МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ КЫРГЫЗСКОЙ РЕСПУБЛИКИ

КЫРГЫЗСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. И. РАЗЗАКОВА

ISSN 1694-5557

ИЗВЕСТИЯ

КЫРГЫЗСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ им. И. РАЗЗАКОВА

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

2016 № 2(38)

Бишкек

Издательский центр «Текник» 2016

Редакционная коллегия:

Журнал выходит ежеквартально.

Все материалы, поступающие в редколлегию журнала, проходят независимое рецензирование.

MINISTRY OF EDUCATION AND SCIENCE OF THE KYRGYZ REPUBLIC

KYRGYZ STATE TECHNICAL UNIVERSITY NAMED AFTER I.RAZZ AKOV

JOURNAL

OF KYRGYZ STATE TECHNICAL UNIVERSITY NAMED AFTER I.RA ZZAKOV

THEORETICAL AND APPLIED SCIENTIFIC TECHNICAL JOURNAL

2016 № 2 (38)

Bishkek

Publishing center "Tehknik" 2016

The journal is published quarterly
All materials that come to the Editorial Board of the journal are subject to independent peer-review

Этот раздел содержит материалы 4-ой Международной конференции по компьютерной обработке тюркских языков "TurkLang 2016"

ИНФОРМАЦИОННЫЕ И ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ И СИСТЕМЫ

This section contains materials of the 4th International conference on Turkic Languages

Processing "TurkLang 2016"

INFORMATION AND TELECOMMUNICATION NETWORKS AND SYSTEMS

ИНФОРМАЦИОННЫЕ И ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ И СИСТЕМЫ

УДК 81.322

МЕТОД ФОРМАЛЬНОГО ОПРЕДЕЛЕНИЯ СМЫСЛА ПРЕДЛОЖЕНИЯ

Азат Абдысадыр уулу, *заведующий сектором канцелярии Аппарата Президента КР, Кыргызстан, 720003, г. Бишкек, пр. Чуй 205, e-mail: azat@adm.gov.kg*

В статье ставится задача определения смысла предложения при помощи системы знаков с индексом пространства, т.е. создания метода формального определения смысла предложения. Задача решается на базе равенства десигната и знака, общего числового семантического поля, а также локального семантического поля. При этом числовые позиции создают композицию с определенным композиционным образом, который является описанием объекта, т.е. смыслом предложения. Данный метод позволяет ставить и решать типовые вопросы создания типологии формального смысла предложения, осуществления формального анализа семантики слова. Метод формального определения смысла предложения является важным инструментом для создания числового семантического поля языка, а также создания формального «образа» значения слова и лексемы.

Ключевые слова: объект, стимул, десигнат, формальные дискретные единицы языка, числовое семантическое поле, сигнатура, композиция.

THE METHOD FORMAL DEFINITION MEANING OF THE SENTENCE

Azat Abdysdyr uulu*, Head of the sector of the Office of the President of the Kyrgyz Republic, Kyrgyzstan, 720003, c. Bishkek, Chui Avenue, 205, e-mail: azat@adm.gov.kg*

In the article to set a task to determine the meaning of the sentence with the help of a system of signs with an index space, i.e., the creation of the method formal definition meaning of a sentence. The task is solved on the basis of equality of sign and designatum, of general numerical semantic space, as well as of local semantic space. Thus numerical position create a certain composition, that is regarded as the object description, i.e. meaning of a sentence. This method allows you to set and solve typical issues of creating a typology of formal meaning of the sentence, of the formal analysis of the semantics of the word. The method formal definition meaning of a sentence is an important tool for creating numerical semantic space of language, as well for creating a formal «image» value of words and lexems.

Keywords: object, stimulus, designatum, formal discrete units of language, numerical semantic space, signature, composition.

I. ВВЕДЕНИЕ

Психическая деятельность человека является идеальным в качестве познавательной деятельности [3]. При этом познавательная деятельность имеет «инструментальный» характер, т.е. высшими функциями человека являются «промежуточные» реакции организма, создающие собственные стимулы организма. Человек модифицирует стимулы [6]. Одна дискретная модификация стимула есть информация, выраженная определенным знаком. Исходя из «инструментального» характера познавательной деятельности, действительность состоит из объектов.

Объект (Y) – это феномен, имеющий признаки, которые познаются в ходе психической деятельности человека (стимулы). Признаки объектов (стимулы) имеют: 1) дискретный характер; 2) обладают свойством отражаемости. Один признак в ходе познавательной деятельности человека воспринимается и передается как один дискретный стимул.

Познавательная деятельность состоит из процессов восприятия, номинации [1] и передачи стимула. При процессе восприятия каждый стимул воспринимается дискретно. К примеру, стимулом может быть звук с определенными свойствами [4]. Множество (или группа) стимулов может восприниматься одновременно. При этом, каждый дискретный стимул отражается по определенному порядку, т.е. происходит порядковое отражение стимулов. Следовательно, номинацию можно представить в виде системы порядковых чисел:

$$\mathrm{N} \ni (1 \ldots \mathrm{n}) \tag{1}$$

При номинации стимул становится десигнатом со формальными признаками: угла (α^0, β^0) и порядка.

При этом, наличие порядковой величины в свое время предполагает наличие индексов T (времени) и $П$ (пространства). Учитывая порядковый показатель десигната, наблюдается равенство десигната (V) и знака (Z).

$$V = Z \tag{2}$$

Для нашего случая показатель $\mathrm{N} \ni (1 \ldots \mathrm{n})$ обладает временным индексом

$$\mathrm{N} \ni 1 \ldots \mathrm{n} \ni T \quad t'(\quad) \quad (\ ) \tag{3}$$

Учитывая индекс T, ставится задача определения смысла предложения при помощи системы знаков с индексом $П$, тем самым цель статьи – создание метода формального определения смысла предложения.

II. РЕШЕНИЕ ЗАДАЧИ

1. Слово и лексема

При равенстве десигната и знака, их соотношение определяется константой номинации (Ω), которая равна к $\sqrt{2}$. Соотношение десигната и знака есть имя (название) стимула (Q):

$$V * \sqrt{2} = Q \tag{4}$$

Слово (S) является знаком [2], производимым от десигната с помощью константы номинации (Ω)

$$(V * \sqrt{2})/\sqrt{2} = S \tag{5}$$

Слово есть парадигматическая единица языка [5]. Омоним слова определяется по следующей формуле:

$$(S * \sqrt{2}) * 3^{p})/\sqrt{2} = S_{op} \quad (6)$$

Где p – порядковое число, $p > 0$;

S_{op} – омоним слова.

Начальная форма имени определяется с помощью Ω

$$S * \sqrt{2} = Q \quad (7)$$

$$S_{op} * \sqrt{2} = Q_{op} \quad (8)$$

Кроме начальной формы имени, которая выражается словом, имеется грамматическая форма имени [6], выражаемая лексемой (l):

$$Q * 2^{n} = q_{n} \quad (9)$$

Где n – порядковое число, $n > 0$;

q – грамматическая форма имени. Тогда с применением Ω:

$$q_{n}/\sqrt{2} = l_{n} \quad (10)$$

На основании (5) и с учетом $V = Z$:

$$(l_{n} * \sqrt{2})/\sqrt{2} = v_{n} \quad (11)$$

Где v_{n} – десигнат с равенством $v_{n} = l_{n}$

$$Q_{\text{op}} * 2^{n} = q_{opn} \quad (12)$$

Где q_{opn} – грамматическая форма имени.

Тогда

$$q_{opn} / \sqrt{2} = l_{opn} \quad (13)$$

Где l_{opn} – лексема для омонима.

$$(l_{opn} * \sqrt{2})/\sqrt{2} = v_{opn} \quad (14)$$

Где v_{opn} – десигнат с равенством $v_{opn} = l_{opn}$

Следовательно, создаются формальные дискретные единицы языка:

$$D_{n} \ni (V, Q, S);\ D_{g} \ni (v, q, l) \quad (15)$$

Где $D_{n} = S$; $D_{g} = l$.

При этом для D_{n} определяются позиции нечетных порядковых чисел, а для D_{g} определяются позиции четных порядковых чисел.

2. Числовое семантическое поле

Формальные дискретные единицы языка могут быть объединены в группы на базе двухмерного евклидова пространства (E^{2}). С учетом параметра номинации $\mathrm{N} \ni (1 \ldots \mathrm{n})$, каждая формальная дискретная единица языка занимает порядковую позицию в E^{2}. Общая группа, т.е. общее количество формальных дискретных единиц языка (M) на базе E^{2} является общей числовой семантической полей языка (Π_{o}).

$$\mathrm{E} \parallel M\left[D_{n};\ D_{g}\right] \rightarrow \Pi_{\mathrm{o}}^{\varphi} \quad (16)$$

При этом максимально большим числовым показателем обладает D_{g}:

$$\mathrm{D_g} \ni (v_{\max}\ q_{\max}\ l_{\max}) \quad (17)$$

Кроме того, возможны локальные семантические поля (Π_{s}^{φ}) на базе E^{2} с определенным количеством формальных дискретных единиц языка (M_{s}).

$$\mathrm{E}^{2} \parallel M_{s}\left[D_{n};\ D_{g}\right] \rightarrow \Pi_{\mathrm{s}}^{\varphi} \quad (18)$$

3. Сигнатуры и композиции

Благодаря t' на базе $П_{о}^{\varphi}$, а также $П_{s}^{\varphi}$ могут сформироваться отдельные группы D_g. Эти группы называются сигнатурами (G). Они имеют определенные композиции (P). Сигнатуры возникают на позициях четных чисел, при этом в сигнатурах не участвуют нулевые позиции.

$$П_{о}^{\varphi} \parallel G \left[D_g \right] \ni P \quad (19)$$

$$П_{s}^{\varphi} \parallel G \left[D_g \right] \ni P \quad (20)$$

Сигнатура выражается предложением (K). Следовательно, основной функцией предложения является феноменологическая функция, т.е. объединение десигнатов в определенную группу – описание объекта. Тогда, учитывая, что сигнатура имеет композицию, образ композиции (W) есть описание объекта. Числовые позиции D_g на базе $П_{о}^{\varphi}$ или $П_{s}^{\varphi}$ создают композицию (P) в виде определенной числовой группы (композиционный образ), которое является описанием объекта, т.е. смыслом предложения $(¥)$. Следовательно, определение W есть формальное описание объекта при помощи его стимулов т.е. определение смысла $(¥)$ предложения .

$$K \Vdots (\quad) \quad (20)$$

Тогда

$$K_1 (U) - K_2 (U) = H \quad (21)$$

Где U – количество D_g в составе K. При $H \neq 0$ – неравнокомпозиционность предложений. Если $H = 0$, тогда

$$\begin{vmatrix} K_1 (D_{g1}) - K_2 (D_{g1}); K_1 (D_{g2}) - K_2 (D_{g2}) \\ K_1 (D_{g3}) - K_2 (D_{g3}); K_1 (D_{g4}) - K_2 (D_{g4}) \ldots \end{vmatrix} = \begin{vmatrix} \tau_1 + \tau_2 \\ \tau_3 + \tau_{4\ldots} \end{vmatrix} = \delta \quad (22)$$

при $\delta = 0$ – равнокомпозиционность предложений;

$\delta \neq 0$ – неравнокомпозиционность предолжений.

III. ВЫВОДЫ

Таким образом, разработан метод формального определения смысла предложения. Данный метод позволяет ставить и решать вопросы: 1) создания типологии формального смысла предложения; 2) осуществления формального анализа семантики слова, при этом является важным инструментом в следующих практических приложениях:

1. Для создания числового семантического поля языка; 2. Для создания формального «образа» значения слова и лексемы.

Список литературы

1. Апресян Ю.Д. Языковая номинация (общие вопросы). М., 1977.
2. Моррис Ч. Значение и означивание//Семиотика М., 1983.
3. Рубинштейн С.Л. Принципы и пути развития психологии. М., 1959.
4. Сыдыков Т., Токтоналиев К. Азыркы кыргыз тили: фонетика жана фонология. Б., 2015.
5. Ф. де Соссюр. Курс общей лингвистики. М., 2006.
6. Luria A.R. The Making of Mind: A. Personal Account of Soviet Psychology M. Cole & S. Cole, eds. Cambridge, 1979.

УДК 811.11+811.512.133:81`322.4

THE BASES OF AUTOMATIC MORPHOLOGICAL ANALYSIS FOR MACHINE TRANSLATION

***Nilufar Abdurakhmonova**, doctoral student, Tashkent State university of Uzbek language and literature named after Alisher Navoi, Tashkent city, abdurahmonova.1987@mail.ru*

The aim of this article is to show how automatic morphological analyzer identifies clarification of the verbs in English and Uzbek languages. Verbs are very complex natured category in both of languages. The linguistic database of given program should not only include pure grammar, but also some morphological algorithms of different languages. English morphology depends on syntactic analyzing in machine translation. That is way the problems of machine translation in inflected and agglutinative languages is often required to be solved in morphological analyze.

Keywords: Uzbek language, automatic morphological analyze, natural language processing, lexicon

ОСНОВЫ АВТОМАТИЧЕСКОГО МОРФОЛОГИЧЕСКОГО АНАЛИЗА ДЛЯ МАШИННОГО ПЕРЕВОДА

***Нилюфар Абдурахманова** докторант, Ташкентский Государственный Университет Узбекского языка и литературы им. Алишера Навой abdurahmonova.1987@mail.ru*

Данная статья рассматривает классифицирование глаголов в английском и узбекском языках с помощью автоматического морфологического анализатора. Глагол в обеих языках является сложной характерной категорией. Кроме того, как утверждает автор, лингвистическая база любой переводческой программы должна включать не только чистой грамматики, но и морфологические алгоритмы разных языков. В машинном переводе морфология английского языка зависит от синтактического анализа. Исходя из данной проблемы, в статье сделана попытка найти решение морфологического анализа машинного перевода флективных и агглютинативных языков.

Ключевые слова: узбекский язык, автоматический морфологический анализ, обработка естественного языка, лексикон, субкатегорическая парадигма

I. Introduction

In Uzbekistan Computational linguistics appeared as the subject at the beginning 2000s, it began to investigate new researches, scientific works. Now machine translation has become very important issue as one of the directions of computational linguistics. It has some problems depending on analysis. First of all, it needs morphological analyzer in translation of English texts into Uzbek. Then it should be subcategorized paradigms parts of speech. In spite of lack of resource formal language of Uzbek, it has rich linguistic database description of so many literatures for that.

Within the process of automatic analyzing of the each word is to be considered morphological surface form of the word as well. We are of the opinion that, the active and passives morphological lexicon is used in translation system. Active morphological form contains the list of word stem and suffixes combinations. Grammatical rules and analysis are input to passive morphological analysis.

This paper presents the first step to lay the foundation for automatic morphological analyzer. Naturally, to input all forms of words is impossible, because there are so many combinations of word forms. That's way that is necessity to study combination word and suffixes of agglutinative

language. Paradigms of part of speech are handful to adapt languages. It should be taken the specificity and general rules of those languages and it needs to be given their formal definition, especially in machine translation for not related languages. That's the reason, some problematic situations between Eastern (agglutinative) languages and European (inflected) languages have created the new and big critical approach in linguistics. Mainly these types of problems are observed in Krivonosov's works [1] concerning syntaxes and translating eastern languages into European languages and vice versa in Marchuk's works [2].

Uzbek language is morphologically rather complicated and rich in inflectional form (form with endings). For example, Noun: bola+jon(1)+lar(2)+im(3)+dagi(4)+lar(5)+ niki(6)+mas(7)+mi(8)+kan(9) (shortened form of ekan) +a(10); **Verb**: o'qi+t(1)+tir(2)+ ma(3)+ gan(4)+lig(5) (changed form of -lik)+im(6)+dan(7)+mas(8)+mi(9)+kan(10)+a(11). As we see, there is long enough chain of suffixes of inflection. Particularly when the text is translated from Uzbek into English, it will be difficult to give all the meaning of the sentence because of diverse structure. The English words are rather transparent as the morphemes are easily segmented and associated with appropriate grammatical meanings. The Uzbek word forms are considerably less transparent. Very often it is impossible to decide unambiguously whether we have morphological formative and stem element: burnim=>burun+im. Morphological analysis lies at the found of all programs of automatic text processing of the Uzbek language. Next we would like to point out a few spheres in which a morphological analyzer is indispensible.

The morphological analyzer is used in various systems for special functions: text editors (spell checker), information retrieval, automatic annotation, speech recognition and machine translation. On the one hand, there are certain linguistic problems. A text body would offer a better opportunity for studying actual language usage, a field still rather poorly cultivated in Uzbek. New prospects are opened up in the studying of a) grammar: the usage of word forms, phrases and collocations; b) lexicography: frequency dictionaries, dictionary of individual styles, authors, dialect, concordance instead of card files, as well as; c) textology and stylistics: grammatical and lexical peculiarities of different text types[1]. The morphology part of every Uzbek grammar are, without exception, synthesis-oriented. They provide rules for the formation of the inflectional forms, but say nearly nothing about the usage those forms. In the actual usage, one word form usually dominates over its parallel forms. Of some words only the singular, or plural, or just a couple of concrete forms are used. Some forms occur only in certain fixed word of large text corpora.

The different strategies underlying morphological analyses are based on the following properties of morphological units and their relationships [3]:

-integrity of word forms
-segmental structure of word forms
-variability of units
Regularity and irregularity of relation

In order to analyze a word form it is first necessary to segment it into units which will then have to be **transformed** into the shape of their initial forms to be, in turn, searched for in dictionaries. The result of the analysis is generated from information attached to the initial forms. In morphological analysis it is not possible to consider unit variation on the level of an individual unit, instead, the word form must be treated as a member of a paradigm. The paradigmatic approach serves as basis for the model of classificatory morphology. The variability of the stem appears within the paradigm, i.e. it is revealed if we compare different inflectional forms of the same word. The variability of the formatives, in the contrary, is revealed inter paradigmatically, i.e. if we compare one and the same inflectional form across different words.

There are about 207 types suffixes (including variation) of parts of speech in Uzbek language and 130 of them are defined as verbs. In order to add endings to the bases of each words it needs to

separate one or another part of speech into paradigms. We separated the verb into following paradigms:

1. According to the features of adding voice endings:

1.1. Causative voice of verbs:

V_1:-ar is added only two verbs=>V: chiq+ar, qayt+ar

V_2:-giz{-g‘iz}is added verbs that is ended voiced consonant =>yur+giz,tur+g‘iz

V_3:-dir {-tir} is added to verbs are ending with vowel and voiced consonant=>ye+dir, yoy+dir

V_4:-ir is added to verbs are ending with **t, ch, sh** consonants=>ich+ir, shosh+ir, tush+ir

V_5:-iz is added to verbs are ending **q, m** consonants=>oq+iz, tom+iz, em+iz

V_6:**-t/it** –is added ending vowels of two or many-syllabled words: ishla+t, tuga+t, boshla+t, o‘qi+t

The causative voice ending in Uzbek language looks like into the following grammatical form in English language. **Have** / **Get** something V_{III}=>Uzbek verbs vocabulary similar to the above mentioned groups V1, V2, V3 are entered into the linguistic database. For example, I have my lesson done –Men darsimni qildirdim. In translation process for Uzbek language we use left-to right structure. The verb is translated. According which group does it belong to: one or multy syllabled, ended with voiced consonant, ended with vowel we can put correct endings.

