UDC 81 https://doi.org/10.33619/2414-2948/58/46

MODELING OF MORPHOLOGICAL ANALYSIS AND SYNTHESIS OF WORD FORMS OF THE NATURAL LANGUAGE

©Kochkonbaeva B., *Osh Technological University named by M.M. Adyshev, Osh, Kyrgyzstan, buajar@mail.ru*
©Egemberdieva Zh., *Kyrgyz-Uzbek University, Osh, Kyrgyzstan, Egemberdieva8787@mail.ru*

МОДЕЛИРОВАНИЕ МОРФОЛОГИЧЕСКОГО АНАЛИЗА И СИНТЕЗА СЛОВОФОРМ ЕСТЕСТВЕННОГО ЯЗЫКА

©Кочконбаева Б. О., *Ошский технологический университет им. М. М. Адышева, г. Ош, Кыргызстан, buajar@mail.ru*
©Эгембердиева Ж. С., *Киргизско-Узбекский университет, г. Ош, Кыргызстан, Egemberdieva8787@mail.ru*

Abstract. This article discusses the modeling of morphological analysis and synthesis of the Kyrgyz language. Models of creating word forms and rules for connecting affixes to the base of a word are considered. As a result of studying the structure of natural language, an algorithm of module was created that performs automatic morphological analysis and synthesis of word forms on a personal computer.

Аннотация. В статье рассматривается моделирование морфологического анализа и синтеза кыргызского языка. Рассмотрены модели создания словоформ и правила подключения аффиксов к базе слова. В результате изучения структуры естественного языка был создан алгоритм модуля, который выполняет автоматический морфологический анализ и синтез словоформ на персональном компьютере.

Keywords: affixes, natural language, roots, morphological analysis.

Ключевые слова: аффиксы, естественный язык, основа, морфологический анализ.

At the present stage of development of science, technology and culture, preference is given to information processing processes, which occupy leading positions in the process of social production and penetrate into all spheres of human activity. Methods and means of processing information in a natural language are becoming increasingly important — from simple document preparation systems to information retrieval systems, machine translation systems, and natural language communication programs.

As a result of studying the structure of natural language, we created a system module that performs automatic morphological analysis and synthesis on a personal computer.

Modeling of word formation in the Kyrgyz language

In view of the fact that the words of the Kyrgyz language consist of a root and affixes, we denote the word as S, then as a function we define them as follows:

$$S = U + Km + Um, \quad (1)$$

here, S is a linear function.
U — the root of the word
Km — word-forming affixes
Um — inflectional affixes.

Т. 6. №9. 2020
https://doi.org/10.33619/2414-2948/58

According to the formula (1), *S* depends on the root, word-forming affixes, and inflectional affixes.

Inflectional affixes can reach up to eight, in other words

$$Um = Um_1 + Um_2 + \ldots + Um_8 \ (2)$$

Rule 1: If $Km=\varnothing$, $Um=\varnothing$, then the *S* function will be equal to the root of the word, and the entered word will not be divided into morphemes.

Inflectional affixes

Inflectional affixes change the grammatical meaning of words, but do not change the lexical meaning [1, p. 13–37].

By grouping the above morphological categories into sets of affixes, we get the following list:

J = {-нын, -га, -ны, -да, -дан} set of case affixes (Noun Cases).

T = {-ым, -ың, -ыңыз, -сы, -ы, -быз, -ңар, -ңыздар} set of possessive affixes (Possessive).

K = {-лар} set of plural affixes *(Pl)*.

Zh = {-мын, -быз, -сың, -сыңар, -сыз, -сыздар} many of the affixes of the person (Personal).

Ch = {-ды, -ган, -ыптыр, -чу, ...} set of time affixes (Verb Tenses).

In = {-са, -гай, ...} set of mood affixes (Imperatives).

Neg = {ба} set of negative affixes (negative aspect of the Verb Tenses category).

Q = {бы} set of affixes of interrogative meaning (aspect of the interrogative category Verb Tenses).

If we say that *Um* is a set of inflectional affixes, then It consists of the following parts:

Um = {J, T, K, Zh, Ch, In, Neg, Q}.

Rules for connecting affixes

Nominative words are called nouns, numerals, adjectives, pronouns [2].

Rule 2: If $U \in (Z \vee C \vee San \vee At)$, then as shown in (1) the formula $U + Um$, $U + Um + Km$ sum does not hold, in other words, after inflectional affixes, the word-forming affixes do not connect.

The rules for inflectional affixes are also preserved:

$$Um = K + T + J + Zh + Q \ (3)$$

On the basis of formula (3) $U \in (Z \vee C \vee San \vee At)$ for time (1) can be written as:

$$S = U + Km + K + T + J + Zh + Q, \ (4)$$

In this formula, some elements of the plural of inflectional affixes can be equal to arbitrary affixes.

For example:

S = 'аталар', U = 'ата', Km = ∅, Um = K = 'лар'.

S = 'аталарыбыз', U = 'ата', Km = ∅, Um = K + T = 'лар'+'ыбыз'.

