition for the implementation of rhythm. Within a long line, it is possible to identify intonation units – rhythmic groups and syntagmas.

One of the most common forms of verse is an eleven-syllable verse, which is found in larger poetic compositions. An example of an eleven-syllable, multi-verse poetic work is Magzhan Zhumabayev's poem "Batyr Bayan":

Жүрегім, мен зарлымын жаралыға,
Сұм өмір абақты ғой саналыға.

52 № 1(146)/2024

Л.Н. Гумилев атындағы ЕҰУ Хабаршысы. Филология сериясы
ISSN: 2616-678X, eISSN: 2663-1288

Қызыл тіл, қолым емес, кісеңдеулі,
Сондықтан жаным күйіп жанады да.
Қу өмір қызығы жоқ қажытқан соң,
Толғанып қарауым сол баяғыға.
Түйіннің тоқсан түрлі шешуі бар
Әдемі ертегідей баяғыда.
Әдемі өткенді ойлап айнымасам,
Сұм өмір күшті уын аяды ма?

жүрөгүм↑// мен зарлымын жаралыға↑//
сұм өмүр / абақты ғой / саналыға↓
қызыл тіл↑// қолұм емес↑// кісендеулү↑//
сондұқтан / жаным г̃үйүб̃ / жанады да↓
құ'өмүр / қызығы жоқ / қажытқан соң↑//
толғанып / қарауұм сол / байағыға↓
түйүннүң / тоқ̃сан түрлү / шешүүү б̃ар↑
әдем'ертегідей / байағыда↓
әдем'өткөндү / ойлаб̃ а̃йнымасам↑//
сұм өмүр / гүштү ұуұн / айады ма↑

This is how the first ten-line of the eight-nine verse poem begins, in which throughout the entire work there is a stable rhythmic structure with a clear end rhyme, manifested in syllabic and sound repetitions. To confirm this, here are two more verses:

Ертегі уатпай ма баланы да,
Сөз сиқыр ғой, жазбай ма жараны да?
Ақын да бір бала ғой айға ұмтылған,
Еркімен өзі-ақ отқа барады да.
Жай тақтақ жабайыдан жол қалғанда,
Қанды ор боп ақын жолы қалады да.
Ойлайды, күңіренеді, толғайды ақын,
Күрсініп көзіне жас алады да.
Ақында адамзаттан дос болмайды,
Жалғыз-ақ сырын сөйлер қаламына.

ертегі / ұуатпай ма / балады ма//↑
сөс сыйқыр ғой↑// жазбай ма / жараны да↓
ақын да / бір бала ғой / айғ'ұмтұлған ↑//
еркімен / өз'ақ- отқа / б̃арады да↓
жай тақтағ / жабайыдан / жол ғалғанда↑//
қанд'ор боб̃ / ақын жолұ / қалады да↓
ойлайды↑// күңүрөнеди↑// толғайд'ақын ↑//
күрсүнүп / көзүнө жас / алады да↓
ақында / адамзаттан / дос п̃олмайды↑//
жалғыз-ақ / сырын сөйлөр / ғаламына↓

Мен де ойды ағытамын қаламыма,
Арқаның көз жіберсем алабына:
Сарыарқа – сары дария, қиыры жоқ,
Көз болсын қандай қыран, талады да.
Ішінде сары дария көз тоқтатар
Көкшетау – Сарыарқаның аралы да.

Көкшеде күні кеше қойдай өрген
Түрлі аң: бөрі, бұғы, маралы да.
Айрылып асау, ерке аңдарынан,
Көкшенің тас жүрегі жаралы да!

мен д'ойдұ / ағытамың / ғаламыма↑//
арқаның г͡өж / жіб͡ерсем алабына ↓
сарыарқа / сары дарыйа↑// қыйыры жоқ↑//
көз б͡олсұң / ғандай ғыран↑// талады да↓
ішінде / сары дарыйа / г͡өз тоқтатар↑
көк̊шөтау / сарыарқаның аралы да↓
көк̊шөдө / күнү кеше / қойдай өргөн↑
түрл'аң бөрү↑// бұғұ↑// маралы да↓
айрылыб͡ а͡сау↑// ерк'аңдарынан ↑//
көк̊шөнүң / таш шүрөгү / жаралы да↓