1.2. The endings of passive and reflexive voices.

The passive voice in English language looks like to a “category” in Uzbek language. That’s way their formula is entered into the database.

S+am/is/are+V_{III}=>S+V+PV+TS+PS[2]: The book is written- Kitob o‘qiladi.

During translation into English it should be input to lexicon due to homonym suffixes passive and reflexive voices in Uzbek. For instance, I wash-Men yuvinaman=>yuv+in+a+man. In

[1]ÜlleViks. A morphological analyzer for the estonian language: the possibilities and impossibilities of automatic analysis http://www.eki.ee/teemad/morfoloogia/viks1.html

[2]1. PV(passive voice), 2. TS(tense suffix), 3. PS(personal suffixes), 4) $V_{v\text{-verb voice}}$, 5) NP_1-noun plural, 6)NA- animate object (boyboys), NIP-noun irregular plural (child-children)

English the meaning must be like wash-1) yuvmoq, 2) yuvinmoq; close-ochmoq, ochilmoq, beginboshlamoq, boshlanmoq and others. And they are put in the discrete paradigms (V_v) such kind of verbs. But it should be given some grammars for these verbs: if S+V_v+Noun=>active, if S+V_v+nonNoun=>reflexive voice

1.3. Cooperative voice (Birgalik nisbat)

–sh (-ish) suffixes are belongs to the voice. What kind of English verb forms suit to the voice. It can be synonym to the plural form in Uzbek as well. For example, Bolalar kelishdi (keldilar)-The children came. So we use plural form as it comes like: NP_1+V => they/ NA+s {-es}/ NIP

2. Functional forms of Uzbek verbs (non-finite forms of the verb)

There are three types functional forms of Uzbek: participle (sifatdosh), harakat nomi, adverbial participle (ravishdosh). English has three types: gerund, infinitive and participle. The characteristics and capacity both of the languages are dissimilar. And they are not suit for each others. In the chart pointed out versions different functions of languages.

	Suffixes	Infinitive (to)	Gerund (v+ing)	Participle (V_{III})

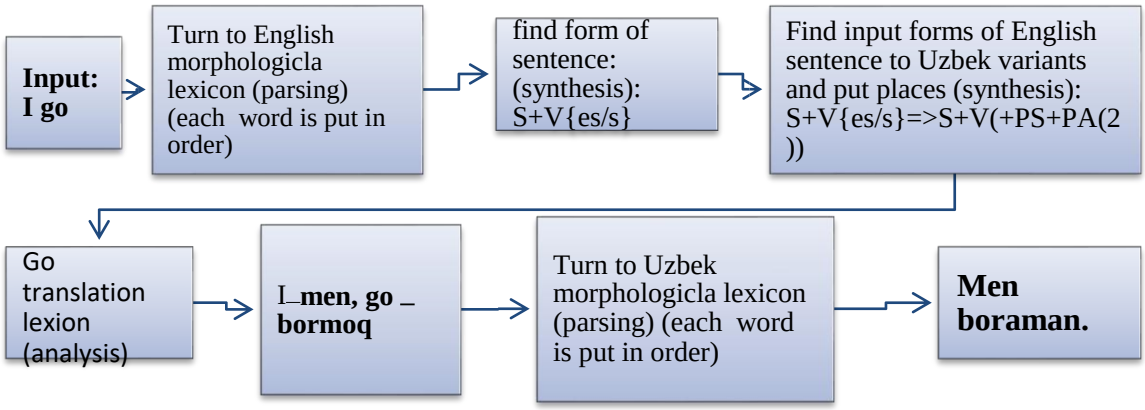
Harakat nomi	-sh,-ish,-v,-uv, -moq, -maslik	+	+	-
Sifatdosh	-gan,-kan,-qan, -yotgan,ayotgan,-ydigan, -adigan, -mas	-	+	-
Ravishdosh	-guncha, -kuncha,quncha, -gach, -kach, -qach, b, -ib, -a, -y, may, -mayin, ma	-	Till (until), by, after	+
Harakat nomi	O‘qish foydali	To read is useful.	Reading is useful.	-
Sifatdosh	Yonayotgan olov ajoyib.	-	Burning fire is wonderful.	
Ravishdosh	Ish qilinguncha vaqt tugaydi. Ish tugatilgach uyga ketdik.	-	Till doing work, time will be over.	After having finished work, we went home.

The morphologic analysis of English is identified coming words in order. It should be responded so that to solve some matters.

1. Verb comes after subject, and then it is considered as a predicate. Then checked simple and complexity of verbs (gerund, infinitive, modal, phrasal verb, have+noun, make+noun, do+noun, take+noun, have+noun+verb)
2. To identify tenses (present, past, future, future in the past)
3. Types of sentences (darak-declerative (Dec.), inkor-negative (Neg), so‘roq-interrogative (Int.), buyruq-imperative (Imp), so‘roq-inkor (IN), undov-exclamatory (EX).
4. To identify transitive and intransitive verbs. It helps to clarify accusative (case) in English. For example, to read Øthe book–kitob**ni** o‘qimoq, go Øhome-uy**ga** bormoq. As we see there isn’t any preposition in English but in Uzbek two different cases.
5. Next step to formulate sentence in two languages (Firstly we take simple sentences). Some of them are below:

	PS PS	**PC**	**PP**	**PPC**
Dec.	S+V{es/s}=> S+V(+PS+ $PA_{(2)}$)	S+am{‘m}/is{s}/are{‘re}+V(-ing) => S+V+(PC+ $PA_{(2)}$)	S +have{‘ve}/has {‘s}+ $V_{(III,-ed)}$=> S+V+(PP+ $PA_{(1)}$)	S+ have{‘ve}/has {‘s}+been+V(ing)
Neg.	S+do not {don’t}/does not {doesn’t}+V=>+V(+ NA+PS +$PA_{(2)}$	S+am/is/are+not / {isn’t/aren’t}+V(-ing) => S+V+(NA+PC+ $PA_{(2)}$+	S+have+not/{haven’t}/has +not/{hasn’t}+ $V_{(III,-ed)}$=> S+V+(PP+NA+PastA+ $PA_{(1)}$)	S+have+not+/{haven’t}/has +not/{hasn’t}+ been+$V_{(ing)}$=> S+V+(PP+NA+PastA+ $PA_{(1)}$)

Int.	Do/Does+S+V=?=> S+V(+PS+ $PA_{(2)}$+QA=?)	Am/is/are +S+ V (-ing)=? => S+V+ (PC+$PA_{(2)}$+QA=?)	Have/has +S+ $V_{(III,-ed)}$=> S+V+(PP +PastA+ $PA_{(1)}$+QA=?)	Have/has +S+ been+$V_{(-ing)}$=> S+V+(PP +PastA+ $PA_{(1)}$+QA=?)
Int. N	Don't/Doesn't+S+V= ?=> S+V(+NA+PS+ $PA_{(2)}$+QA=?	Am/is/are+S+not+ V(-ing)=? => S+V+(NA+PC+$PA_{(2)}$+QA=?)	Have/has +S+ not+$V_{(III,-ed)}$=> S+V+(NA+PP PastA+$PA_{(1)}$+QA =?)	Have/has +S+been+$V_{(-ing)}$=> S+V+(NA+PP +PastA+ $A_{(1)}$+QA=?)



1. I go=>:S+V{es/s}=> S+V{es/s}=>S+V(+PS+PA(2))=>Men boraman.

1.1. I don't go=>S+do not {don't}/does not {doesn't}+V=> S+V(+NA+PS+PA(2))=> Men bormayman.

1.2. Do I go?=> Do/Does+S+V=?=> S+=> S+V(+PS+PA(2)+QA=?)=> Men boramanmi?

1.3. Don't I go? => Don't/Doesn't+S+V=? =>S+V(+NA+PS+PA(2)+QA=?) => Men bormaymanmi?

2. I am going=>S+am{'m}/is{s}/are{'re}+V(-ing)=> S+V+(PC+PA(2))=> Men boryapman.

2.1. I am not going=>S+am/is/are+not / {isn't/aren't}+V(-ing) => S+V+(NA+PC+PA(2))=> Men bormayapman.

2.2. Am I going ?=> Am/is/are +S+ V(-ing)=? => S+V+(PC+PA(2)+QA=?)=>Men boryapmanmi?

2.3. Am I not going ?=>Am/is/are +S+not+ V(-ing)=? => S+V+(NA+PC+PA(2)+QA=?)=>Men bormayapmanmi?

3. I have gone=>S+have{'ve}/has {'s}+$V_{(III,-ed)}$=> S+V+(PP+PA(1))=> Men borib bo'ldim.

3.1. I have not gone=>S+have+not/{haven't}/has +not/{hasn't}+V(3,-ed)=> S+V+(PP+NA+PastA+PA(1))=> Men borib bo'lmadim.

3.2. Have I gone?=> Have/has +S+ $V_{(III,-ed)}$=> S+V+(PP+NA+PastA+PA(1))=> Men borib bo'lmadim.

3.3. Have I not gone?=> Have/has +S+ not+$V_{(III,-ed)}$=> S+V+(NA+PP PastA+$PA_{(1)}$+QA=?)=> Men bormaganmidim?

We have presented a rule-based morphological analysis system for English-Uzbek translation system. As we admitted that it is initial (opening) stage of translation system. Using theories of

typological grammar we create deep principles of morphological analysis. Rich lexicon, full based grammar rules, the base of terms are all of them help to improve analyzing text translation process. And we hope the next researches on linguistic database of translation program will be advanced within the next few years.

References 1. Кривоносов А.Т. (2001) Система классов слов как отражение структуры языкового создания. Москва –Нью –Йорк: Че-ро, 846 с.

2. Marchuk Y.N. (2003) The Burdens and Blessings of Blazing the Trail. In Journal of quantitative linguistics. Trier, Swets, Zeitlinger, Vol. 10, No. 2 Aug. p 81-87

3. ÜlleViks. A morphological analyzer for the estonian language: the possibilities and impossibilities of automatic analysis http://www.eki.ee/teemad/morfoloogia/viks1.html

4. Idem

5. Miriam Butt, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The parallel grammar project. COLING'02, Workshop on Grammar Engineering and Evaluation.

6. Eugene Charniak, Kevin Knight, and Kenji Yamada. 2003. Syntax-based language models for statistical machine translation. MT Summit IX.

УДК 81.32

EXPERIMENTS WITH RUSSIAN TO KAZAKH SENTENCE ALIGNMENT

Zhenisbek Assylbekov, *Nazarbayev University, Astana, Kazakhstan zhassylbekov@nu.edu.kz*
Bagdat Myrzakhmetov, Aibek Makazhanov, *National Laboratory Astana, Astana, Kazakhstan bagdat.myrzakhmetov@nu.edu.kz, aibek.makazhanov@nu.edu.kz*

Sentence alignment is the final step in building parallel corpora, which arguably has the greatest impact on the quality of a resulting corpus and the accuracy of machine translation systems that use it for training. However, the quality of sentence alignment itself depends on a number of factors. In this paper we investigate the impact of several data processing techniques on the quality of sentence alignment. We develop and use a number of automatic evaluation metrics, and provide empirical evidence that application of all of the considered data processing techniques yields bitexts with the lowest ratio of noise and the highest ratio of parallel sentences.

Keywords: sentence alignment, sentence splitting, lemmatization, parallel corpus, Kazakh language

ЭКСПЕРИМЕНТЫ ПО ВЫРАВНИВАНИЮ ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ ДЛЯ РУССКО-КАЗАХСКИХ ПРЕДЛОЖЕНИЙ

Асылбеков Женисбек Аманбаевич, *Назарбаев Университет, Астана, Казахстан zhassylbekov@nu.edu.kz*
Мырзахметов Багдат Омаралиевич, Макажанов Айбек Омиржанович, *Национальная Лаборатория, Астана, Казахстан,* *bagdat.myrzakhmetov@nu.edu.kz, aibek.makazhanov@nu.edu.kz*

Выравнивание параллельных текстов по предложениям является заключительным этапом построения параллельного корпуса и возможно оказывает наибольшее влияние качество конечного продукта и на точность систем машинного перевода, использующих этот корпус для обучения. Качество же выравнивания по предложениям, в свою очередь, также зависит от ряда факторов. В данной статье мы исследуем влияние некоторых способов обработки данных на качество выравнивания по предложениям. Мы разрабатываем и используем несколько автоматических метрик оценки качества, и приводим эмпирические доказательства того, что совокупное использование всех рассмотренных способов обработки данных приводит к получению параллельных корпусов с наименьшей долей шума и наибольшей долей параллельных предложений.

Ключевые слова: выравнивание по предложениям, разбивка по предложениям, лемматизация, параллельный корпус, казахский язык

1. Introduction

Sentence alignment (SA) is the problem of identification of parallel sentences (pairs of sentences that constitute translations from a source to a target language) from given a given pair of source and target documents, where the target document is assumed to be a translation of the source (mutual translation assumption is also common). More formally, given a source document D_s and a target document D_t represented as lists of sentences S and T respectively, SA is the task of building a list of pairs P, where each pair p contains 0 or more (ideally one) source sentence(s) aligned to 0 or more (ideally 1) target sentence(s). Approaches that consider sentence length correlations [1, 2], bilingual lexicon-based solutions [3], and combinations of the two [4] have been proposed in the past to solve this problem in a sufficiently accurate and efficient manner. In this paper we do not offer a new solution to the problem, nor do we try to improve the existing approaches. Our goal is to investigate what could be done to the input data (not to the methods) to improve the quality of SA.

We begin by asking a few questions, which are inspired directly by the definition of the problem and by ways of solving it. First, a formal definition of SA problem assumes that documents to be aligned are already split into sentences. However, in practice it is almost never the case, and one has to perform splitting before SA. Assuming that one uses for that a statistical approach that requires training, e.g. punkt splitter [5], a question regarding the choice of training data arises: *does it suffice to train the splitter on any data, or would it be beneficial to train on a sample drawn from a target domain?* Second, assuming one uses a lexicon based approach to SA, *should one bother trying to reduce typos and data sparsity of the input, and what lexicon to use automatically induced or handcrafted?* Lastly, *after sentences have been aligned can we still increase the portion of parallel pairs?* In an attempt to answer these questions, we propose to employ the following five data processing techniques: (i) *domain adapted sentence splitting*; (ii) *error correction*; (iii) *lemmatization* (to reduce sparsity); (iv) *use of handcrafted bilingual lexicons*; (v) *junk removal*.

The objective of this work is to assess the impact of the proposed data processing techniques on SA accuracy and find the combination of thereof which maximizes the quality of parallel corpora produced by SA.

2. Data Collection

For our experiments we have crawled three websites, `akorda.kz`, `strategy2050.kz`, `astana.gov.kz`, using our own Python scripts to download only specific branches of these sites - mainly news and announcements. The choice of these specific websites is motivated by the fact that all of them provide built-in document alignment, i.e. each news article or announcement in Russian contains a link to a corresponding translation into Kazakh (sometimes translation direction may be opposite). Rare exceptions to this behavior include cases where translation link is absent or broken. Such pairs are not included into the data set. The obtained pairs of HTML documents are parsed in a

site-specific manner with the help of Python `BeautifulSoup` library to produce raw Russian and Kazakh texts aligned on the document level.

2. Baseline Sentence Alignment and Data Processing Techniques

Let us describe the basic sentence alignment (BSA) procedure that does not assume any of the data processing techniques (DPTs) that we propose. At the sentence splitting stage BSA uses NLTK `punkt` tool [5] trained on approximately 200-250 Mb of plain texts from Russian and Kazakh Wikipedias. Next we tokenize the documents with a Perl-script from an SMT-toolkit Moses [6]. Sentence alignment is performed on tokenized and lowercased texts using `hunalign` [4]. After sentences are aligned we restore their original, non-tokenized and non-lowercased, format. In what follows we describe the implementation of the DPTs that we propose.

Domain adapted sentence splitting. To see whether we can gain any improvements at sentence splitting step, we train `punkt` on 350-370 Mb of text from the news domain rather than on random Wikipedia texts and supply it with a list of abbreviations in Russian and in Kazakh.

Error correction. In this work we consider a light-weight error correction procedure, which involves normalization of scripts (alphabets) used in a given text. Electronic texts written in Cyrillic (Russian and Kazakh alike), especially those which were digitized in 1990s, sometimes suffer from mixed scripts, i.e. when Latin letters are used instead of Cyrillic ones and vice versa: e.g. in a Kazakh word “ecipткi” it is possible to replace the letters ‘e’ (u+0435), ‘c’ (u+0441), ‘i’ (u+0456), ‘p’ (u+0440) with their Latin homoglyphs, ‘e’ (u+0065), ‘c’ (u+0063), ‘i’ (u+0069) and ‘p’ (u+0070), which allows a total of 2^5=32 spelling variations. To reduce data sparseness that may result from this, we developed a tool which tries to resolve unambiguous cases.

Lemmatization. Another possible way to improve sentence alignment is a prior lemmatization of texts. Theoretically this should decrease data sparseness and be helpful when combined together with automatic construction of a bi-dictionary. Also, one can try to align sentences when both texts and handcrafted dictionary are lemmatized. For Russian-side lemmatization we use an open source tool Mystem [7], and for Kazakh – morphological disambiguation tool developed by Makhambetov et al. [8].

Adding bilingual dictionaries for sentence alignment. In the baseline approach no bilingual dictionaries are provided to `hunalign`, and very often such dictionaries are not available, especially for low-resourced languages. In such cases one can construct and exploit rough bi-dictionaries in three steps: (1) apply the baseline sentence alignment; (2) use `hunalign` again to align already aligned texts with the `-autodict` option - the byproduct of this step is a bidictionary; (3) finally, apply `hunalign` to the original non-aligned texts with the obtained bidictionary. We experiment with both options, using as a handcrafted dictionary a compilation of resources obtained from `Bitextor` [9], `Apertium-kaz` [10], and `www.mtdi.kz`.

Junk removal. Finally, we believe that removing the following sentence pairs should benefit the final corpora (hereinafter such pairs are called “junk”):

- at least one of the sides (Kazakh or Russian) is empty;
- at least one of the sides does not contain any letters (Latin and Cyrillic);
- both sides are identical after tokenization and lowercasing.

3. Evaluation Metrics

The most reliable way to evaluate the quality of SA is to perform human evaluation by checking the output of an automatic SA method, and calculating its accuracy, i.e. percentage of correct alignments in the total number of aligned pairs. To evaluate the baseline SA in this fashion, we ran the baseline on our data set and on the data crawled from an additional page-aligned website (`adilet.zan.kz`). We then randomly sampled 800 sentence pairs (including null alignments produced by `hunalign`) and asked three annotators to label each pair as parallel or not. The

inconsistencies were resolved by the third annotator. In Table 2 we present the results of this procedure (averages are calculated excluding the results for `adilet.zan.kz` for comparison purposes).

Table 1. Accuracy of the baseline SA, per website and per annotator

Web-site	**Annotator 1**	**Annotator 2**	**Annotator 3**	**Annotator 4**
`adilet.zan.kz`	0.9375	0.9425	0.8825	**0.9075**
`akorda.kz`	0.9525	0.9450	0.8675	**0.9050**
`astana.gov.kz`	0.7925	0.7950	0.7325	**0.7400**
`strategy2050.kz`	0.7700	0.7525	0.6575	**0.6900**
Average	0.8383	0.8308	0.7525	**0.7783**

As we can see, on a sample of our data set (the latter three websites) the baseline SA method achieves the average accuracy of ~78%. As we will show later application of the DPTs can increase the accuracy. But to show that, we need to develop a more efficient way of computing SA, because to test all configurations of the DPTs would require us to perform expensive human evaluation procedure up to 10 times.