S = 'аталарыбыздын', U = 'ата', Km = ∅, Um = K + T + J = 'лар' + 'ыбыз' + 'дын'.

S = 'аталарсыңар', U = 'ата', Km = ∅, Um = K + T + Zh = 'лар' + 'сыңар'.

S = 'аталарыбызсыңар', U = 'ата', Km = ∅, Um = K + T + J + Zh = 'лар' + 'ыбыз' + 'сыңар'.

S = 'аталарыбызсыңарбы', U = 'ата', Km = ∅, Um = K + T + J + Zh + Q = 'лар' + 'ыбыз' + 'сыңар' + 'бы'.

Rules for placing vocabulary affixes of verb words

Since verb words are abstract concepts and differ from nominal words based on their presence in the grammatical category.

There are also two rules for verb words, and $U + Um_1 + Um_2 + \ldots$

Т. 6. №9. 2020
https://doi.org/10.33619/2414-2948/58

For example: in the words *созгула-* (hold out -), *кирин-* (swim-), *сүйүн-* (rejoice-) word-forming affixes (*-гыла, -ын*) entered the word basis (roots).

So, in verbal words, the location of word-forming affixes is carried out in the following order. Base (or root), affixes of relation, negative meaning affixes, time affixes (or time affixes + auxiliary verb), time and mood indicators, person and number indicators, request and question meaning affixes.

Modeling of morphological analysis and word synthesis

The morphological development of natural language texts can be viewed from a human and computer perspective.

If for us the morphology of a word is the root of a word, affixes, parts of speech, then speaking in computer language, in other words, an automated system working on a natural language, we can get various information about different degrees of structure of this language.

We can consider 3 stages of morphological analysis:

1. Defining only the grammatical meaning of a word.
2. Defining only the basic word.
3. Determining the grammatical meaning and basis of the word.

A detailed or incomplete study of morphological analysis depends on the task at hand.

Morphological analysis is the initial stage of various tasks related to natural language, and therefore, its precise implementation is of great importance.

Morphological analysis methods can be divided into 3 types:

–to analyze with a dictionary of affixes.

–to analyze with the help of a dictionary of affixes and bases.

–to analyze using a dictionary of the word system.

In the method of analysis using the dictionary of affixes [3, p. 424–429], we consider the selection of affixes from a word and search through the dictionary, and on this basis, reveal the grammatical meaning of the word. Only the grammatical meaning is selected as the result of this type of analysis. In the course of morphological analysis, the dictionary of affixes and the research method in the text of natural language are not particularly used. This is due to the fact that using only the dictionary of affixes, it is impossible to conduct morphological synthesis.

In the analysis method, which consists of a dictionary of affixes and bases, the base and affixes of the word are selected and searched through the dictionary, based on this, the grammatical and semantic meaning of the word is taken.

Only the basis [5, p. 119–133] or grammatical meaning is considered to be the type of result of such morphological analysis.

The negative side of this method is considered when adding affixes to some bases, because of the drop of letters, the basis turns out to be incomplete.

For example, let's consider the word *балдар*, the basis of *бала* + *дар*, it would be wrong to analyze the word *бал*, because the root of this word is *бала*.

The third method of morphological analysis is considered to be the use of a dictionary of the word system [4, p. 98–101]. Here you need to find the form of the desired word and return the corresponding grammatical meaning. With this method of analysis, you have to work with a very large vocabulary. This method of analysis is performed for inflectional languages, not for agglutinative languages. In other words, it is not suggested to use this method in Kyrgyz language texts.

Among the above methods of analysis, performing morphological analysis of natural language texts using the dictionary of basics and affixes was considered in the Embercadero RAD

Studio programming environment. A dictionary made up of basics includes a class of basics and their various forms. This is applied in order to eliminate the above-mentioned shortcomings.

We said that in the Kyrgyz language affixes are divided into word-forming and inflectional affixes. Word-forming affixes form a word with a new meaning. Therefore, at the first stage, we should not refer to the root meaning, but consider it correct to conduct a morphological analysis in accordance with its meaning in the text. For example, if the word "*тарбиячылар*" occurs in the text, it is considered as the word *тарбиячы* + *лар*, *тарбиячы* is a noun without a person, and *лар* is analyzed as an affix defining the plural. And if we consider from the point of view of word formation, then *тарбия* is a noun, without a person, *чы* is a nominative word – forming affix, *лар* is analyzed as an affix that defines the plural. When compiling the program, due to the fact that many letters of the Kyrgyz language differ from the letters of the Russian language, the UTF8 format was used.

Program testing

A test program was created based on the algorithm in the Embarcadero RAD Studio programming environment. The system interface is shown in Figure.

The morphological analyzer can be used in all programs that work with text for initial analysis.

These programs can be classified as follows:

–anti-plagiarism systems.

–machine translation systems.

–information retrieval systems.

–question and answer systems.