Each line of the poem corresponds to a syntagma consisting of three rhythmic groups. Odd-numbered syntagmas, as a rule, are characterized by an ascending direction of movement of the main tone, while even-numbered ones have descending melodic contours. At the junctions of syntagmas, in addition to a slight pause, a break is noted in the melodic contour, which is formed by positive and negative tone intervals. Regardless of the melodic form, syntagmas are characterized by identical tonal ranges and realized in average tonal levels. Syntagmas are usually separated by pauses in speech. Pauses take part in the transmission of certain syntactic and semantic relationships. In natural speech, there are grammatical pauses that separate the intonation-formatted parts of a phrase and pauses of hesitation (uncertainty) from each other.

The boundary of a syntagma can be marked not only by a physical break in the speech signal, but also by a sharp change in pitch and (or) other prosodic characteristics, which are perceived as a violation of the smooth flow of speech. It is important to note that the process of syntagmatic segmentation must satisfy the solution of two main tasks: to establish the boundaries of syntagmas in those places where they must be present, and not to establish the boundary of a syntagma where it can disrupt the semantic perception of speech. The final syllable of a line is accentuated by a syntagmatic stress through tonal, temporal, and dynamic parameters. Each verse (shumak), consisting of two paired syntagmas, represents an ascending-descending homogeneous melodic contour, repeated throughout the entire poetic work.

The prosodic marking of poetic subcorpora indicates that prosodic elements serve as a means of expressing punctuation marks in oral speech. Considering that prosodic means are closely related to semantics and syntax, one can infer their significant role in conveying punctuation marks in oral speech. It is essential to bear in mind that the written form of speech does not have direct equivalents in the oral form. The primary purpose of punctuation marks is to delimit sentences and their parts in the text and establish semantic relationships between them.

The punctuation used in written speech interacts with the rhythmic segmentation of poetic text in oral speech; in other words, the segmentation of oral speech through prosodic means is reflected in writing through punctuation marks. Additionally, punctuation marks in written speech, as a rule, correlate with the *intonemes* of the Kazakh language.

In addition to the grammar of a language, there is a distinct grammar of speech with its own units and specific rules of functioning. Therefore, grammatical and intonational means often do not coincide, which complicates the prosodic analysis of oral texts. The work related to the interpretation of spoken language is exceptionally labor-intensive, but thanks to special programs it is feasible.

Thus, modern punctuation in the Kazakh language, phonetic in its foundation, serves as a means of delimiting the text and expresses semantic-syntactic and intonation-meaningful segmentation in written speech. In some cases, semantic-syntactic indicators coincide with

54   № 1(146)/2024

Л.Н. Гумилев атындағы ЕҰУ Хабаршысы. Филология сериясы
ISSN: 2616-678X, eISSN: 2663-1288

prosodic and punctuation means, while in other cases the only means of expressing syntactic meaning becomes the rhythmic-melodic organization of the voiced poetic text.

Conclusion. The prosodic marking of the poetic subcorpus was carried out on the basis of an the auditory analysis of the poem. The analysis and prosodic marking of the poetic text occurred in several stages. At the first stage, the text was segmented into syntagmas and the designation of intonemes using a special automated program. At the second stage, rhythmic groups ('bunaq' – 'бунақ') were indicated for each line of the verses. At the third stage, the orthoepic version of the poetic text was presented.

Poetic speech in Kazakh language, as in other Turkic languages, is based on the syllabic structure, where a stable number of syllables in a line is a necessary condition for the existence of rhythm, closely related to the prosodic system of the language. Rhythmic groups (bunaq), uniting with each other through the same combination of syllables, form a line (tarmaq), which as a rule, corresponds to a semantic-intonation and syntactic unit – syntagma. Punctuation marks, as well as changes in prosodic speech characteristics, serve as markers for the boundaries of syntagmas and intonemes. The combination of lines (tarmaq) were combined into verses (shumaq), on which the tonal, dynamic and temporal components of intonation were realized in interconnection and mutual influence. Tonality parameters of prosodic annotation were used for the analysis of this poetic subcorpus as an indication of the direction of the main tone movement.