Table 2. Features used in a learning-based SA accuracy estimator

#	**DC**	**Feature**	---------	**#**	**DC**	**Feature**
1,2	S,T	length in characters		19,20	S,T	count of personal initials
3	ST	MMR(F1,F2)		21	ST	COS(F19*,F20*)
4,5	S,T	length in tokens		22,23	S,T	ratio of alphanumerics
6	ST	MMR(F4,F5)		24	ST	MMR(F22,F23)
7,8	S,T	count of symbols		25,26	S,T	count of words in quotes
9	ST	COS(F7*,F8*)		27	ST	MMR(F25,F26)
10,11	S,T	count of numerals		28,29	S,T	count of words in parenthesis
12	ST	COS(F10*,F11*)		30	ST	MMR(F25,F26)
13,14	S,T	count of digits		31	ST	num. of tokens in identical pairs
15	ST	COS(F13*,F14*)		32	ST	min-max ratio between unique tokens in source and target sentences
16,17	S,T	count of latin alphanumerics				
18	ST	COS(F16*,F17*)		33-35	ST	`Hunalign` score: absolute, relative, min-max scaled.

To compute the accuracy estimate of SA more efficiently we cast the SA problem as a classification task, where given a pair of source and target sentences, a supervised learning algorithm estimates the probability of the pair being parallel. We design a set of 35 features listed in Table 3, where each feature has a domain of calculation (DC), and can be calculated for the (S)ource or the (T)arget sentence, or for both (ST). MMR refers to min-max ratio, e.g. MMR(F1,F2) means that the minimum of features 1 and 2 is divided by the maximum of the two. Similarly, COS refers to cosine similarity calculated for the count-vectors of a given pair of features.

We extract these features from the annotated sample that was used for human evaluation and perform a five-fold cross-validation using a range of classifiers implemented in Python

`scikitlearn` library. Gradient Boosting classifier achieved the highest F-measure of 0.94 (per-fold average) and the lowest variance of 0.08. Therefore, we use this classifier as a rough estimator of SA accuracy as follows. Given the alignment pairs produced by SA, the estimator classifies each pair as parallel or not. The ratio of pairs classified as parallel to the total number of pairs provides the accuracy estimate.

4. Experiments and Results

We compare different configurations of DPTs. To refer to a specific technique we use the following abbreviations: adapted splitting - AS, error correction - EC, automatically obtained dictionary - AD, handcrafted dictionary - HD, lemmatized handcrafted dictionary - LHD, lemmatization - L, junk removal - JR.

We measure the quality of produced bitexts in the total number of parallel pairs (P) and the automatic accuracy estimation (P/T). As it can be seen from Table 3, combined application of all DPTs (AS+EC+ L+LHD+JR) achieves the highest accuracy per-site and on average improves ~6% over the baseline. It also produces about 5.5K more parallel sentences (total) than the baseline, and only 40 pairs less than the same configuration shy of JR. However, notice how drastically junk removal increases the accuracy of SA, more than 5%. Hence, we indeed can increase the portion of parallel sentences after SA has been performed, through the removal of the pairs which are very unlikely to be parallel. We also notice that text lemmatization applied without the use of a handcrafted dictionary (AS+EC+L) produces far less parallel sentences than the baseline and is only 0.28% more accurate. Perhaps more surprisingly adding an automatically obtained dictionary to this configurations (AS+EC+L+AD) makes matters even worth.

Table 3. Qualities of bitexts produced using various processing techniques

Method	`akorda.kz`		`astana.gov.kz`		`strategy2050.kz`		**total**	**average**
	P	**P/T**	**P**	**P/T**	**P**	**P/T**	**P**	**P/T**
Baseline	70,956	0.9116	63,731	0.7215	201,678	0.6650	336,365	0.7660
AS+EC	70,860	0.9171	63,777	0.7310	202,354	0.6740	336,991	0.7740
AS+EC+AD	71,002	0.9193	63,298	0.7266	200,319	0.6681	334,619	0.7713
AS+EC+HD	71,062	0.9199	**64,210**	0.7365	204,716	0.6818	339,988	0.7794
AS+EC+LHD	71,089	0.9203	64,159	0.7356	204,857	0.6822	340,105	0.7794
AS+EC+L	70,605	0.9138	63,260	0.7246	199,675	0.6661	333,540	0.7682
AS+EC+L+AD	70,862	0.9178	62,929	0.7213	198,500	0.6638	332,291	0.7676
AS+EC+L+HD	71,114	0.9204	64,119	0.7345	206,225	0.6880	341,458	0.7810
AS+EC+L+LHD	**71,129**	0.9208	64,029	0.7333	206,797	0.6899	**341,955**	0.7813
AS+EC+L+LHD+JR	71,115	**0.9488**	64,014	**0.7823**	**206,786**	**0.7667**	341,915	**0.8326**

To measure the level of noise (the lower the better) in the produced bitexts, we calculate proportion of short pairs[1] among parallel pairs (S/P), and proportion of junk pairs among all pairs (J/T). From Table 3 we notice that a complete set of DPTs achieves the second lowest S/P ratio after the AS+EC+L+AD configuration, which also achieves the second lowest J/T ratio. Thus, using auto-

[1] Short pairs are defined as those where both sides, Kazak and Russian, contain *three* or less words. Usually such text chunks are dates, titles, enumerations, etc., and they do not qualify as full sentences.

induced dictionary on lemmatized text produces least amount of parallel sentences and the lowest ratio of thereof, but resulting bitexts actually come out less noisy than in other DPT configurations. We will study this strange behavior in the future.

Table 4. Noise level in bitexts produced using various processing techniques

Method	akorda.kz		astana.gov.kz		strategy2050.kz		average	
	S/P	J/T	S/P	J/T	S/P	J/T	S/P	J/T
Baseline	0.0190	0.0294	0.0098	0.0610	0.0202	0.0982	0.0163	0.0629
AS+EC	0.0172	0.0291	0.0084	0.0586	0.0195	0.0969	0.0150	0.0615
AS+EC+AD	0.0169	0.0294	0.0078	0.0588	0.0193	0.0974	0.0147	0.0619
AS+EC+HD	0.0172	0.0292	0.0084	0.0589	0.0193	0.0980	0.0150	0.0620
AS+EC+LHD	0.0171	0.0294	0.0084	0.0591	0.0193	0.0981	0.0149	0.0622
AS+EC+L	0.0171	0.0296	0.0084	0.0613	0.0198	0.0979	0.0151	0.0630
AS+EC+L+AD	**0.0167**	0.0297	**0.0072**	0.0620	**0.0191**	0.0950	**0.0143**	0.0622
AS+EC+L+HD	0.0170	0.0295	0.0084	0.0627	0.0192	0.0997	0.0149	0.0640
AS+EC+L+LHD	0.0170	0.0294	0.0085	0.0629	0.0192	0.1002	0.0149	0.0642
AS+EC+L+LHD+JR	0.0168	**0.0000**	0.0082	**0.0000**	0.0192	**0.0000**	0.0147	**0.0000**

5. Conclusion

In this work we have shown that various techniques of data processing can increase the accuracy of sentence alignment and reduce the level of noise in the resulting bitexts. We provided empirical evidence that combined application of five simple data processing techniques before and after sentence alignment results in production of parallel corpora with the lowest ratio of noise and the highest ratio of parallel sentences.

Acknowledgements. This work has been funded by the Nazarbayev University under the research grant №064-2016/013-2016, and by the Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan under the targeted program O.0743 (0115PK02473).

References

1. Peter F. Brown, Jennifer C. Lai, and Robert L. Mercer. “Aligning sentences in parallel corpora”. ACL, 1991.

2. William A. Gale and Kenneth Ward Church. “A program for aligning sentences in bilingual corpora”. ACL, 1991.

3. Martin Kay and Martin Roscheisen. “Text-translation alignment”. Computational Linguistics, 19(1), 1993.

4. Varga, Daniel et al. “Parallel corpora for medium density languages”. AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4 292 (2007): 247.

5. Kiss, Tibor, and Jan Strunk. “Unsupervised multilingual sentence boundary detection”. CL 32.4 (2006): 485-525.

6. Koehn, Philipp et al. “Moses: Open source toolkit for statistical machine translation”. Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions 25 Jun. 2007: 177-180.

7. Segalovich, I. “A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine”. MLMTA. (2003)

8. Makhambetov, O., Makazhanov, A., Sabyrgaliyev, I., Yessenbayev, Z. “Data-driven morphological analysis and disambiguation for Kazakh”. CICLing 2015, 151–163.

9. Espla-Gomis, Miquel, and Mikel Forcada. “Combining content-based and url-based heuristics to harvest aligned bitexts from multilingual sites with bitextor”. The Prague Bulletin of Mathematical Linguistics 93 (2010): 77-86.

10. Washington, Jonathan, Ilnar Salimzyanov, and Francis M Tyers. “Finite-state morphological transducers for three Kypchak languages”. LREC 2014: 3378-3385.

УДК 681.3

АЛГОРИТМ ОБРАЗОВАНИЯ СЛОВОФОРМ ДЛЯ АВТОМАТИЗАЦИИ ПРОЦЕДУРЫ ПОПОЛНЕНИЯ БАЗЫ ДАННЫХ СЛОВАРЯ

***Бакасова Пери Султановна**, магистрант КГТУ им. И. Раззакова, Кыргызстан, 720044, г.Бишкек, пр. Айтматова Ч., 66,* e-mail: *bakasovap@mail.ru*
***Исраилова Нелла Амантаевна,** к.т.н.,зав. кафедрой ИВТ, КГТУ им. И.Раззакова, Кыргызстан, 720044, г.Бишкек, пр. Айтматова Ч., 66,* e-mail: *inela@mail.ru*

Цель статьи – исследовать алгоритм формирования словоформ по морфологическим признакам части речи. В данной статье рассматривается принципы формирования форм глаголов исходя из следующих морфологических признаков: спряжение (первый, второй), лицо (первое, второе, третье), число (единственное, множественное число), время (прошлое, настоящее, будущее).

Ключевые слова: алгоритм, машинный перевод, морфологические признаки, часть речи, словоформа, морфологический анализ

ALGORITHM OF FORMATION WORD FORMS FOR AUTOMATION REPLENISHMENT OF DICTIONARY DATABASES.

***Bakasova Peri Sultanovna**, undergraduate KSTU. I. Razzakova, Kyrgyzstan, 720044, Bishkek, av. Aitmatov Ch, 66, e-mail: bakasovap@mail.ru*
***Israilov Nella Amantaevna**, PhD, Head. the Department of ICT, KSTU. I.Razzakova, Kyrgyzstan, 720044, Bishkek, av. Aitmatov Ch, 66,e-mail: inela@mail.ru*

The purpose of the article - the investigate of algorithm to the formation of word forms on morphological characteristics of the speech. This article discusses the principles of the verb forms the basis of the following morphological features: conjugation (first, second), a person (first, second, third), number (singular, plural), time (past, present, future).

Keywords: algorithm, machine translation, morphological features, part of speech, word forms, morphological analysis

Для обеспечения гибкости при переводе текстов (предложений, слов) необходим словарь, содержащий различные формы слова и их переводы. Для пополнения данного словаря используем алгоритм формирования слов по морфологическим признакам части речи, к которой относится исходное слово на русском языке(RU) и его перевод на кыргызском языке(KG). Суть алгоритма заключается в формировании разных форм

исходного слова и его перевода по морфологическим признакам части речи, к которой относится исходное слово и слово- перевод.

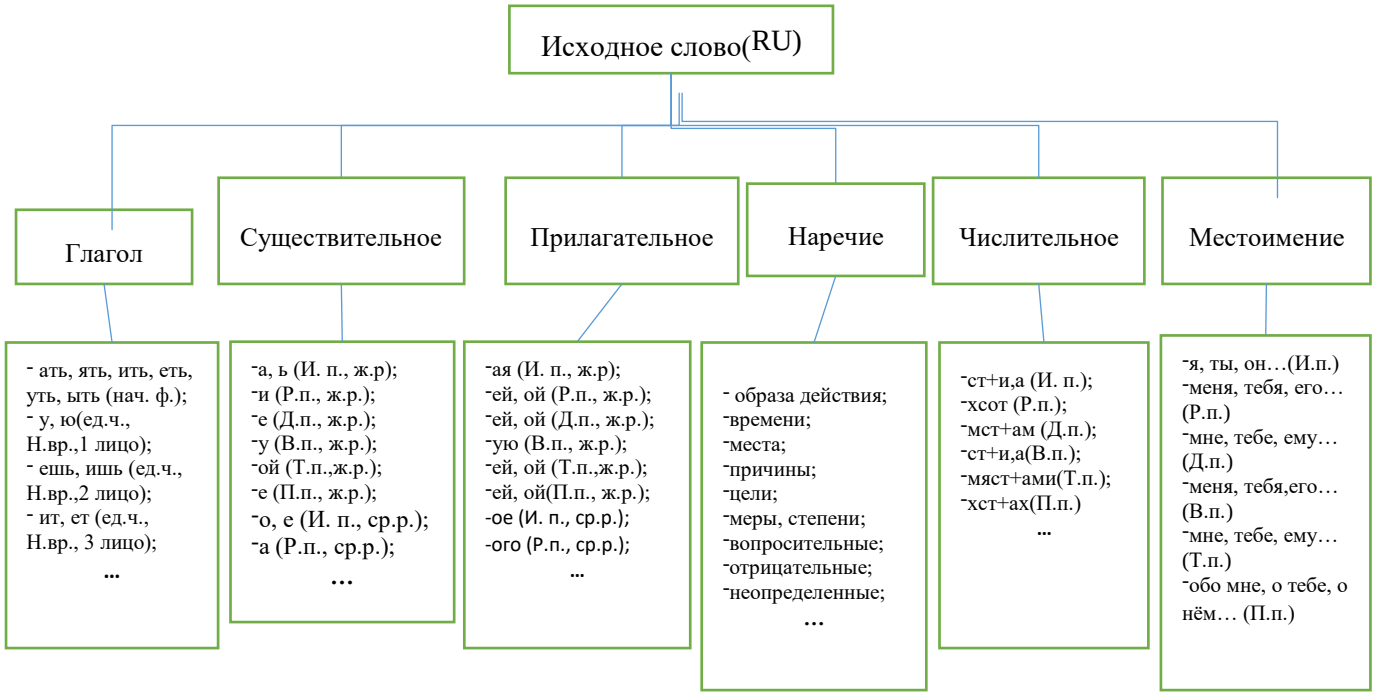


Рис. 1 Схема формирования словоформ по морфологическим признакам части речи в русском языке

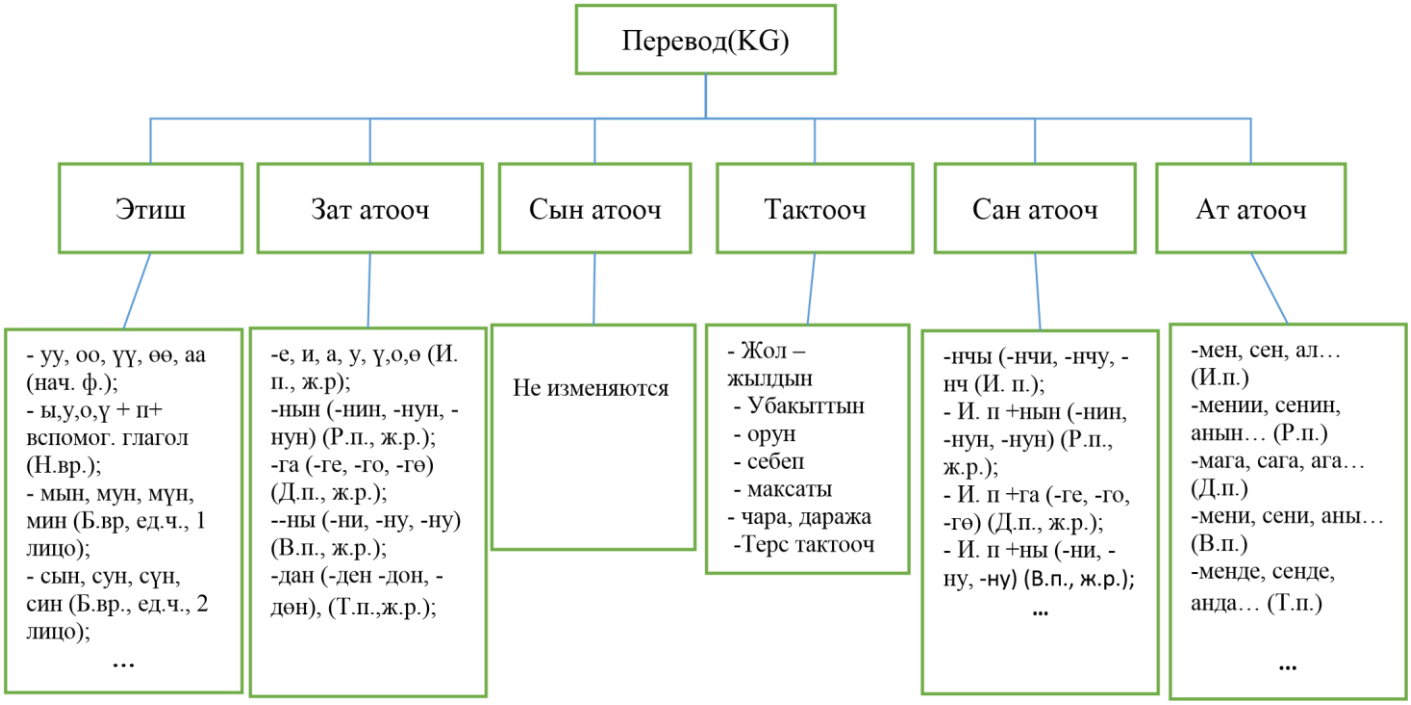


Рис. 2 Схема формирования словоформ по морфологическим признакам части речи в кыргызском языке

Формирование форм глаголов в русском языке

Алгоритма формирования слов по морфологическим признакам глаголов позволяет распознать к какому спряжению относится искомый глагол и образует формы слова следующим образом.

Формы глаголов в русском языке образуются спряжением глаголов по числам и лицам.

Формы глагола первого спряжения настоящего времени образуются спряжением глагола по лицам и числам добавляя личные окончания глаголов, показанные в таблице 1:

Таблица 1. – Личные окончания глаголов первого спряжения в настоящем времени

	Ед. ч.	Мн. ч.
1 лицо	ю	ем
2 лицо	ешь	ете
3 лицо	ет	ют

Формы глагола второго спряжения настоящего времени образуются спряжением глагола по лицам и числам добавляя личные окончания глаголов, показанные в таблице 2:

Таблица 2. - **Личные окончания глаголов второго спряжения в настоящем времени**

	Ед. ч.	Мн. ч.
1 лицо	ю	им
2 лицо	ишь	ите
3 лицо	ит	ят

Глаголы в прошедшем времени спрягаются по родам и числам с помощью окончаний показанных в таблице 3.

Таблица 3. – **Окончания глаголов в прошедшем времени**

	Ед. ч.	Мн. ч.
м.р.	л	ли
ж.р.	ла	ли
ср.р.	ло	ли

Формы глаголов прошедшего совершенного вида образуются добавлением приставки “с” и личных окочаний глаголов, показанных в таблице 3.