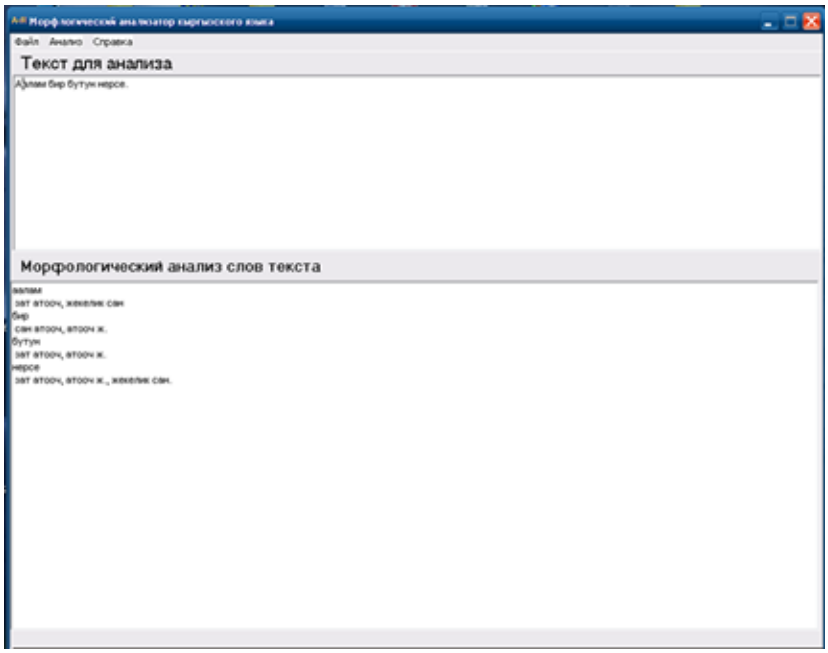

Figure. Testing process.

Conclusion

We have considered the example of the Kyrgyz language morphological analyzer algorithms and methods of analysis.

As we have seen, the Kyrgyz language has several features:

–Like all Turkic languages, Kyrgyz is one of the agglutinative languages.

–Words of the Kyrgyz language are formed using word-forming and inflectional affixes.

–Word-forming affixes are never behind inflectional affixes.

–The law of synharmonism is fulfilled.

–To the appropriate extent, inflectional affixes are subordinated to the appropriate order.

Taking these features into account, the initial morphological analyzer system was developed in the sphere of Embarcadero RAD Studio. This system was implemented using methods of working with the dictionary of basics and affixes.

The system also calculates the weight of words. As a result, when developing a morphological analyzer program, the features of each language are taken into account.

But, in General, since the methods of word formation in the Turkic languages are the same, changing the vocabulary base, you can apply the compiled system of dictionaries for other languages as well.

References:

1. Abduvaliev, I., & Sadykov, T. (1997). Sovremennyi kyrgyzskii yazyk. Bishkek. (in Russian).

2. Batmanov, I. A. (1936). Chasti rechi v kyrgyzskom yazyke. Frunze. (in Russian).

3. Nozhov, I. M. (2000). Prikladnoi morfologicheskii analiz bez slovarya. *In KII-2000: Trudy konferentsii. Moscow. V. 1. 424-429*. (in Russian).

4. Polyakov, V. N. 1996. Programma “Podlesok”: kognitivnyi podkhod k izucheniyu prirody smyslovykh svyazei estestvennogo yazyka. *In Konferentsiya po iskusstvennomu intellektu KII-96. Kazan*, *98-101*. (in Russian).

5. Prutskov A. V., & Rozanov A. K. (2014). Metody morfologicheskoi obrabotki tekstov. *Prikaspiiskii zhurnal: upravlenie i vysokie tekhnologii,* (3), 119-133. (in Russian).

Список литературы:

1. Абдувалиев И., Садыков Т. Современный кыргызский язык. Бишкек, 1997.

2. Батманов И. А. Части речи в кыргызском языке. Фрунзе, 1936.

3. Ножов И. М. Прикладной морфологический анализ без словаря // КИИ-2000: Труды конференции. М.: Физматлит, 2000. Т. 1. С. 424-429.

4. Поляков В. Н. Программа «Подлесок»: когнитивный подход к изучению природы смысловых связей естественного языка // Конференция по искусственному интеллекту КИИ-96. Казань, 1996. С. 98-101.

5. Пруцков А. В., Розанов А. К. Методы морфологической обработки текстов // Прикаспийский журнал: управление и высокие технологии. 2014. №3. С. 119-133.

Работа поступила
в редакцию 03.08.2020 г.

Принята к публикации
07.08.2020 г.

Ссылка для цитирования:

Kochkonbaeva B., Egemberdieva Zh. Modeling of Morphological Analysis and Synthesis of Word Forms of the Natural Language // Бюллетень науки и практики. 2020. Т. 6. №9. С. 435-439. https://doi.org/10.33619/2414-2948/58/46

Cite as (APA):

Kochkonbaeva, B., & Egemberdieva, Zh. (2020). Modeling of Morphological Analysis and Synthesis of Word Forms of the Natural Language. *Bulletin of Science and Practice, 6*(9), 435-439. https://doi.org/10.33619/2414-2948/58/46