The formation of sound subcorpora and their multi-level marking have become possible recently due to the development of information technologies in the humanities. Thus, it should be noted that electronic corpora play a major role in the development of applied research, serving as their main conductor of innovative technologies and a means of accelerated modernization.

The article was written in the scope of the project "Improving the National Corpus of the Kazakh Language as a Tool of Cultural Communication and Expanding Its Subcorpus"

Contribution of the authors. *Zeinep Bazarbayeva* – creating a concept of the work, conducting experimental research, using methods of research, *Akmaral Srailova* – a selection of materials and work with literature, analysis of research results, *Zhanar Zhumabayeva* – analysis and synthesis of research results and work with manuscript text

References

1. Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – СПб., 2005. – 48 с.

2. Плунгян В.А. (ред.). Национальный корпус русского языка: 2003-2005. - М., 2005. – 150 с.

3. Жубанов А.К. Национальный корпус казахского языка и проблемы метаразметки. Tiltanym. 2015;(1):21-29. – URL: https://www.tiltanym.kz/jour/article/view/312/145?locale=ru_(20.12.2023).

4. Жанабекова А.Ә., Пірманова К., Барменқұлова А. Ұлттық корпустарға негізделген лингвистикалық зерттеулер жүргізу (қазақ, орыс, ағылшын тілі материалдары негізінде // ҚазҰУ Хабаршысы. Филология сериясы. – 2022. - № 3. – Т. 187.

5. Базарбаева З.М. Қазақ тілі: интонология, фонология. - Алматы, 2008. - 326 б.

6. Radloff W. Phonetik der nordlichen Turksprachen. - Leipzig, 1883.

7. Ахметов З. Казахское стихосложение. - Алма-Ата, 1964.

8. Щерба Л.В. Фонетика французского языка. - М., 1948.

9. Байтұрсынов А. Шығармалары: өлеңдер, аудармалар, зерттеулер.   Алматы, 1989.

References

1. Zakharov V.P. Korpusnaya lingvistika [Corpus linguistics], Educational and methodological manual (Saint Petersburg, 2005, 48 p.). [in Russian]

2. Plungyan V.A. (edit.). Natsionalnyi korpus russkogo yazyka. 2003-2005 [National Corpus of the Russian Language. 2003-2005) (Moscow, 2005, 150 p.). [in Russian]

3. Zhubanov A.K. Natsionalnyi korpus kazakhskogo yazyka I problem metarazmetki [National corpus of the Kazakh language and problems of meta-marking]. Tiltanym. 2015; (1):21-29). Available at: https://www.tiltanym.kz/jour/article/view/312/145?locale=ru_(20.12.2023).
 [in Russian]

4. Zhanabekova A.A., Pirmanova K., Barmenkulova A. Ulttyk korpustarga negizdelgen lingvistikalyk zertteuler zhurgizu (kazakh, orys, agylshyn tili materialdary negizinde) [Conducting linguistic research based on national corpora (based on Kazakh, Russian, English language materials)], KazUU Habarshysy. Filologija serijasy [Bulletin of KazNU. Philology series], 3 (187) (2022) [in Kazakh]

5. Bazarbayeva Z.M. Kazakh tili: intonologiya, phonologiya [Kazakh language: intonology, phonology] (Almaty, 2008, 326). [in Kazakh]

6. Radloff W. Phonetik der nordlichen Turksprachen (Leipzig, 1883).

7. Akhmetov Z. Kazakhskoe stihoslozhenie [Kazakh versification] (Almaty, 1964). [in Russian]

8. Shcherba L.V. Phonetika phrantsuzskogo yazika [Phonetics of the French language] (Moscow, 1948). [in Russian]

9. Baitursynov A. Shygarmalary: olender, audarmalar, zertteuler [Works: poems, translations, research] (Almaty, 1989). [in Kazakh]

З.М. Базарбаева¹, А.Б. Сраилова², Ж.Т.Жұмабаева¹
¹Ахмет Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Казахстан
²Абылай хан атындағы Қазақ халықаралық қатынастар және әлем тілдері университеті, Алматы, Қазақстан

Поэтикалық ішкорпус: просодикалық белгіленім

Аңдатпа. Мақала қазақ тіліндегі поэтикалық мәтіндердің просодикалық белгіленуін сипаттауға арналған. Мәтіннің просодикалық белгіленуі синтагмаларға, синтагмаларды ритмикалық топтарға (бунақтарға) бөлуден және белгілі бір ережелерге сәйкес синтагмалардың интонациялық түрін анықтаудан тұрады. Синтагма интонациялық мазмұны бар фразаның бір бөлігі немесе жалпы фразаның өзі болып табылады.