Глаголы в будущем времени имеют те же окончания что и глаголы в настоящем времени, но к ним еще и добавляются такие приставки как – *с, по, раз, у* и т.д.

§ Формирование форм глаголов в кыргызском языке

Формы глаголов в кыргызском языке, как и в русском образуются спряжением глаголов по числам и лицам.

Формы глагола настоящего времени образуются с помощью двух глаголов — основного и вспомогательного: к основному глаголу в исходной форме — в повелительном наклонении — прибавляется окончание -п, если глагол оканчивается на гласную, или -ып (ип, -уп, -үп), если глагол оканчивается на согласную.

В качестве вспомогательных чаще всего берутся глаголы жат - лежи и жүр - иди; реже глаголы тур - встань и отур - садись в форме настоящего времени, которые теряют свое значение и на русский язык не переводятся.

При спряжении глаголов настоящего времени основной глагол не изменяется, а вспомогательный принимает личные окончания будущего времени,

Формы глаголов будущего времени образуются путем прибавления к основе глагола (повелительному наклонению) личных окончаний

Таблица 4. – Личные окончания глаголов в настоящем времени

	Единственное число	Множественное число
1 лицо	-мың (-миң, -муң, -мүң)	-быз (-биз, -буз, -бүз)
2 лицо	-сың (-сиң, -суң, -сүң)	-сыңар (-сиңер, -суңар, -сүңөр)
2 лицо (в.ф.)	-сыз (-сиз, -суз, -сүз)	-сыздар (-сиздер, -суздар, -сүздөр)
3 лицо	-т	-шат (-шет, -шот, -шөт)

При спряжении глагола в будущем времени между основой глагола (повелительным наклонением) и личным окончанием появляется звук -Й-, если глагол оканчивается на гласную, или звук -А- (-Е-, -О-, -Ө-), если глагол оканчивается на согласную. Если исходная форма глагола заканчивается на Й, то при сочетании этих глаголов с личными окончаниями будущего времени во всех лицах вместо ЙО пишется Ё, вместо ЙА — Я.

В кыргызском языке глаголы прошедшего определенного времени образуются путем прибавления к основе глагола суффиксов -ды (после гласных и звонких согласных) или -ты (после глухих, согласных) плюс личные окончания глагола.

Таблица 5. – Окончания глаголов в прошедшем совершенном времени

	Единственное число	Множественное число
1 лицо	-дым (-дим, -дум, -дүм)	-дык (-дик, -дук, -дүк)
2 лицо	-дың (-диң, -дуң, -дүң)	-дыңар (-диңер, -дуңар, -дүңөр)
2 лицо (в.ф.)	-дыңыз (-диңиз, -дуңуз, -дүңүз)	-дыңыздар (-диңиздер, -дуңуздар, -дүңүздөр)
3 лицо	-ды (-ди, -ду, -дү)	-шты (-шти, -шту, -штү)

Прошедшее обычное время образуется при помощи суффикса -ган (-ген? -гонг -ген), если основа глагола оканчивается на гласный и звонкий согласный, или -кан (-кен, кон, кон), если основа глагола оканчивается на глухой согласной.

Таблица 6. – Окончания глаголов в прошедшем обычном времени

	Единственное число	Множественное число
1 лицо	-ганмың (-генмиң, -гонмуң, -гөнмүң)	-ганбыз (-генбиз, -гонбуз, -гөнбүз)
2 лицо	-гансың (-генсиң, -гонсуң, -гөнсүң)	-гансыңар (-генсиңер, -гонсуңар, -гөнсүңөр)
2 лицо (в.ф.)	-гансыз (-генсиз, -гонсуз, -гөнсүз)	-гансыздар (-генсиздер, -гонсуздар, -гөнсүзд
3 лицо	-ган (-ген, -гон, -гөн)	-ышкан (-ишкен, -ошкон, -үшкөн)

ө

р

Е

сли глагол оканчивается на согласную, то в 3 л. мн, числа между основой и личным окончанием вставляется показатель совместимости -ыш (-иш, -уш, -уш).

§ Запись в словарь

Слова записываются в словарь в виде:

Слово1(RU)|Перевод1(KG)/

Слово2(RU)|Перевод2(KG)/

Также образованные формы слова сохраняются в базе по частям речи.

Например, глаголы на русском языке хранятся в базе глаголов, а их переводы на кыргызском в базе глаголов кыргызского языка в следующем формате:

Глагол(RU)/Форма1/Форма2/Форма3/...
Этиш(KG)/Форма1/Форма2/Форма3/...

Вывод:

Алгоритм формирования слов по морфологическим признакам части речи обеспечивает быстрое пополнение базы словаря за счет автоматического словообразования на основе искомого слова и его перевода и обеспечивает гибкость при переводе.

Недостатком является возможность возникновения ошибки при формировании словоформ. Поэтому дальнейшие оптимизации и модификации данного алгоритма являются актуальными.

Список литературы

1. Грамматика кыргызского языка: краткий справочник для студентов, Бишкек 2002 2. Глазунова О. Грамматика русского языка в упражнениях и комментариях (Морфология + Синтаксис).
3. Ө. Калыева Разговорник русско-кыргызский - Орусчакыргызча сүйлөшмө
4. Исраилова Н.А. Принципы организации морфологического анализатора в трансляторе. /Известия КГТУ им. И. Раззакова-2010, №20, С233-236
5. Исраилова Н.А. Организация морфологического анализа в трансляторах. /Вестник Восточно-Казахстанского государственного технического университета им. Д. Серикбаев, Усть-Каменогорск, №1, март, 2012 г.-С 97-101
6. Исраилова Н.А. Алгоритмы отладки процесса трансляции. /Известия КГТУ им. И. Раззакова-2011, №22, С278-279

УДК 004.432.4

AUTOMATING OF THE TEXT GENERATION WITH A GIVEN REPRESENTATIVENESS PHONETIC UNITS BY A FORMALIZATION OF PHONOLOGICAL RULES

***Aigerim Buribayeva**, PhD, L.N.Gumilyov Eurasian National University, 010008, Astana, Pushkin str. 2, e-mail: buribayeva@mail.ru*

Arman Kaliyev, *L.N.Gumilyov Eurasian National University, 010008, Astana, Pushkin str. 2, e-mail: kaliyev.arman@yandex.kz*
Banu Yergesh, *L.N.Gumilyov Eurasian National University, 010008, Astana, Pushkin str. 2, e-mail: saturn_banu@mail.ru*

The article describes formalization of phonological rules of Kazakh language and use it to automate the process of formation carried the text of the material with a given phonetic units of representativeness (in particular diphones). This is necessary, particularly in the development of automatic speech synthesis. Using this versatile program, it will be able to get the text material with full coverage of all possible diphones for all Turkic language.

Keywords: Speech synthesis, text analyzer, sound units, diphones, statistics, acoustic database, text body.

АВТОМАТИЗАЦИЯ ФОРМИРОВАНИЯ ТЕКСТОВОГО МАТЕРИАЛА С ЗАДАННОЙ ПРЕДСТАВИТЕЛЬНОСТЬЮ ФОНЕТИЧЕСКИХ ЕДИНИЦ С ПОМОЩЬЮ ФОРМАЛИЗАЦИИ ФОНОЛОГИЧЕСКИХ ПРАВИЛ

Бурибаева Айгерим Кеулимжаевна*, PhD, ЕНУ им. Л.Н. Гумилева, 010008, Астана, Пушкина 2, e-mail: buribayeva@mail.ru*
Калиев Арман Куанышевич, *ЕНУ им. Л.Н. Гумилева, 010008, Астана, Пушкина 2, e-mail: kaliyev.arman@yandex.kz*
Ергеш Бану Жантуганкызы, *ЕНУ им. Л.Н. Гумилева, 010008, Астана, Пушкина 2, e-mail: saturn_banu@mail.ru*

В статье описана формализация фонологических правил казахского языка и на ее основе осуществлена автоматизация процесса формирования текстового материала с заданной представительностью фонетических единиц (в частности дифонов). Это необходимо, в частности, при разработке систем автоматического синтеза речи. Так, используя данную универсальную программу, появится возможность получить текстовый материал с полным покрытием всех возможных дифонов для любого тюркского языка.

Ключевые слова: Синтез речи, текстовый анализатор, звуковые единицы, дифоны, статистика, акустическая база, текстовый корпус

Introduction: Many modern trends of development speech technologies involve the use of speech databases. These databases are based on the texts in the utterance of one or more speakers. The criterion of suitability speech serves as the base, above all, the fullness its speech elements. For example, for the development of speech synthesis systems, such elements may be different depending on the selected base units. Most often it is diphones or allophones. For the diphone's synthesis requires a database, containing all possible for a given language binomial combination of phonemes (allophones).

In this connection with the above, particular importance is the preparation of text material in problems where a set of necessary elements of speech pre-defined framework, especially in cases where the resource for processing and structuring of the speech material is limited.

In addition to the phonetic representation, using a special text material provides a fixed amount of compactness. This further allows to significantly reduce the time to process it. It is believed that the use of large amounts of textual material completeness of the coverage units is achieved without special selection, however, this approach inevitably arises redundant database

(which may in the future to require post-processing of the material to exclude duplicates). Also, some rare phonetic combinations that occur at the junction of words, may not once meet. In order to simplify and automate the process of formation of the texts with a given phonetic representativeness, assessing the completeness of coverage of items in a given text, as well as reducing the body text of redundancy due to the removal of a recurring items, was developed by the phonetic analyzer text material.

A feature of this program is its versatility. It can be used for virtually any Turkic language, even without the formalization of phonological rules, it is enough just to have a complete pronouncing dictionary rather than spelling.

The formalization of the phonological rules of sound combinations in Kazakh language: For the development of phonetic transkriptor were investigated orthoepic rules of the Kazakh language. For the conveniance of the reader in this text, the rules are divided into groups that are numbered:

1. In the Kazakh language, if the word begins with vowel «е», then in front of it pronunciation is heard as «й», if the word starts with the vowel «о», «ө», then the pronunciation in front of them formed a brief insert «у» for example, «ет» – «йет», «он» – «уон», «өнер» – «уөнер».
2. If a word begins with a consonant «р» or «л», then the pronunciation of these sounds could be heard before the vowel «ы», «і», depending on the hardness or softness of consonants here «r», «l» means soft analogs "«р» and «л». For example, «рас» – «ырас», «рет» – «ірет», «лас» – «ылас», «лезде» – «ілезде».
3. When pronouncing borrowed sound «ю» as part of word is heard «йүу», «йұу»,, depending on the hardness or softness of the other vowels in syllables. For example: «қою» – «қойұу», «түю» - «түйүу»;;
4. When pronouncing borrowed sound «я» as a part of a word is heard «йа», «йә», depending on the hardness or softness of the other vowels in syllables. For example: «аян» – «айан», «әлия» – «әлійә»;
5. When pronouncing borrowed sound «и» in a consisting of words heard «ый», «ій», depending on the hardness or softness of the other vowels in syllables. For example, «ине» – «ійне», «жина» – «жыйна». If before or after «и» go according to «қ», «ғ» with descender, that the pronunciation of the sound «и» always heard «ый». For example, «қиын» – «қыйын», «қиғаш» – «қыйғаш».
6. When the pronunciation of the diphthong «у» as a part of a word is heard «ұу», «үу», depending on the hardness or softness of the other vowels in syllables. For example, «туыс» – «тұуыс», «күту» – «күтүу».
7. The Vowels «ұ», «ү», «о», «ө» at the beginning or the first syllable of the word in the pronunciation change in the next syllable vowel sounds «ы», «і» on the vowels «ұ», «ү» respectively. For example, «қолтық» – «қолтұқ», «құлын» – «құлұн», «күлкі» – «күлкү», «көлік» – «көлүк»;
8. Vowels «ү», «ө» in the beginning or in the first syllable of the word during the pronunciation changes in the following syllables vowel «е» to the next vowel «ө», for example, «үлкен» – «үлкөн», «өнер» - «өнөр».
9. Vowels «ә», «ү», «і» in the beginning or in the first syllable of the word during the pronunciation changes in the following syllables vowel «а» to its allophone «ә», for example, «ләззат» – «ләззәт», «діндар» – «діндәр».
10. If in a word sounds «с» and «ш», «с» and «ж» or «з» and «ш» meet in succession, then instead of them is pronounced the sound of double «шш». Also, instead of borrowed sound «щ» is pronounced «шш». For example: «досжан» - «дошшан», «басшы - башшы», «сөзшең – сөшшөң», «көшсең»-«көшшөң», «ащы»-«ашшы».

11. If in a word after sounds «з» and «ж» meet in succession, instead of them pronounce dual sound «жж», if sounds «з» and «с» meet in succession, instead of them pronounce dual sound «сс». For example: «бозжорға» - «божжорға», «азсыну - ассынұу».

12. If in a word after sound «н» встречается «б» или «п», then the pronunciation of sound «н» replaced to «м». For example: «мінбер – мімбер», «ойынпаз» – «ойымпаз».

13. If in a word after sound «н» meet «г», «ғ», «к» or «қ» then pronounce of «н» replaced to «ң». For example: «түнгі» – «түңгү», «қашанғы» - «қашаңғы», «зиянкес» – «зыйаңкес», «сәнқой» – «сәңқой».

14. When pronouncing the word in the composition of sound combinations мл, ғн, ғл between two sounds is formed a brief insertion of vowels «ы», «і», depending on the hardness and softness corresponding syllable. For example, «мемлекет» – «мемілекет», «бағлан» – «бағылан», «яғни» – «йағыный».

15. Uncombinable sounds found in many compound words are replaced by the pronunciation sound. For example, «шашбау» - «шашпау», «атбегі»-«атпегі», «атжалман» - «атшалман», «Көпбосын» - «Көппосұн», «түпдерек» - «түбдөрөк», «көпжиын» - «көбжыйын», «көпмүше – «көбмүшө», «түпнегіз» - «түбнегіз», «тасбауыр» - «таспаұуұр» и т.д.

Transkriptor implemented as a program that replaces some other characters in accordance with the rules contained in the control file. Rules are written in accordance with the each item of orthoepic rules of Kazakh language:

1. #е=йе, #о=уо, #ө=уө;
2. #л^а=ыл^а, #л^о=ыл^о, #л^ұ=ыл^ұ, #л^ә=іл^ә, #л^ү=іл^ү, #л^е=іл^е, #л^і=іл^і, р^а=ыр^а, #р^о=ыр^о, #р^ұ=ыр^ұ, #р^ә=ір^ә, #р^ү=ір^ү, #р^е=ір^е, #р^і=ір^і;
3. аю=айұу, ою=ойұу, ұю=ұйұу, ыю=ыйұу, үю=үйүу, ею=ейүу, қию=қыйұу, #тию#=тійүу, кию=кійүу, #сию#=сыйұу, #жию#=жыйұу, а^ию=а^ыйұу, о^ию=о^ұйұу, ұ^ию=ұ^ұйұу, ы^ию=ы^ыйұу, ә^ию=ә^ійүу, ө^ию=ө^үйүу, ү^ию=ү^үйүу, і^ию=і^ійүу, е^ию=е^ійүу;
4. ая=айа, оя=ойа, ұя=ұйа, ыя=ыйа, қия=қыйа, #сия=сыйа, #жия=жыйа, #мия=мыйа, #зия=зыйа, а^ия=а^ыйа, о^ия=о^ұйа, ұ^ия=ұ^ұйа, ы^ия=ы^ыйа, ә^ия=ә^ійә, ү^ия=ү^үйә, ия^а=ыйа^а;
5. #ми=мый, #жи=жый, а^и=а^ый, о^и=о^ый, ұ^и=ұ^ый, ы^и=ы^ый, ә^и=ә^ій, ө^и=ө^ій, ү^и=ү^ій, і^и=і^ій, е^и=е^ій, и^а=ый^а, и^о=ый^о, и^ұ=ый^ұ, и^ы=ый^ы, и^ә=ій^ә, и^ө=ій^ө, и^ү=ій^ү, и^і=ій^і, и^е=ій^е, қи=қый, ғи=ғый, иқ=ыйқ, иғ=ыйғ;
6. а^у=а^ұу, о^у=о^ұу, ұ^у=ұ^ұу, ы^у=ы^ұу, ә^у=ә^үу, ө^у=ө^үу, ү^у=ү^үу, і^у=і^үу, е^у=е^үу, у^а=ұу^а, у^о=ұу^о, у^ұ=ұу^ұ, у^ы=ұу^ы, у^ә=үу^ә, у^ө=үу^ө, у^ү=үу^ү, у^і=үу^і, у^е=үу^е;
7. о^ы=о^ұ, ұ^ы=ұ^ұ, ө^і=ө^ү, ү^і=ү^ү;
8. ө^е=ө^ө, ү^е=ү^ө;
9. і^а=і^ә, ү^а=ү^ә, ө^а=ө^ә;
10. сш=шш, сж=шш, зш=шш, шс=шш, щ=шш;
11. зж=жж, зс=сс;
12. нб=мб, нп=мп;
13. нг=ңг, нғ=ңғ, нк=ңк нқ=ңқ;
14. мл = міл, ғн=ғын, ғл=ғыл;
15. шб=шп, тб=тп, тж=тш, пб=пп, пд=бд, пж=бж, пм=бм, пн=бн, сб=сп, сд=ст, қб=қп, қг=кг, қд=ғд, қж=ғж, қз=ғз, қм=ғым, қн=ғын, кб=кп, кг=кк, кд=кт, кж=гж, кз=гз, км=гм, кн=гн, зк=зг, зп=зб, зт=ст, қл=ғыл.

Each substitution rule is composed of two parts separated by a sign «=». From the left of this sign are original alphabetic character for word recording, on the right - the characters that should be replaced in the transcription.

For transcription of the given word consistently searches next occurrence of the left part of rule in it. If any of it detected, then instead of it, inserted the right part of the rule.

As a transcription symbols for vowels used mainly relevant Kazakh letters. Solid consonants are transcribed as Kazakh letters and the relevant soft consonants with the analogous Latin letters.

«#»means the beginning or end of word, depending on the location of: if«#» standing in front of characters, then it is the beginning of word; if «#» stands after the characters, it's the end.

«^»means any characters in any number between two sounds.

Each substitution rule is composed of two parts separated by a sign «=». From the left of this sign are original alphabetic character recording word on the right - the characters that should be replaced in the transcription.

For transcription of given word consistently searches next occurrence of the left part of rule in it. If any of it is detected, then it is inserted along the right side of the rule.

It is recommended that in the control file of these groups in numerical order, without changing the order of the rules in groups, because the order of substitutions is obviously important.

Also orthoepic rules were included to transkriptor rules defining the softness and labial consonants:

1. әб=әb, әг=әg, әд=әd, әж=әv, әз=әz, әй=әj, әк=әk, әл=әl, әм=әm, ән=ән, әң=әq, әп=әf, әр=әr, әс=әs, әт=әt, әу=әu, әш=әw, еб=eb, ег=eg, ед=ed, еж=ev, ез=ez, ей=ej, ек=ek, ел=el, ем=em, ен=en, ең=eq, еп=ef, ер=er, ес=es, ет=et, еу=eu, еш=ew, іб=ib, іг=ig, ід=id, іж=iv, із=iz, ій=ij, ік=ik, іл=il, ім=im, ін=in, ің=iq, іп=if, ір=ir, іс=is, іт=it, іш=iw, бә=bə, гә=gə, дә=də, жә=və, зә=zə, йә=jə, кә=kə, лә=lə, мә=mə, нә=nə, ңә=qə, пә=fə, рә=rə, сә=sə, тә=tə, уә=uə, шә=wə, бе=be, ге=ge, де=de, же=ve, зе=ze, йе=je, ке=ke, ле=le, ме=me, не=ne, ңе=qe, пе=fe, ре=re, се=se, те=te, ше=we, бі=bi, гі=gi, ді=di, жі=vi, зі=zi, йі=ji, кі=ki, лі=li, мі=mi, ні=ni, ңі=qi, пі=fi, рі=ri, сі=si, ті=ti, ші=wi.

2. өб=өb, өг=өg, өд=өd, өж=өv, өз=өz, өй=өj, өк=өk, өл=өl, өм=өm, өн=өn, өң=өq, өп=өf, өр=өr, өс=өs, өт=өt, өу=өu, өш=өw, үб=үb, үг=үg, үд=үd, үж=үv, үз=үz, үй=үj, үк=үk, үл=үl, үм=үm, үн=үn, үң=үq, үп=үf, үр=үr, үс=үs, үт=үt, үу=үu, үш=үw, бө=bө, гө=gө, дө=dө, жө=vө, зө=zө, йө=jө, кө=kө, лө=lө, мө=mө, нө=nө, ңө=qө, пө=fө, рө=rө, сө=sө, тө=tө, уө=uө, шө=wө, бү=bү, гү=gү, дү=dү, жү=vү, зү=zү, йү=jү, кү=kү, лү=lү, мү=mү, нү=nү, ңү=qү, пү=fү, рү=rү, сү=sү, тү=tү, уү=uү шү=wү;

Let us explain marks used in the replacement rules. Latin characters in a group of 16 means that the sound is soft and non-labial, 17 in group "2" after the consonant means that the consonant - solid lip and in the group of 18 the number "3" after the consonant means that the consonant - soft labial.

In general phonetic transkriptor was about 400 rules. Similar rules have been further used for the transcription of sounds borrowed from the Russian language, formalized in [1].

Description of the program: The program involves automating the process of forming the texts of the material with a given phonetic units of representativeness (in particular diphones). This is necessary for development of automatic speech synthesis. Thus, using this program, you will be able to get a text material with full coverage of all possible diphones in Kazakh language.

Of course, we are not talking about the original generation of coherent text – we cannot afford this machine. This refers to the layout of the text matching words contained in pronouncing dictionary. Selection of the program will be implemented in such a way as to minimize redundancy (repetition) of units in the text.

Thus, the program allows the following tasks:

- the generation of the primary list of diphones submitted by Alphabet;

- checking for the specified diphones in the pronouncing dictionary, and a definitive list of the correct diphones;
- generation of mini-set of the text covering all sorts of diphones;
- evaluation of the degree of phonetic in formativeness of the words included in the text material.

As mentioned above, the program can be used for virtually any Turkic language, without the formalization of phonological rules, replacing the spelling dictionary to the pronouncing dictionary. For Kazakh language, we used the spelling dictionary of 45,000 word forms.

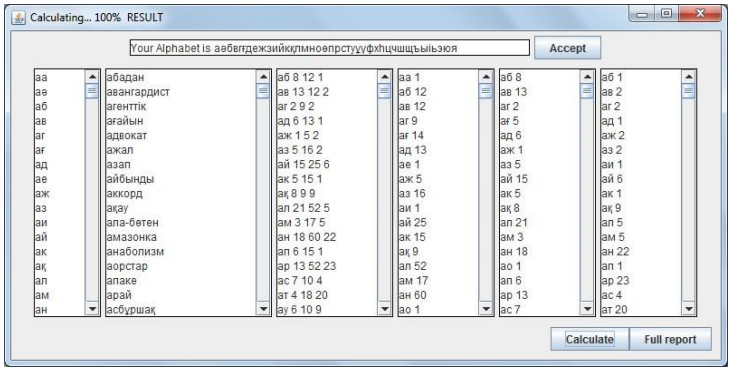


Fig. 2. Text analyzer's window

The first field is displayed diphones occurring at the beginning, in the middle and at the end of the words in the second - found only at the beginning of the word in the third - found only in the middle of the word in the fourth - found only at the end of the words, with appropriate statistics. By clicking on the "Full report" we get a detailed report with a set of text.

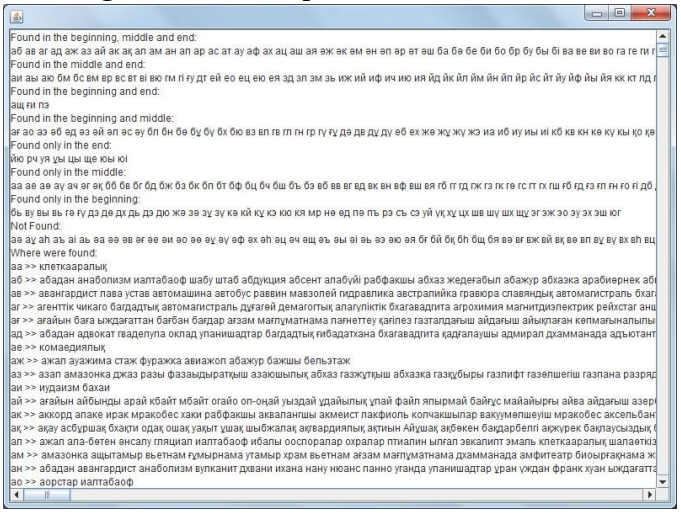


Fig. 3. Full report's window

A selection of the text carried by the following algorithm:

1. Get over the words for a particular diphones;
2. The words are divided into diphones;
3. Each diphone in words is checked by a priority (the number of occurrences of words chosen before);
4. Selection of the words with the lowest priority;
5. New priorities are assigned to each diphone in the selected word.

Of course, there are common diphones, for example, diphone «ан» in the text set occurs 100 times, when diphone «ае» occurs only 1 time. Yet, due to the method of assigning priorities to phonetic units, we obtain a set of text less redundant and compact.

Results:

Quantity of diphones in the primary list 412=1681;

After checking on existence in their pronouncing dictionary remained – 1003;

From them:

Meeting at the beginning, in the middle and at the end of the word – 297;

Meeting at the beginning and in the middle of the word – 144;

Meeting at the beginning and at the end of the word – 3;

Meeting in the middle and at the end of the word – 159;

Meeting only at the beginning of the word – 45;

Meeting only in the middle of the word – 347;

Meeting only at the end of the word – 8; Quantity of words in a text set – 6131.

Found at the beginning, in the middle and at the end:

аб ав аг ад аж аз ай ак ақ ал ам ан ап ар ас ат ау аф ах ац аш ая әж әк әм ән әп әр әт әш ба бә бе би бо бр бу бы бі ва ве ви во га ге ги го гу ға ғы да де дж ди до др ду ды ді ев ег ед еж ез ек ел ем ен еп ер ес ет еу еф еш жа же жи жо жу жы жі за зе зи зо зу зы зі ив иг ид ие из ии ик ил им ин ио ип ир ис ит их иш йе йо йі ка ке ки кл ко кр кс ку кх кі қа қи қу қы ла ле ли лл ло лу лы ль ма мб ме ми мн мо му мы мі на нә не ни но ну ны ні нь ню об ов ог од ож оз ой ок оқ ол ом он оо оп ор ос от оф ош ою өз өк өл өн өс па пе пи по пр пс пт пу пы пі ра рә ре ри ро ру ры са се си ск см со сс ст су сы сі сь та те ти то тр ту ты ті ть уа уг уе уз уи ук ул ум ун уп ур ус ут уш уы уі ұз ұқ ұл ұм ұн ұп ұр ұс ұя үз үй үк үл үн үр үс үт үш үю фа фе фо фр фт фь ха ца ча чи ша ше ши шт шу шы ші ыб ыз ық ыл ым ын ыр ыс ыт ыш із іл ін ір іс іш эз эл эн эр эт юк юм юп юр яг яд як ян яр яс яш

Found in the middle and at the end:

аи аы аю бм бс вм вр вс вт ві вю гм гі ғу дт ей ец ею ея зд зл зм зь иж ий иф ич ию ия йд йк йл йм йн йп йр йс йт йу йф йы йя кк кт лд лк лқ лп лс лт лі ля мм мп мс мт нг нд нж нз нк нн нр нс нт нч нш ня өй өп пп рб рв рг рд рк рқ рл рм рн рп рс рт рұ рф рх рц рш рщ рі рь рю тл тм тс тт тч уб уд уу уф ух фм фф цо шь щы ый ык ып ік ің ім іп іт іу ьб ьд ье ьз ьк ьм ьн ьс ьт ьф ья юд юз юй юс ют юф яж яз яй яқ ял ям ят яу

Found at the beginning and at the end:

ащ ғи пэ

Found at the beginning and in the middle:

аға ао аэ әб әд әз әй әл әс әу бл бн бө бұ бү бх бю вз вл гв гл гн гр гү ғұ дә дв дұ дү еб ех жө жұ жү жэ иа иб иу иы иі кб кв кн кө кү кы қо қө құ қү лә лұ лү лю мә мө мұ мү мэ мю нұ нү оғ ох оц оч оя өб өг өж өм өр өт өш пл пн пұ пь рұ сә св сл сн сө сп сұ сү сф сх сц сю тә тв тө тұ тү тю уә ув уқ уо уч ұғ ұд ұж ұй ұт ұш үб үг үғ үд үж үм үп фи фл фу хе хи хл хо хр ху хэ це ци цу че чу шә шк шл шм шн шо шө шп шр шұ ығ ыд ыж іб эб эв эд эй эк эм эп эс эф юб юв юл юн яп ях

Found only at the end:

йю рч уя ұы цы ще юы юі

Found only in the middle:

аа ае аө аү ач әг әқ бб бв бг бд бж бз бк бп бт бф бц бч бш бъ бэ вб вв вг вд вк вн вф вш вя гб гг гд гж гз гк гө гс гт гх гш ғб ғд ғз ғл ғн ғо ғі дб дг дғ дд дк дл дм дн дп дс дц дш дъ еа еә еғ ее еи ең еө еұ еү еч ещ еі еэ жб жғ жд жк жн жс жт жш зб зв зг зғ зж зз зк зқ зн зө зп зр зс зт зх зш зъ зю иә иғ иқ иө иц ищ иэ йа йб йв йг йғ йж йз йқ йө йұ йх йц йш йэ кг кд кж кз кқ км кп кф кц кч кш кб кғ қд қе қж қк ққ қл қм қн қп қр қс қт кф қх қш лб лв лг лғ лж лз лм лн лө лф лх лц лч лш лэ мв мг мғ мд мж мз мк мқ мл мф мх мш мь нб нв нғ нқ нл нм нп нф нх нц нъ нэ оа оә ое ои оө оу оұ оү оы оі пб пв пд пж пк пқ пм пф пх пц пч пш пю рғ рж рз рө рр ръ ря сб сг сд сж сз сқ ср сш ся тб тг тғ тд тж тз тк тқ тн тп тф тх тц тш тэ тя уғ уж уө уұ уү уц уэ ұб ұх ұщ ұю үа үә үе үф үх үы фв фг фк фс фш фы фя хб хв хг хм хн хс хт хх хш хы хь цг цд цз цк цл цм цт цц чж чк чм чт чч шб шг шғ шд шж шқ шс шш шю ща щи ъе ъю ыа ыв ыг ые ыи ыо ыұ ых ыщ ыч ыя іг ід іж іи іо іх ьв ьг ьл ьо ьп ьр ьх ьц ьч ьш ьэ ью юа юж юи юо юц юч юш юэ яб яв яғ яц

Found only at the beginning: бь ву вы вь гә ғү дз дө дх дь дэ дю жә зә зұ зү кә кй құ кэ кю кя мр нө өд пә пъ рэ съ сэ

уй үқ хұ цх шв шү шх щұ эг эж эо эу эх эш юг

It is very important to know item distributions of vowel sounds in a word as in Kazakh language the accent falls on a final syllable. It considerably simplifies and reduces diphone base. For example, the sound **О** meets only at the beginning of a word and all diphones with sound participation **О** have no shock analogs.

Vowel sounds meeting at the beginning and in the middle of a word sound almost equally that also does diphone base of more compact.

Conclusion: Further development of the program involves automating the generation of other phonetic units (allophones, Trifonov, syllables). The criterion can serve not only phonetic (segment) text properties, but also communicative and syntactic component (for punctuation): for example, you can specify the proportion of interrogative sentences, or sentences containing nonfinal syntagma (by commas). Obviously, the efficiency of selection depends on the parameters of the reference body text: the more complete and varied it is, the more informative and compact it will be obtained on the basis of textual material.

It is also planned to automate the analysis of statistics of occurrence in the text of the speech units of different levels: phonemes, allophones, syllables, sound sequences of words, the development of automatic phonetic transkriptor Kazakh speech text to speech synthesizer Kazakh.

References

1. Shelepov V.U. Leksii o raspoznavanii rechi. – Donetsk: IPSHI «Nauka i osvita», 2009. – 196 p.

УДК 81’33

МНОГОФУНКЦИОНАЛЬНАЯ МОДЕЛЬ ТЮРКСКОЙ МОРФЕМЫ: БАЗОВЫЕ ФОРМАЛИЗМЫ

Сулейманов Джавдет Шевкетович *д.т.н., директор, Институт прикладной семиотики Академии наук Республики Татарстан, Казанский федеральный университет. E-mail: dvdt.slt@gmail.com*

Гатиатуллин Айрат Рафизович *к.т.н., зав.отделом, Институт прикладной семиотики Академии наук Республики Татарстан. E-mail: agat1972@mail.ru*

Статья является продолжением описания многофункциональной модели тюркской морфемы, которая включает структурно-функциональную модель морфем, базу данных модели и информационно-программную оболочку, технологический инструментарий, для заполнения базы данных. В статье рассматриваются базовые формализмы, которые включают выражения для алломорфов, морфем и семантем тюркских языков. Многофункциональная модель может быть использована, как ресурсная база для программных продуктов, осуществляющих компьютерную обработку тюркских языков, информационно-справочная система и инструментарий для исследований ученых-тюркологов, в частности, для сравнительного анализа тюркских языковых единиц.

Ключевые слова: многофункциональная модель тюркской морфемы, морфема, алломорф.

MULTIFUNCTIONAL COMPUTER BASED LINGUISTIC MODEL: FORMALIZMES

Dzhavdet Suleymanov *Tatarstan Academy of Sciences Institute of Applied Semiotics, Kazan Federal University E-mail: dvdt.slt@gmail.com*
Ayrat Gatiatullin *Tatarstan Academy of Sciences Institute of Applied Semiotics, E-mail: agat1972@mail.ru*

This article is a continuation of the description of the multi-functional model of Turkic morpheme, which includes structural and functional model of morphemes, the database model and the information and the program shell, the technological tools to fill the databases. The article discusses the basic formalisms which include expressions for allomorphs, morphemes and semantemes of Turkic languages. The multifunctional model can be used as a resource base for the software which carries out the computer processing of the Turkic languages; information and reference system and tools for researches of Türkologists, in particular, for the comparative analysis of the Turkic language units.

Keywords: multifunctional model of Turkic morpheme, morpheme, allomorph.

Введение

В последние 25 лет наблюдается значительное увеличение количества разработок, а также публикаций в области компьютерной обработки тюркских языков [1, 6, 7]. Этому способствует как интерес к информации на тюркских языках в Интернет-пространстве, так и научно-исследовательская и прикладная активность носителей языка по компьютерной поддержке тюркских языков. Как показывает анализ имеющихся работ, наименее развитым и представленным в публикациях является отсутствие соответствующего формального инструментария, адекватной спецификации и семиотических (лексико-грамматических и семантических) моделей, отражающих структурно-функциональные особенности тюркских языков. Очевидно, целый ряд вопросов, возникающих у разработчиков программного инструментария для работы с тюркскими языками, связанных с неопределенностью понятий, многозначностью языковых единиц, могли бы разрешаться уже на этапе формального описания моделей.

Морфологическая и синтаксическая близость между тюркскими языками позволяет описывать соответствующие языковые явления схожими лингвистическими моделями и использовать для их обработки практически один и тот же программный инструментарий. Очевидно, в основе своей на 70-80 и более процентов они являются общими для всех тюркских языков, как при разработке структуры и функционала электронных корпусов языков, грамматических анализаторов, так и машин поиска и машинных переводчиков. А это

означает, что возможно создать единую модель, наиболее полно отражающую лингвистические особенности тюркских языков.

Авторы статьи считают, что одним из таких объединяющих механизмов может стать многофункциональная модель тюркской морфемы. В данной статье предлагаются базовые формализмы этой модели.

1. Многофункциональная модель

Многофункциональная модель тюркской морфемы включает структурнофункциональную модель морфем [2], базу данных модели и информационно-программную оболочку, технологический инструментарий, для заполнения базы данных.

Способы применения многофункциональной модели:

- в качестве ресурсной базы для программных продуктов, осуществляющих компьютерную обработку тюркских языков.
- в качестве информационно-справочной системы, содержащей практически полную информацию о тюркских языковых единицах – морфемах.

– в качестве инструментария для исследований ученых-тюркологов, в частности, для сравнительного анализа тюркских языковых единиц.

Использование в качестве подобной ресурсной базы именно модели морфем обусловлено исключительной значимостью морфологического языкового уровня при обработке естественно-языковых текстов. Это особенно актуально для языков агглютинативного типа с богатой морфологией, к которым относятся все языки тюркского семейства.

Зададим базовые формализмы модели.

Пусть L_i – языки тюркской группы.

Например:

L_1 – татарский язык, L_2 – казахский язык, L_3 – турецкий язык и т.д.

Обозначим Σ_i – алфавит языка L_i. $a_{(i,j)} \in \Sigma_i$ - символ языка L_i.

Например, алфавит языка L_1 (татарского языка): Σ_1 = {‘аәбвгдеёжҗзийклмнңоөпрстуүфхһцчшщъыьэюя’}

Расширим алфавит языка L_i пустым символом (пробелом) e = {‘ ’}.

Σ_i’ = Σ_i + {e} – (расширенный) аналитический алфавит языка L_i

Тогда словоформа языка L_i:

$w_{(i,k)} = (a_{(i,1)}\ a_{(i,2)}\ a_{(i,2)}\ ..\ a_{(i,n)})$ – k-е слово языка L_i, $a_{(i,j)} \in \Sigma_i$; $a_{(i,j)} \neq e$

Например:

$w_{(1,k)}$ = ‘баралар’ – k-е слово татарского языка

Множество всех возможных слов языка L_i обозначим Σ_i*.

2. Алломорфы

Рассмотрим обозначения морфологических единиц языка. В.А. Плунгян в своей работе [4] определяет три модели морфологии:

1) элементарно комбинаторная;

2) элементарно процессная; 3)

словесно парадигматическая.

Элементарно комбинаторная модель – это модель, основным инструментом которой является линейная сегментация. В языках с элементарно процессной моделью морфологии некоторые алломорфы рассматриваются как исходные, а другие — как производные, которые могут быть получены из первых путем применения различных операций типа

«фонологических процессов». В словесно парадигматических моделях вообще происходит отказ от морфемного членения при описании словоизменения. Именно словоформа и оказывается минимальной единицей грамматического описания в парадигматической модели.

Согласно этой классификации тюркские языки относятся к языкам элементарнокомбинаторного типа, комбинируемыми элементами являются морфы и алломорфы.