Синтагманың арасын белгілеу семантикалық мазмұн мен интонема түрінің берілуіне әсер етеді. Сондықтан мәтінді синтагмаларға бөлу кезінде оның мағынасына қарап, мазмұны жағынан дұрыс қабылдануын ескеру қажет. Сөйлеу үстінде синтагмалардың шекарасын белгілеу үшін пунктуациялық, фонетикалық талдауға негізделген синтагманы ажыратудың белгілі бір ережелері қолданылады.

Сонымен, просодикалық талдау басқа түркі тілдеріндегідей қазақ тілінде де поэтикалық сөйлеудің буындық құрылымға негізделгенін, бір қатардағы буын санының тұрақты болуы ырғақ болуының қажетті шарты екенін көрсетті. Поэтикалық субкорпораның просодикалық таңбалануы просодикалық элементтердің ауызша сөйлеуде тыныс белгілерін білдіру құралы екенін көрсетеді.

Түйін сөздер: поэтикалық ішкорпус, синтагма, просодикалық белгілеу, ауызша сөз, корпустық лингвистика, аннотациялау, метабелгіленім.

З.М. Базарбаева¹, А.Б. Сраилова², Ж.Т.Жумабаева¹
¹Институт языкознания имени Ахмета Байтурсынова, Алматы, Казахстан
²Казахский университет международных отношений и мировых языков имени Абылай хана, Алматы, Казахстан

Поэтический подкорпус: просодическая разметка

Аннотация. Настоящая статья посвящена описанию просодической разметки поэтических текстов в казахском языке. Просодическая разметка текста заключается в его членении на синтагмы, разметке синтагм на ритмические группы (бунак) и маркировке интонационного типа синтагм в соответствии с определёнными правилами. Под синтагмой понимается самостоятельная в интонационном смысле часть фразы или вся фраза.

Установка границ синтагм влияет на передачу смыслового содержания и типа интонемы. При разбиении текста на синтагмы важно не ставить границу синтагмы там, где она может нарушить смысловое восприятие речи. *Для установки границ синтагм в естественной речи используются определённые правила синтагматического членения, базирующиеся на пунктуационном, фонетическом анализе текста.*

Таким образом, просодический анализ показал, что поэтическая речь в казахском языке, как и в других тюркских языках, основана на силлабическом строе, когда стабильное количество слогов в строке является необходимым условием существования ритма. Просодическая разметка поэтических подкорпусов свидетельствует о том, что просодические элементы являются средством выражения знаков препинания в устной речи.

Ключевые слова: поэтический подкорпус, синтагма, просодическая разметка, звучащая речь, корпусная *лингвистика*, аннотирование, метаразметка.

Information about authors:

Bazarbayeva Z.M. – Doctor of Philological Sciences, Professor, Academician of National Academy of Science Republic of Kazakhstan, Institute of linguistics named after A. Baitursynov, Almaty, Kazakhstan.

Srailova A. B. – Master of philological sciences, Senior lecturer, Kazakh Ablai Khan University of International Relations and World Languages, Almaty, Kazakhstan.

Zhumabayeva Z.T. – Candidate of Philological Sciences, Leading Researcher at the Institute of Linguistics named after A. Baitursynov, Almaty, Kazakhstan.

Базарбаева З.М. – ф.ғ.д., профессор, ҚР ҰҒА академигі, А. Байтұрсынов атындағы Тіл білімі институтының бас ғылыми қызметкері, Алматы, Қазақстан.

Сраилова А.Б. – ф.ғ.м., аға оқытушы, Абылай хан атындағы Қазақ халықаралық қатынастар және әлем тілдері университеті, Алматы, Қазақстан.

Жұмабаева Ж.Т. – ф.ғ.к., А. Байтұрсынов атындағы Тіл білімі институтының жетекші ғылыми қызметкері, Алматы, Қазақстан.