Мельчук [3] дает им следующее определение:

Морфа (морф) – элементарный сегментный знак – минимальная сущность с одним и тем же морфологическим поведением, которое достаточно случайным образом связано с ее формой.

Алломорф - совокупность морфов одной морфемы, имеющих одинаковый фонемный состав.

В нашей модели будем использовать в качестве элементов комбинаторики – алломорфы:

$m_{(i,j)} = (a_{(i,1)}\ a_{(i,2)}\ a_{(i,2)}\ ..\ a_{(i,k)})$ – алломорф с номером j языка L_i, где $a_{(i,j)} \in \Sigma_i$

Плунгян [4] определяет словоформы как морфемный комплекс, между составными частями которого существуют особенно тесные связи — на порядок более тесные, чем между самими этими комплексами.

$w_{(i,k)} = (m_{(i,1)}\ m_{(i,2)}\ m_{(i,3)}\ ..\ m_{(i,p)})$ - слово языка L_i из p – алломорфов, где $p \geq 1$.

Если p=1, то словоформа состоит из одного алломорфа.

Каждый алломорф обладает рядом определенных свойств. Введем обозначения для этого набора свойств:

$C(m_{(i,j)}) = \{c_1(m_{(i,j)}), c_2(m_{(i,j)}), c_3(m_{(i,j)}), \ldots c_t(m_{(i,j)})\}$ – свойства алломорфа $m_{(i,j)}$

Одними из основных свойств алломорфов являются свойства, описывающие контекст этого алломорфа. Эти свойства делятся на два типа контекст алломорфа в пределах словоформы и контекст алломорфа за пределами словоформы.

Определим свойства 1 и 2 как свойства сочетания алломорфов в словоформе. Эти свойства определим множествами алломорфов, которые могут встретиться в словоформе слева и справа от искомого алломорфа:

$c^1(m_{(i,j)}) = \{ m_{(i,1)}, m\ m_{(i,2)}, \ldots m\ m_{(i,k)}\}$ – множество алломорфов, которые могут следовать в словоформе слева $c^2(m_{(i,j)}) = \{ m_{(i,1)}, m\ m_{(i,2)}, \ldots m\ m_{(i,k)}\}$ – множество алломорфов, которые могут следовать в словоформе справа

Например:

c^2(‘га’) = {‘мы’, ‘мыни’, ‘дыр’, ‘рак’, е} c^2(‘ка’) = {‘мы’, ‘мыни’, ‘дыр’, ‘рак’, е}

Наличие в свойстве символа е показывает, что алломорф может быть последним в словоформе. Если символ е отсутствует в свойстве, то этот алломорф не может быть последним в словоформе.

Например:

c^2(‘китап’) = {‘ка’, ‘тан’, ‘та’, …, e} c^2(‘китаб’) =
{‘ым’, ‘ың’, ‘ы’, ‘ын’, ‘ыбыз’, ‘ыгыз’}

Если $c^1_i(m_{(i,j)}) = \{e\}$, то этот алломорф может быть в словоформе только первым.

Если $c^2_i(m_{(i,j)}) = \{e\}$, то этот алломорф может быть в словоформе только последним.

Например: c^2(‘мы’) = {e}

Обозначим $Mi^* = \{m_{(i,j)}\}$ – множество всех алломорфов языка L_i.

$M_i^* = M_i' + M_i''$

$M_i' = \{m_{(i,j)}: c^1(m_{(i,j)}) = \{e\}\}$ – множество всех алломорфов языка L_i, которые могут стоять в словоформе только первыми.

$M_i'' = \{m_{(i,j)}: e \notin c^1(m_{(i,j)})\}$ – множество всех алломорфов языка L_i, которые не могут стоять первыми в словоформе.

3. Семантемы

У Мельчука языковая единица определяется тройкой: означающее, означаемое и синтактика. Соответственно, каждому алломорфу соответствует определённое значение, называемое семантемой: $m_{(i,j)} \rightarrow s_k$ или $m_{(i,j)}(s_k)$.

Введем обозначение $S^* = \{s_k\}$ – множество всех семантем языка, которые можно выразить отдельными алломорфами.

Множество семантем делится на подмножества, в соответствии с обозначаемыми концептами: $S^* = E + R$

E – множество всех сущностей языка; R – множество отношений между сущностями

$E = O + V + A$

O – множество объектов; A – множество действий; V – множество значений

Алломорфы, используемые для выражения одной и той же семантемы являются синонимами.

Например:

L_1: $m_{(1,j1)}(s_k)$ = ‘сандугач’; $m_{(1,j2)}(s_k)$ = ‘былбыл’

L_2: $m_{(2,j3)}(s_k)$ = ‘бұлбұл’

L3: $m_{(3,j4)}(s_k)$ = ‘bülbül’

4. Морфемы

Морфема в лингвистике определяется, как минимальная значащая часть слова, совокупность морфов (алломорфов), имеющих одинаковое значение и ряд других общих признаков:

$M_{(i,j)}(s_k) = \{m_{(i,j1)}(s_k) .. m_{(i,jp)}(s_k): c_k(m_{(i,j1)}(s_k) = c_k(m_{(i,j2)}(s_k)\}$.

Этим другим свойством, согласно работам московской фонологической школы, является свойство регулярного чередования [3].

В каждом из тюркских языков эти множества чередований свои:

$G(L_i) = G_i = \{G_{(i,1)}, G_{(i,2)}, \ldots G_{(i,t)}\}$

Например:

G_1 = {{‘г’,’к’}; {‘д’,’т’}; {‘л’,’н’}; {‘а’,’ә’}; {‘у’,’ү’}; {‘б’,’п’}; … } – чередования в татарском языке.

Подставив требование чередования к определению морфемы, получаем следующую формулу:

$M_{(i,j)}(s_k) = \{m_{(i,j1)}(s_k) .. m_{(i,jp)}(s_k): (a_{(i,j1)} \in m_{(i,j2)}) \& (a_{(i,j1)} \in m_{(i,j2)}): (a_{(i,j1)} = a_{(i,j2)}) \wedge ((a_{(i,j1)} \in G_{(i,p)}) \& (a_{(i,j2)} \in G_{(i,p)})\}$

Согласно этой формуле, одной морфеме принадлежат только алломорфы с чередованием фонем. Таким образом, алломорфы -к и -быз оба являются показателями предикативности 1 лица множественного числа, однако они не будут принадлежать одной морфеме, так как у них нет свойства чередования.

При p = 0, получается пустая морфема, а при p=1 морфема, состоящая из одного алломорфа.

Если следовать определению морфемы, приведенному выше, то морфема –[Г]А в следующем примере - это разные морфемы –[Г]А с одинаковым обозначением (знаком).

Мин урманга гөмбәгә ике сәгатькә барам. Я в лес за грибами на два часа пойду.

В нашей модели морфем предлагается считать одной морфемой те варианты, которые совпадают по правилам чередования и синтактике, но различаются по выражаемым семантемам (т.е. многозначная морфема).

Тогда согласно нашей модели –[Г]А в предыдущем примере, это одна и та же морфема –[Г]А, используемая для выражения значения разных семантем. А в словоформах барды-лар и алма-лар это алломорфы разных морфем –ЛАр$_1$ и –ЛАр$_2$, так как у них разные синтактики, то есть правила следования в словоформе.

Обозначим M_i** - множество всех морфем языка L_i

5. Корневые и аффиксальные морфемы

Как отмечает Плунгян [4], различие между корневыми и аффиксальными морфемами представляется интуитивно очевидным, но в действительности оно с трудом поддается формализации. Одним из основных признаков, отличающих корень и аффикс, является то, что корень может образовывать словоформу, а аффикс - нет. Другая формулировка такова: словоформа может состоять только из одного корня, но не может состоять только из аффиксов.

Согласно морфологической классификации языков А.П. Володина [5] все языки делятся на две большие группы:

1. Языки, в которых представлена алтайская линейная модель словоформы – слово всегда начинается с корня. Алтайская модель имеет два вида:
1. алтайская 1: R+(m) - запрещена префиксация, композиция и инкорпорация (тюркские, эскимосский);
2. алтайская 2: (r)+R+(m) - запрещена префиксация, в некоторых языках (финноугорские, корейский, японский) разрешена инкорпорация существительного.
2. Языки, в которых представлена американская линейная модель словоформы - слово не всегда начинается с корня. Американская модель также имеет три вида:
1. американская 1: (m)+(r)+R+(M) - айну, нивхский, чукотский, кетский, баскский, шумерский, многие языки американских индейцев;
2. американская 2: (m)+(r)+R+F+(m) - индоевропейские языки, F - флексия;
3. американская 3: (m)+R+(M) - ительменский, некоторые из языков американских индейцев.

Здесь R, r - корневые морфемы, M, m - аффиксы, Скобки означают, что данный элемент может быть представлен в конкретной словоформе более одного раза или вообще отсутствует.

Таким образом для тюркских языков $M_i' = \{m_{(i,j)}: c^1(m_{(i,j)}) = \{e\}\}$, определенное выше, совпадает с множеством корневых алломорфов.

Множество $M_i'' = \{m_{(i,j)}: e\ !\epsilon\ c^1(m_{(i,j)})\}$ – совпадает с множеством аффиксальных алломорфов.

Аналогично алломорфам все морфемы имеют конечный набор свойств. Свойства морфем формируются из свойств алломорфов, из которых состоит морфема:

$C(M_{(i,j)}) = \{c_1(m_{(i,j)}), c_2(m_{(i,j)}), c_3(m_{(i,j)}), \dots c_t(m_{(i,j)})\}$ – свойства морфемы $M_{(i,j)}$ языка L_i.

Совокупность всех тюркских морфем, их свойств, семантем и соответствий между морфемами и семантемами образует многофункциональную модель тюркской морфемы:

$MM = \{ M_{(i,j)}, C(M_{(i,j)}), s_k\}$

Заключение

В статье дается описание базовых элементов многофункциональной лингвистической модели тюркских морфем. Весьма конструктивным и продуктивным представляется использование данной многофункциональной и многоязычной модели тюркских морфем в качестве одного из центральных, ядерных, модулей в едином веб-портале для тюркских языков. Авторы статьи выражают также надежду, что данный проект послужит интеграции усилий ученых-тюркологов для расширения базы данных описаниями различных тюркских языков, что обеспечит эффективное использование многофункциональной модели в качестве технологического инструментария и межязыкового модуля в системах компьютерной обработки тюркских языков.

Список литературы

1. Труды Первой международной конференции «Компьютерная обработка тюркских языков». – Астана: ЕНУ им. Л.Н. Гумилева, 2013. – 345с.
2. Сулейманов Д.Ш., Гатиатуллин А.Р. Структурно-функциональная компьютерная модель татарских морфем. – Казань: Фэн, 2003. – 220с.
3. Мельчук И.А. Курс общей морфологии. Т. IV. / Пер. с фр. Е.Н. Саввиной под общ.ред. Н.В. Перцова. – М.: Вена: Языки славянской культуры: Венский славистический альманах, 2001. – 584с.
4. Плунгян В.А. Общая морфология: Введение в проблематику: Учебное пособие. М.: Эдиториал УРСС, 2000. – 384с.
5. Володин А.П. Палеоазиатские языки // Языки мира М., 1996.
6. Proceedings of the International Conference on Turkic Language Processing (TURKLANG-2014). (Istanbul, November 6-7, 2014). - Istanbul: Özkaracan MatbaacılıkBağcılarr, 2014. – 135 pp.
7. Proceedings of the International Conference “Turkic Languages Processing: TurkLang2015”. – Kazan: Academy of Sciences of the Republic of Tatarstan Press, 2015. – 488 pp.

UDC 004.8+81.322.2

SENTIMENT ANALYSIS OF KAZAKH PHRASES BASED ON MORPHOLOGICAL RULES

Yergesh Banu, *L.N.Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: b.yergesh@gmail.com*

Sharipbay Altynbek, *L.N.Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: sharalt@mail.ru*

Bekmanova Gulmira, *L.N.Gumilyov Eurasian National University, Astana, Kazakhstan, E-mail: gulmira-r@yandex.ru*

Lipnitskii Stanislav, *United Institute of Informatics Problems NAS Belarus, Minsk, Belarus*
E-mail: lipn@newman.bas-net.by

Abstract. Sentiment analysis of texts in natural languages is one of the fastest growing technologies of natural language processing. There is no any study has been carried out on the sentiment analysis Kazakh texts, to date. We examined the texts in the Kazakh language and identified the parts of speech, which determines the text sentiments. In this work we described the linguistic approach for sentiment analysis of phrases in the Kazakh language, based on the morphological rules.

Keywords: sentiment, tonality, sentiment analysis, Kazakh language, classification, production rules, morphological rules

АНАЛИЗ ТОНАЛЬНОСТИ КАЗАХСКИХ ФРАЗ НА ОСНОВЕ МОРФОЛОГИЧЕСКИХ ПРАВИЛ

Ергеш Бану, *Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, E-mail: b.yergesh@gmail.com*

Шарипбай Алтынбек, *Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, E-mail: sharalt@mail.ru*

Бекманова Гульмира, *Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан, E-mail: gulmira-r@yandex.ru*

Липницкий Станислав, *Объединенный институт проблем информатики НАН Беларуси, Минск, Беларусь, E-mail: lipn@newman.bas-net.by*

Ключевые слова: сентимент, сентимент анализ, казахский язык, анализ тональности, классификация по тональности, морфологические правила.

Аннотация. Сентимент анализ текстов на естественном языке один из быстро развивающих технологий обработки естественного языка. До этого времени не проводились работы по сентимент анализу казахских текстов. Мы изучили тексты на казахском языке и определили части речи, которые придают сентимент тексту. В данной работе описывается лингвистический подход сентимент анализа текстов на казахском языке, основанный на морфологических правилах.

1 Introduction. Sentiment analysis of the text is a class of content analysis methods in computational linguistics, designed for the automated detection of emotional vocabulary and emotional evaluation used by the author, in order to understand his/her attitude (opinion) to the particular object. The emotional component expressed at the lexeme level or a communicative fragment is called a lexical sentiment or tonality. The tonality of the whole text generally can be determined as a function (in the simplest case, the sum) of lexical sentiment that constitutes its units (sentences) and the rules of their collocations [1, 2].

The systems of sentiment analysis have been widely accepted in the commercial and social spheres. It can be noticed that the number of blogs, reviews, forums, web pages of social networks are growing by the day in worldwide network. Therefore, manual processing such a big amount of data becomes impossible, thus different linguistic and machine learning methods are used [1-11].

The sentiment analysis of texts written in Kazakh language has not been studied yet. This work can be considered as an introduction and an attempt to apply the linguistic approach for sentiment analysis of the texts written in the Kazakh language. For that reason, this paper describes

the main methods used in sentiment analysis and approaches used to determine the sentiment of Kazakhs phrases by formalizing the morphological rules.

2 The main methods use in sentiment analysis

According to the work described in [1] the automatic analysis of a sentiment of texts in the natural language carried out by applying the methods such as *machine learning methods and linguistic methods*.

The sentiment analyses of tonality based on machine learning methods are "trained" on a collection of pre-marked texts. These methods include a support vector machine (SVM), logistic regression, naive Bayes classifier, maximum entropy, k nearest neighbor (k-NN) and other methods.

Linguistic methods usually use morphological analysis, specifically designed sentiment dictionaries of words and phrases as well as set of linguistic rules [11].

3 Determining the sentiment of phrases in Kazakh language

Determination of sentiment of phrases in Kazakh language is based on a classification of texts by two features: *positive* (POS or 1) and *negative* (NEG or 0). For this purpose, a dictionary of sentiment words was developed which participates in determining the tonality of the text. In Kazakh language the tonality of a phrase is given by parts of speech as an *adjective*, *verb* and *adverb*. After that, morphological rules of parts of speech are formalized that are involved in determining the tonality of the sentence: words and/or phrases are extracted from the sentence which contains evaluative words. The overall sentiment of the text is evaluated according to the sentence/phrase tonality.

By conducting the analysis of the Kazakh texts it has been identified that adjectives mainly determine the semantic orientation (tonality) of the text, and the noun plays a role of an aspect (object) of discussion. From the extracted phrases we can determine the tonality of the whole text. The tonality of evaluative words might depend on the context and subject area. Also, the tonality can be changed or intensified depending on *adverbs*, *verbs* and *conjunctions*. According to research, the following phrases can be used to define the tonality:

1) [ADJECTIVE]+[NOUN]
2) [ADJECTIVE] + [Negation]+[NOUN]
3) [ADJECTIVE] + [VERB]
4) [ADJECTIVE]+[VERB]+ [Negation]
5) [ADVERB] + [ADJECTIVE]

For determining of morphological features we used work described in [12]. Here we explained how semantic hyper-graphs are used to describe ontological models of morphological rules of the Kazakh language. On the basis of this work, morphological analyzer was built. This morphological analyzer is used for extraction of morphological information from texts.

4 Rules used to define the sentiment

Here we are giving some rules for defining the phrases sentiment in the Kazakh language. They described using production rules. For that we introduce the following meta notations (Table 1).

Table 1 – Meta notations

Notation	Destination
W	Set of words of language
L	Set of sentences of language
A	Set of adjectives
N	Set of noun words
V	Set of verbs

V^{-1}	Set of verbs with negative form
D	Set of superlative and comparative adverbs
sent	Predicate of sentiment
¬	Operation of negation words emes/zhok (not/no)
.	Operation of concatenation
$\zeta, \ldots, \xi$	Any string of words, including empty

The production rules for determining the sentiment of an adjective and noun phrases are given below in the sequential form $\frac{A}{B}$, where A – antecedent, B - consequent.

1) If the extracted word is an adjective with positive orientation, and the next word is a noun, then the sentiment of this phrase is positive.

$$\frac{\omega \in L,\ \omega = \zeta \cdot \alpha \cdot \beta \cdot \xi\ ,\ \alpha \in A,\ sent(a) = 1,\ \beta \in N}{sent(\omega) = 1}$$

here and below $\zeta,\ \xi$ - any string of words, including empty.

Example, қызық (positive adjective) кино (noun) (an interesting movie) ⟶ POS.

2) If there is a negation word “емес” (not) between the positive adjective and noun, then the sentiment of this phrase is become negative.

$$\frac{\omega \in L,\ \omega = \zeta \cdot \alpha \cdot \beta \cdot \gamma \cdot \xi\ ,\ \alpha \in A,\ sent(a) = 1,\ \beta = \neg, \gamma \in N}{sent(\omega) = 0}$$

Example, қызық (positive adjective) емес (negation) кітап (noun) (a book is not interesting) ⟶ NEG.

The production rules for defining the sentiment of an adjective and verb phrases are following:

3) In the word combinations (phrases) consisting of an adjective and verb if the adjective is negative and and the next word is a verb, then the sentiment of this phrase is negative.

$$\frac{\omega \in L,\ \omega = \zeta \cdot \alpha \cdot \beta \cdot \xi\ ,\ \alpha \in A,\ sent(a) = 0,\ \beta \in V}{sent(\omega) = 0}$$

Example, жаман (negative adjective) дайындалған (verb) (bad prepared) ⟶NEG.

4) Also, there are cases when the verb has a negative form. If there is a negative form of the verb after negative adjective, then the sentiment of the phrase is positive.

$$\frac{\omega \in L,\ \omega = \zeta \cdot \alpha \cdot \beta \cdot \xi\ ,\ \alpha \in A,\ sent(a) = 0,\ \beta \in V^{-1}}{sent(\omega) = 1}$$

Example, жаман (negative adjective) бол-мады(negative form of the verb) (was not bad)⟶ POS.

The production rules which defines the sentiment of an adverb and adjective phrases are given below.

5) if the superlative or comparative adverbs comes before adjective with positive sentiment, then the sentiment of this phrase is positive.

$$\frac{\omega \in L,\ \omega = \zeta \cdot \alpha \cdot \beta \cdot \xi\ ,\ \alpha \in D,\ \beta \in A,\ sent(\beta) = 1}{sent(\omega) = 1}$$

For example, ең (adverb) әдемі (positive adjective) (the most beautiful)⟶POS.

In addition, the sentiment might also depend on the conjunctions between words or sentences

- if there are connecting conjunctions, the sentiment does not change;

- if between words or sentences comes dividing or adversative conjunctions, then semantic orientation of sentence changes to opposite. For example, бұл кәмпит тәтті, бірақ қатты екен (this candy is tasty, but hard).

Conclusion. In this paper for the first time discusses the study carried out on the sentiment analysis of the Kazakh phrases which was based on the vocabulary and linguistic (morphology) rules of the Kazakh language. It was also planned to apply this method for defining sentiment of sentences ant text in the Kazakh language in future works. For this purpose, we will consider the conjunctions (және, немесе, бірақ(and, or, but, etc.)) and apply different logics to formalize them.

References

1. Bing Liu. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012 (in Eng.).
2. Pang B., Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval. Now Publishers, 2008 (in Eng.).
3. Loukachevitch N.V., Chetviorkin I.I. Evaluating Sentiment Analysis Systems in Russian. Artificial intelligence and decision-making, 2014, vol.1, pp.25-33 (In Russ.)
4. Chetviorkin I., Braslavskiy P., Loukachevich N. Sentiment Analysis Track at ROMIP 2011. In Proceedings of International Conference Dialog-2012, 2012, vol. 2, pp. 1-14 (in Eng.).
5. Chetvirokin I., Loukachevitch N. Sentiment Analysis Track at ROMIP 2012. In Proceedings of International Conference Dialog-2013, vol. 2, 2013. pp. 40-50 (in Eng.).
6. Chetviorkin I., Loukachevitch N. Extraction of Russian Sentiment Lexicon for Product Meta-Domain In Proceedings of COLING 2012, pp. 593-610 (in Eng.).
7. Steinberger J., Lenkova P., Ebrahim M., Ehrmann M., Hurriyetogly A., Kabadjov M., Steinberger R., Tanev H., Zavarella V., Vazquez S. Creating Sentiment Dictionaries via Triangulation. In Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-HLT, 2011, pp. 28-36 (in Eng.).
8. Akba F., Uçan A., Sezer EA., Sever H. Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. In: 8th European Conference on Data Mining, 2014, pp. 180-184 (in Eng.).
9. Ezgi Yıldırım, Fatih Samet Çetin, Gülşen Eryiğit, Tanel Temel. The Impact of NLP on Turkish Sentiment Analysis. In Proceedings of the TURKLANG'14 International Conference on Turkic Language Processing, Istanbul, 2014, (in Eng.). 10. Gülşen Eryiğit, Fatih Samet Çetin, MeltemYanık, TanelTemel, İlyasÇiçekli. TURKSENT: A Sentiment Annotation Tool for Social Media. In Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse, ACL 2013, Sofia, Bulgaria (in Eng.).

УДК 81.322.4:811.512.122:811.581

EXAMINING THE PROGRAM OF MACHINE TRANSLATION FROM KAZAKH TO CHINESE LANGUAGE BASED ON THE RULES OF KAZAKH LANGUAGE

Kaden Kenzhekhan, *Doctorate of Information technologies department University of Xinjiang, Urumchi , China Kenjehan89@mail.ru*

Gulila Altenbek, *College of Information Science and Engineering, Xinjiang University, Urumqi, Xinjiang, 830046, China Xinjiang Laboratory of Multi-language Information Technology The Base of Kazakh and Kirghiz Language of National Language Resource Monitoring and Research Centre Minority Languages, gla@xju.edu.cn*

Unzila Kamanur *L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan Unzila.88@mail.ru*

Machine translation is a type of translation from one language to another performed by a computer. In order to solve issues arisen in the area of machine translation it is significant to have a clear understanding of the context and rules of particular language. The presence of vocabulary with sufficient amount of words guarantees the quality of the translation results. Furthemore, the greatest difficulty incountered in this systems lies on creating this vocabulary. Nowadays, especially at the end of twenty first century the scope and area of translation have widened considerably. It can be noticed that many years before the concentration was on the machine translation from Chinese to Kazakh language. However, today due to the wide demand on worldwide exchange of information to be in a high quality translation from European and East languages to Kazakh language has become popular trend in all spheres of life. As a result, in order to create wide amount of vocabulary the documents from legal and social political areas have been translated. The science of linguistic translation of world's linguistic mainly copes with issues such as terms and term phrases in this sphere. According to the statistics today, there are more than 2000 labels in Chinese language related to the translation. However, in Kazakh language they are directly translated from Chinese language. The purpose of this paper is to perform good quality translation from Chinese to Kazakh language based on the Kazakh language and machine translation rules taken from the texts .

Keywords: Machine translation, Natural language

ИССЛЕДОВАНИЕ ПРОГРАММЫ МАШИННОГО ПЕРЕВОДА С КАЗАХСКОГО ЯЗЫКА НА КИТАЙСКИЙ ЯЗЫК, ОСНОВАННЫЙ НА ПРАВИЛАХ КАЗАХСКОГО ЯЗЫКА

Каден Кенжехан, *докторант отдела «Информационных технологий» Университет Синцзянь, Урумчи, Китай, Kenjehan89@mail.ru*
Гулила Алтенбек*, колледж информационной науки и разработки, Университет Синьцзян, Урумчи, Синьцзян, 830046, Китай, Синьцзянская лаборатория многоязычной информационной технологии, основа казахского и Киргизского языка, центр контроля ресурса национального языка и исследовательские языки меньшинства. gla@xju.edu.cn*
Унзила Каманур *Евразийский национальный университет им.Л.Н.Гумилева Астана, 010000 Казахстан Unzila.88@mail.ru*

Машинный перевод - вид перевода с одного языка на другой выполняется компьютером. Для того, чтобы решить проблемы, появляющиеся в областях машинного перевода, существенно иметь ясное понимание контекста и правила специфического языка. Присутствие словаря с достаточным количеством слов гарантирует качество результатов перевода. Кроме того, самые большие трудности встречаются в этой системе на создании этого словаря. В настоящее время, особенно в конце двадцать первого столетия возможность и область перевода расширилась значительно. Это может быть замечено, что много лет перед тем, как концентрация была на машинном переводе от китайского на казахский язык. Каккогда-либо, сегодня из-за широкого востребования на всемирном обмене информации, чтобы быть на высоком качественном переводе с Европейских и Восточных языков на Казахский,стало популярной тенденцией во всех областях жизни. В результате, для того, чтобы создать широкое количество словаря, документы от законных и социальных политических областей были переведены. Наука лингвистического перевода миров, лингвистических главным образом, покрывает с проблемами как например термины и фразами срока в этой сфере. Согласно статистике сегодня есть более чем 2000 этикеток на Китайском языке имеющие отношение к переводу. Но на Казахском языке они непосредственно переводятся от Китайского языка. Цель этой статьи - выполнять хороший

качественный перевод от китайского к Казахский язык, основанному на Казахском языке и правилах машинного перевода, взятых из текстов .

Ключевые слова: Машинный перевод, Естественный язык

Introduction

Our purpose is to open new areas of machine programming in the context of translation from Chinese language to Kazakh language, especially we focus on defining the models of rules which will be particularly used in programming languages. More generally, does not matter what languages are used for translation activities it is required to fully change all the texts used in communication context so as the main context is understandable. As a consequence, all the translated texts are used for analysis of defining the unique rules used to translate from one to another particular language. It is noticeable that information and culture exchange between nations are taking place widely. Thus, the society and circumstances demand the systematic language among Turkish language nations which has been automatic in satisfactory level.

1. The importance of the research topic

Kazakh language uses two types of the writing style Arabic and Cirillic. In the Cirillic, the writing starts from the left of the page and continues to the right, whereas Arabic in opposite is known as RTL(right-to-left) language. Thus, Arabic being non similar to Chinese makes difficult of using in computer based translation since all writing stiles are one stylized for letter linkage. The link between words in Kazakh language is relatively complicated. We focus on making the translation from Cyrillic to Chinese in a high level based on the research and investigation, so that when information in Kazakh language is analysed it will be clear understandable.

2. Methods used in machine translation

According to the history of Machine translation from the early time divided as Traditional Machine Translation Method (TMTM) and Emerging Machine Translation Method, қысқаша (EMTM). TMTM is considered as being a complicated type of machine translation with rules and it has been covered in sufficient amounts of publication. This type of translation shows good results based on its rules and vocabulary. EMTM was developed based on TMTM and can be considered as Corpus Based Machine Translation (CBMT). CBMT is divided into two groups: Statistics Based and Example Based machine translation. Our systems mostly use rule-based machine translation system which works by conducting an analysis on sentences by using the rules.

3. The main working principals of machine translation

Our vocabulary consist of more than 20 000 verbs, nouns and phrases.

Data

ID	HY	py	KZ	CX
1252	论文	lùn wén	мақала	n
1253	患者	huàn zhě	науқас	n
1254	热情	rè qíng	ықылас	n
1255	策略	cè lüè	тактика	n
1256	老人	lǎo rén	қарт	n
1257	通讯	tōng xùn	байланыс	n
1258	高中	gāo zhōng	орта мектеп	n
1259	同志	tóng zhì	жолдас	n
1260	昨日	zuó rì	күні кеше	n
1261	外面	wài miàn	сырт	n
1262	部队	bù duì	аскери бөлім	n
1263	大厦	dà shà	ғимарат	n
1264	红色	hóng sè	қызыл	n
1265	意外	yì wài	тұтқиыл	n
1266	大哥	dà gē	үлкен аға	n
1267	天堂	tiān táng	жұмақ	n
1268	兴趣	xìng qù	ынта	n

名称name	表格 form
名词 noun	n
动词 verb	v
形容词 ai	a
时间 time	t
代词 pronoun	p
成语 idiom	I
量词 quantifier	L
数词 numeral	num

Figure 1. Vocabulary

Kazakh language rules of cases has following prefixes added at the end of the words

Figure 2. Rules of Kazakh language

The performance of translation program

```
#region//如果uyan以uawts以结尾
if (U_Y_D(gu.kz) && (guo.hy.Equals("給") || guo.hy.Equals("往")))
{
    if (jwa > jin)
    {
        gu.kz = gu.kz + "ға";
    }
    else
    {
        gu.kz = gu.kz + "Г е";
    }
}
#endregion
#region//如果hatan以结尾
else if (H_T(gu.kz) && (guo.hy.Equals("給") || guo.hy.Equals("往")))
{
    if (jwa > jin)
    {
        gu.kz = gu.kz + "қа";
    }
    else
    {
        gu.kz = gu.kz + "К е";
    }
```

Figure 3. Program for performing the translation

Results taken from the translation

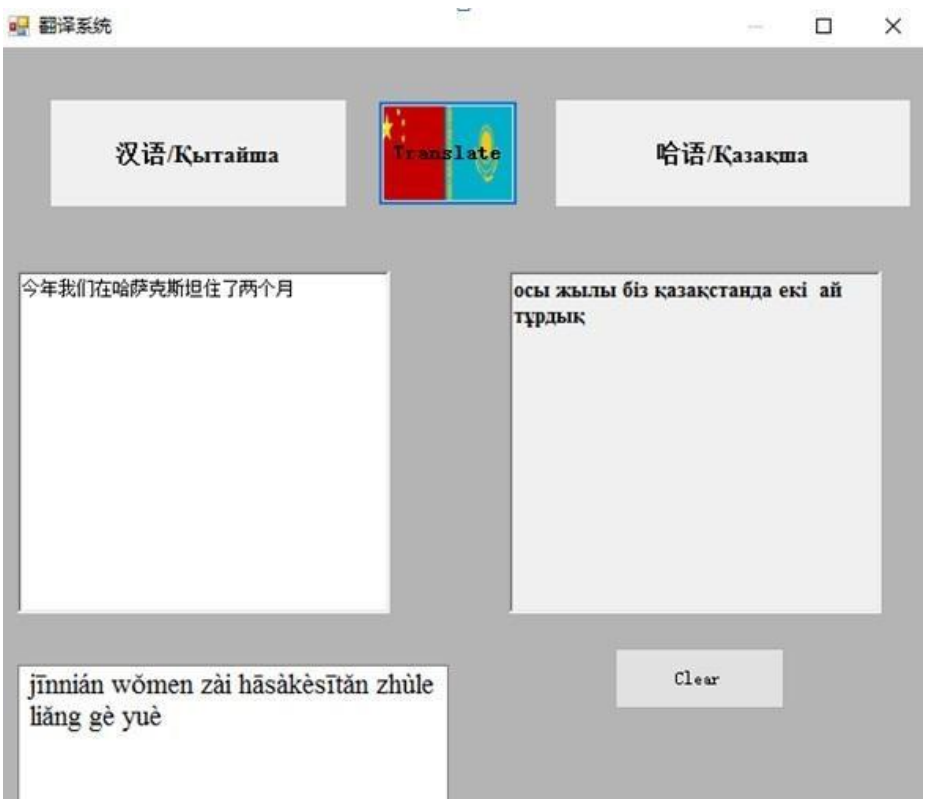


Figure 4. Translation results

Conclusion

The experiments has shown that this system's performance of translation from Chinese to Kazakh language in satisfactory level. In future, it is planned to improve the translation of verb tenses as well as use of the plural nouns, since compared with English language their rules are much more and complicated depending on the vowels and consonants.

Defining the tenses is also considered as a complex process , because it is necessary to perform analysis. Afterwards using taken analysis the program synthesizer has to be created.

Creating the synthesizer include several difficulties since the results needs to be matched with the tense used in original text. Otherwise the sentence meaning can be modified and making the translator confused. All of this abovementioned cases is fully linked to the distinct features of Kazakh language compared with other languages.

.

References

1. Gulila Altenbek. and Dawel,A. and Muheyat,N. 2009. A Study of Word Tagging Corpus for the Modern Kazakh Language, Journal of Xinjiang University, 26(4):323–326 2. http://check.cnki.net/user/,

UDC 004.8

SPOKEN TERM DETECTION FOR KAZAKH LANGUAGE

Zhanibek Kozhirbayev National Laboratory Astana,*Nazarbayev University, Kazakhstan*
e-mail: zhanibek.kozhirbayev@nu.edu.kz
Muslima Karabalayeva *National Laboratory Astana,Nazarbayev University,Kazakhstan,*
e-mail: zhyessenbayev@nu.edu.kz
Zhandos Yessenbayev *National Laboratory Astana,Nazarbayev University, Kazakhstan e-mail: muslima.karabalayeva@nu.edu.kz*

The paper presents a spoken term detection system for Kazakh language in which significant improvements are obtained through modifying speech-to-text process used for generating wordbased lattices. These lattices are indexed and used for the keyword search later. Spoken Term Detection systems quickly discover the occurrence of a term, which might be just a word or sequence of words, in a large audio set of heterogeneous speech records. The paper provides an overview of a speech-to-text and keyword search system architecture built primarily on the top of the Kaldi toolkit and expands on a few highlights. Our aim was to develop a general system pipeline which could be advanced regarding phonological and linguistic features of Kazakh language in order to detect OOV keywords.

Keywords: Speech Retrieval, Lattice Indexing, Spoken Term Detection, Speech Recognition, Keyword Search

ПОИСК РАЗГОВОРНОГО ТЕРМИНА НА КАЗАХСКОМ ЯЗЫКЕ

Кожирбаев Жанибек Мамбеткаримович *Национальная лаборатория Астаны, Назарбаев Университет, Казахстан e-mail: zhanibek.kozhirbayev@nu.edu.kz*
Карабалаева Муслима Хизятовна *Национальная лаборатория Астаны, Назарбаев Университет, Казахстан e-mail: muslima.karabalayeva@nu.edu.kz*

Есенбаев Жандос Аманбаевич *Национальная лаборатория Астаны, Назарбаев Университет, Казахстан e-mail: zhyessenbayev@nu.edu.kz*

В статье представлена система для обнаружения разговорного термина на казахском языке, в котором получены значительные улучшения путем модификации процесса речи в текстовый формат, используемый для создания слов на структурной (lattice) основе. Эта структура индексируется и позже используется для поиска ключевого слова. Система обнаружения разговорного термина быстро обнаруживает возникновение термина, которым может быть просто слово или последовательность слов в большом аудио наборе разнородных речевых записей. В статье дается обзор речи в текстовый формат и архитектура системы поиска ключевых слов, построенной на основе платформы Kaldi и расширяемой на несколько основных моментов. Наша цель состояла в том, чтобы разработать общую систему, которая может продвинуться вперед в отношении фонологических и лингвистических особенностей казахского языка с целью выявления OOV ключевых слов.

Ключевые слова: Поиск речи, Индексация структур, Поиск разговорного термина, Распознавание Речи, Поиск по ключевым словам

1. Introduction

Data processing is nowadays an essential activity of IT industry and recorded materials is considered as core source of that data. Hence, together with increasing volume of audio information, a possibility grew for developing an efficient information retrieval system for audio data storage. Spoken term detection (STD) attempts to set up a specific term (considered as a sequence of single or multiple words) speedily and thoroughly in major heterogeneous audio storages, in order to be utilized solely as input to better compounded information retrieval techniques.

The STD task is based on the detection process, demanding every appearance to be clarified by its beginning and ending times in contradiction to spoken document retrieval. Moreover, systems have to maintain a score for every appearance and a tough decision showing its accuracy. Unprocessed audio materials and a list of search key words together make an input for the task. Despite the fact that the evaluation practically utilizes only small number of data, it is designed to initiate the biggest data condition. Thus, the systems are obliged to be deployed in two stages: indexing as well as searching. Processing of audio data is held meanwhile indexing stage going on, without knowing about search terms. In order to retrieve the terms the output index is kept and admitted in the course of the searching stage. The searching stage might be replicated several times for various terms so the effectiveness of its deployment is very significant

The remainder of this paper is organized as follows. Section 2 describes related works. The system architecture will be discussed in details in Section 3. Section 4 demonstrates the experiments and the obtained results. Finally, the last section concludes the paper and suggests further research in this area.

2. Related works

Languages with morphologically features like Kazakh experience challenges in large vocabulary continuous speech recognition (LVCSR) process by virtue of vocabulary increase. Enormous vocabularies, great frequency of OOV words as well as scattered data for LM are the prominent indications of word-based lexical techniques. A lot of researches have been conducted to solve these challenges regarding OOV issue.

First method to limit the OOV issue is to preventively enlarge the LVCSR dictionary. To be precise, a person augments automatically created pronunciations of a huge amount of words in the LVCSR dictionary prior to lattice creation [1].

Second method to deal with OOV words is though sub-word blocks similar to phones, syllables or morphemes. A subword index is built by producing a sub-word lattice which were presented by [2], or by modifying a word lattice to a sub-word lattice with or without the utilization of a proper phone confusion matrix described in [3].

The OOV challenge also can be solved by introducing the concept of query expansion in text retrieval. Optional words or syllables for OOV can be produced by a confusion matrix utilized in [4]. Consequently, rather than searching primary keywords, the system searches within word or syllable index. In contradistinction to concept of query expansion, in which a person adds potentially lacking word with other terms which are corresponding semantic features, speech search involves different words which sound identically.

Job performed in this paper is based on [5], in which proxy keywords are generated utilizing a confusion matrix. Rather than looking for the proxy keywords in list produced by the LVCSR system, weighted finite state transducer (WFST) is applied on the basis of framework for directly linking several proxies versus the whole LVCSR lattices.

3. Kaldi-based STD system

The Kaldi-based STD system includes two different subsystems, as illustrated in Figure 1: the ASR subsystem and the STD subsystem. This section presents each subsystem in more detail regarding interprocesses and tuning parameters.

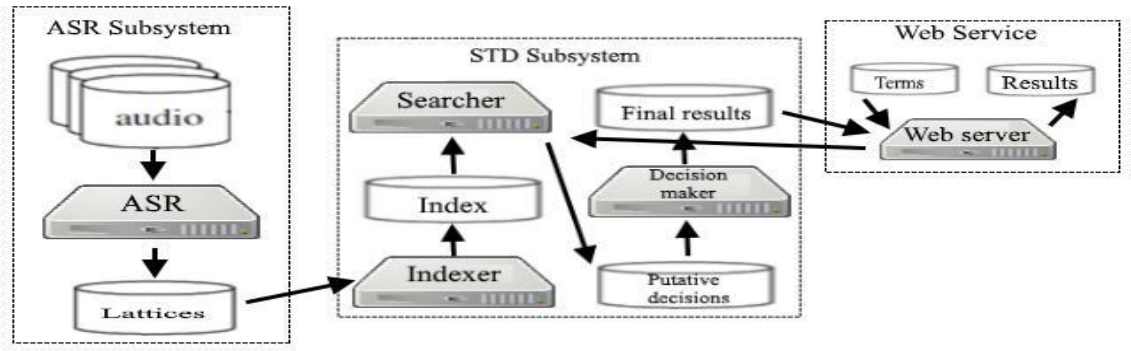


Figure 1: Spoken Term Detection System Architecture

3.1 ASR subsystem

The ASR subsystem utilizes the Kaldi toolkit [6] to gather word lattices from the raw audio data. It applies 13-dimensional Mel-frequency cepstral coefficients, Linear Discriminant Analysis transform as well as Maximum Likelihood Linear Transform rating. A flat-start initialization of context-independent phonetic HMMs begins the training, while speaker adaptive training of stateclustered triphone HMMs along with GMM output densities finishes it. Moreover, the ML-based acoustic model training step is run, which followed by a universal background model. It is created from speaker-transformed training data that is utilized to train an SGMM applied in the decoding step to produce the word lattices [7, 8].

Kazakh speech data from the KazSpeechDB and KazMedia corpora is employed to train the above mentioned acoustic model. The KazSpeechDB corpus as part of Kazakh Language Corpus [9] is a body of utterances consisting of 12675 Kazakh sentences recorded in a sound recording studio, uttered by speakers of different age and gender, from different regions of Kazakhstan. Every audio file is supplied with a text file that contains transcription text of the utterance. The corpus contains 22 hours of speech.

The KazMedia corpus is a body of text and audio data collected from official websites of broadcast news channels “Khabar”, “Astana TV” and “Channel 31”. The audio data is 600 wavfiles, which are actually audio tracks extracted from a number of video news in Kazakh. Every wavfile is

supplied with a txt-file that contains detailed transcription text of the news and a time-aligned annotation file with labels about speaker gender, language and noise. The total duration of these audio files is 13 hours of speech. The total training data amount is about 30 hours of speech.

The dictionary and the language model (LM) of the ASR subsystem were formed on the basis of cumulative text data of both KazSpeechDB and KazMedia sub-corpora. The dictionary contains over 160000 words. The LM was trained using a text database of about 3.5 million words.

A common metric of the performance of experimental speech recognition models is WER (word error rate), which is computed as the ratio of erroneously recognized words to the total number of words. It is commonly believed that a lower WER shows superior accuracy in recognition of speech, compared with a higher WER. The experimental results are given in Table 1.

Table 1: Minimum value of the WER on the train, validation and test sets

Experiment \ Set	train	dev	test1 (Khabar)	test2 (Astana TV)	test3 (Channel 31)
Triphones	6.19 %	6.42 %	6.62 %	14.87 %	19.52 %
SGMM	5.16 %	5.39 %	5.56 %	13.18 %	16.95 %

3.2 STD subsystem

The STD subsystem combines Kaldi term detector [1] that searches for the keyword terms among the lattices generated by the ASR subsystem. Firstly, word lattices of all the utterances in the speech data from individual WFSTs are converted to a single generalized factor transducer pattern that aggregates the start and end times, as well as the lattice posterior probability of each word token as a three-tuple cost in the lattice indexing approach presented by [10]. This factor transducer indicates an inverted index of all the word sequences comprised in the lattices. Hence, with the search term, an ordinary finite state machine which receives the term is built and structured with the factor transducer with an eye to gather all the appearances of the term in the audio data. The posterior probabilities of the lattice relatively to all the words of the search term are piled up, appointing a confidence score to every detection. The decision maker process merely gets rid of those detections with a confidence score which is less than a predetermined threshold.

The Kaldi STD system [1] treats OOV term search by virtue of a technique named proxy words [6]. This approach based on replacement every OOV word of the keyword term to acoustically corresponding IV proxy words, demolishing the necessity of a subword-based system for handling OOV term search.

4. System evaluation and results

An assessment of the STD systems performance on the basis of the term length (IV words) has been conducted (Table 2, 3). Test search queries are divided into three classes: unigrams, bigrams and trigrams. Generally, longer queries should produce greater results than shorter ones as long as these are originally more liable to be confused with speech data.

Table 2: Search term categories (IV test set) Table 3: Search terms statistics (IV test set)

	Terms		
Type	Unigrams	Bigrams	Trigrams
Amount	1000	1000	1000
Total	3000		

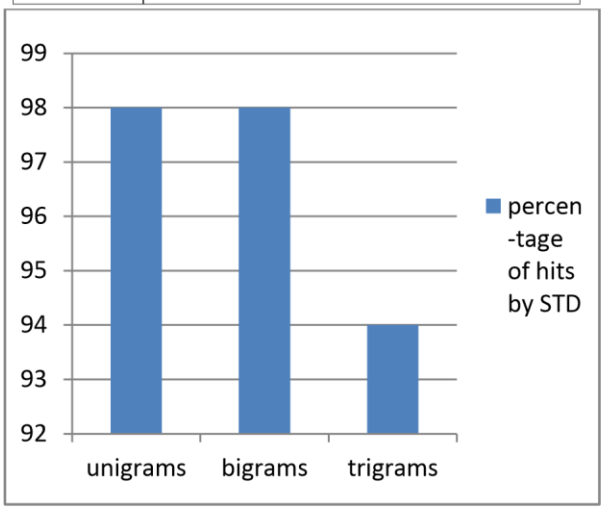


Figure 2: The percentage of hits by STD

Terms	Occurence of IV quesires in test set
unigrams	82874
bigrams	12383
trigrams	3937

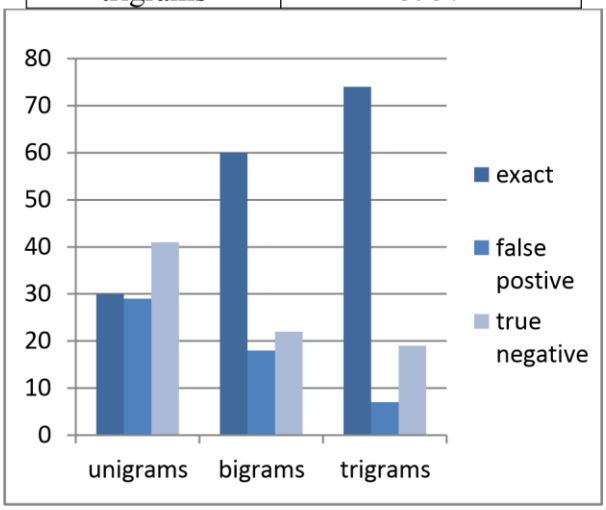


Figure 3: Hits out of 1000 queries for each category

The above figures show that STD system performs in searching IV keywords well. It can be seen that trigrams and bigrams are detected more accurately rather than unigrams as expected (Figure 2). Figure 3 presents the results of exact, false positive and true negative hits out of 1000 queries for each category. As the decision maker process depends on the posterior probability of the occurred term, there might be a false alarm or accurate decision for query terms. Therefore, the exact hits mean that the frequency of query matches with frequency obtained by the STD system, the false positive and true negative hits show that original frequencies are more and less respectively than the gathered ones.

The test set with 295 OOV words was created in order to evaluate the STD system for OOV queries. More precisely, the words are not in the main lexicon but are in the manual transcripts of the test audio set. In order to do fuzzy search, the phone confusions are gathered firstly and a G2P model is trained using G2P Sequitur [11]. Afterwards, this model was applied to the OOV keywords and KWS data directory was built. The system searches the words from the main lexicon with the minimum edit distance score. For example: “ӘЖЕМДІ” → “ӘЖІМДІ”, “АДЫМДАРЫН” → “АДАМДАРЫН”, “ДӘПТЕРІМ” → “ДӘПТЕРІН”. The results of the experiment are shown in Table 4.

Table 4: OOV test set results

OOV queries	Hits by STD	# Yes decisions	# No decisions
295	112	259	114

5. Conclusion and future works

In this paper we present an STD system that benefit significantly from tuning the speech-totext process and applying lattice indexing. Making applied models more diverse, indexing and searching methods produce more advanced results for our system. Even though the proxy keywords approach applied for detecting the OOV terms and it does not show the expected results, there are other approaches for handling the keywords which are not in lexicon. In future, phone, syllable, morph bases techniques and even feature based method will be utilized to enhance this research for Kazakh language.

Acknowledgments

This work was supported by the Ministry of Education of Science of the Republic of Kazakhstan under the research program number 349//049-2016.

References

1. Chen, G, Khudanpur, K, Povey, D, Trmal, J, Yarowsky, D and Yilmaz, O 2013, "Quantifying the value of pronunciation lexicons for keyword search in low resource languages", *in Proceedings of ICASSP 2013*, pp. 8560–8564.

2. Saraclar, M and Sproat, R 2004, "Lattice-based search for spoken utterance retrieval", *in Proceedings of HLTNAACL 2004*.

3. Chaudhari, U & Picheny, M 2007, "Improvements in phone based audio search via constrained match with high order confusion estimates", *in Proceedings of ASRU 2007*, pp. 665–670.

4. Karanasou, P, Burget, L, Vergyri, D, Akbacak, M and Mandal, A 2012, "Discriminatively trained phoneme confusion model for keyword spotting", *in Proceedings of Interspeech 2012*.

5. Chen, G, Yilmaz, O, Trmal, J, Povey, D, Khudanpur, K 2013, "Using proxies for OOV keywords in the keyword search task", *in Proceedings of ASRU*, pp. 416–421.

6. Povey, D, Ghoshal, A, Boulianne, G, Burget, L, Glembek, O, Goel, N, Hannemann, M,
Motlicek, P, Qian, Y, Schwarz, P, Silovsky, J, Stemmer, G, Vesely, K 2011, "The KALDI speech recognition toolkit", in Proceedings of ASRU.

7. Yessenbayev, Zh and Karabalayeva, M 2016, "A baseline system for Kazakh broadcast
news transcription", *in Proceedings of the V International scientific-practical conference information society*, pp. 48-50.

8. Kozhirbayev, Zh and Islam, Sh 2016, "A distributed platform for speech recognition research", *in Proceedings of the V International scientific-practical conference information society*, pp. 38-40.

9. Makhambetov, O, Makazhanov, A, Yessenbayev, Zh, Matkarimov, B, Sabyrgaliyev, I and Sharafudinov, A 2013, "Assembling the Kazakh Language Corpus", *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1022–1031.

10. Can, D and Saraclar, M 2011, "Lattice indexing for spoken term detection", *IEEE Trans. Audio Speech Lang. Process*, pp. 2338–2347.

УДК 681.5

ТАБИГЫЙ ТИЛДЕГИ ТЕКСТТЕРДИ ОРУС ТИЛИНЕН КЫРГЫЗ ТИЛИНЕ МАШИНАЛЫК КОТОРУУДА СӨЗДӨРДҮ АНАЛИЗДӨӨНҮН АЛГОРИТМИН ТҮЗҮҮ

Кочконбаева Буажар Осмоналиевна, *старший преподаватель, ОшТУ им. академика М.М. Адышева, 714018, г. Ош, ул. Исанова 81, e-mail: buajar@mail.ru*

Морфологиялык талдоо жана сөз түзүү процесстери бири бирине жакын болуп эсептелет. Машиналык которуу процессинде бир эле учурда сөздөрдү морфологиялык жактан талдап аны жаңы тилге которуу үчүн сөз жасоого туура келет. Бул процесстер бүгүнкү күндө актуалдуу проблемалардан болуп эсептелет. Анткени бардык эле улут өз тилинин унутулбай автоматташтырылышын жана интернет сайттарына кошулуусу үчүн аракет кылышууда.

Ошондуктан мен бул макалада табигый тилдеги тексттерди орус тилинен кыргыз тилине которуучу программаны түзүүнүн алгоритмин карап чыктым.

Ачкыч сөздөр: машиналык которуу, морфологиялык анализатор, аффикстер, синтаксистик анализатор,семантикалык анализатор, программа, маалыматтар базасы

DEVELOPMENT OF ALGORITHM ANALYSIS OF WORDS IN NATURAL TEXT MACHINE TRANSLATION FROM RUSSIAN INTO KYRGYZ

Kochkonbaeva Buazhar Osmonalievna, *senior lecturer, OshTU after named acad. M.M. Adishev, 714018,c. Osh, Isanov st. 81, e-mail: buajar@mail.ru*

Morphological analysis of words and word-formation are very similar in how to conduct and are closely related to each other. Today these processes are very actual. Because all the nations seek to automate their language and add to the online translators. In this article I will consider the algorithm of a program algorithm machine translation of texts from Russian to Kyrgyz language.

Keywords: machine translation, Morphological analyzer, affixes, syntactical analyzer, semantic analyzer, program, database.

Киришүү

Жасалма интеллект багытындагы адам баласынын изилдоолору кун санап өсүп барат. Азыркы күндө компьютер бир гана эсептөө техникасы эмес, ар кандай багыттагы суроолорго жооп берүүчү эксперттик система, бир тилден экинчи тилге тексттерди которуучу каражат же болбосо интеллектуалдык оюндарды ойноодо экинчи тарап катары жасалма интеллекттин ролун жаратып келүүдө. Бул макалада жогоруда айтылган багыттардын ичинен табигый тилдеги тексттерди орус тилинен кыргыз тилине машиналык которууда сөздөрдү анализдөөнүн компьютердик моделин (морфологиялык анализатор) түзүү маселеси каралды.

Тексттин үстүнөн иштөөнүн этаптары

Машиналык которуу – бул компьютердик программанын жардамында бир табигый тилдеги текстти экинчи табигый тилге которуу болуп эсептелет [1]. Түзүлгөн программанын сапаты которуунун тактыгынан көз каранды. Бүгүнкү күндө коммерциялык багыттагы көптөгөн программалар (PROMPT, Lingvo, Google сайтынын которуу кызматы ж.б.) атайын түзүлгөн сөз айкаштардын корпустук сөздүгүнүн жардамында тексттерди которуп келет. Бул программалардын ичинен Google сайтынын которуу сервистик программасына тексттерди кыргыз тилине которуу мүмкүнчүлүгү кошулду.

Машиналык которуу программасын түзүү өз ичине графометикалык, морфологиялык, синтаксистик жана семантикалык этаптарды камтыйт. Морфологиялык анализдөө модулунда кийирилген орус тилиндеги текст лексемаларга бөлүнөт жана алардын морфологиялык касиеттери (сөз түркүмү, жагы, чагы) аныкталат. Кийинки этап синтаксистик анализаторго берилет да, анда ар бир сөз сүйлөмдө өз ордуна жайгашат. Мисалы кыргыз тилинде сүлөмдөр SOV (Subject – object - verb) тартибине баш ийет. Башкача айтканда ээлик состав сүйлөмдүн башында жайгашат жана баяндоочтук состав сүйлөмдүн аягынан орун алат. Мисалы: На полях работали новые машины. Бул сүйлөмдү кыргыз тилине которууда төмөнкү тартиптеги сүйлөм алынат: Жаңы машиналар талааларда иштешти.

Кийинки этапта семантикалык анализатордун жардамында сүйлөмдөгү сөздөрдүн бири-бири менен байланышы каралат да сүйлөмдүн маанисине көңүл бурулат.

Машиналык которууда ар турдүү маани берүүчү сөздөрдү анализдөө

Бардык түрк тилдери сыяктуу эле кыргыз тили да агглютинативдик тилдерге кирет. Тексттин толуктугу жана сөздөрдүн маниси жактан байланышы уңгуга аффикстердин уланышы менен ишке ашат. Колдонмо лингвистикалык процесстерде сөз түшүнүгүнүн жалпы моделин түзүү өтө татаал, анткени ар бир тилде өзүнө тиешелүү касиеттер жана чектөөлөр кабыл алынган. Сөз жасоонун негизги эрежелерине таянып төмөнкү туюнтманы жазууга болот:

<сөздүн грамматикалык формасы>::=<сөз><грамматикалык көрсөткүчтөрү>

<грамматикалык көрсөткүчтөр>::=<аффикс>|<предлог>|<…>

<сөз>::=<зат атооч>|<сын атооч>|<…>

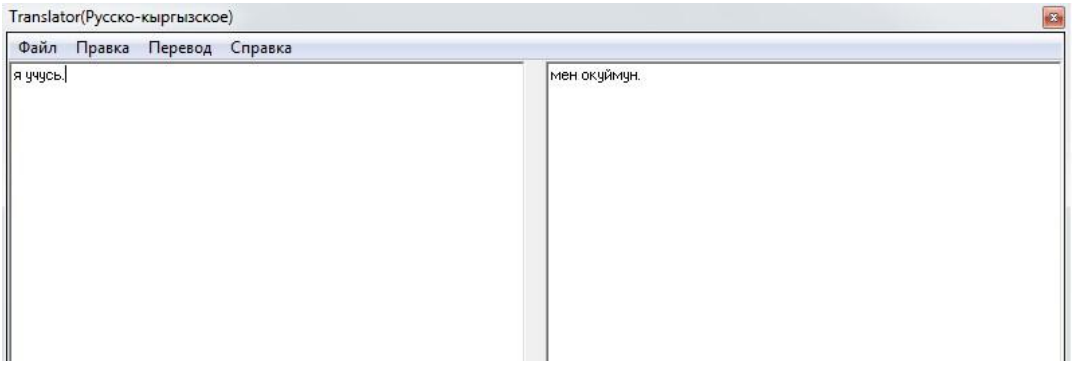
Изилдөөлөрдүн негизинде Delphi чөйрөсүндө морфологиялык анализатордун программасы түзулдү. Сөздөрдүн морфологиялык касиеттерин камтыган таблица Access чөйрөсүндө түзүлүп, төмөнкүдөй мамычалардан турат (1-сүрөт).

ID	word_r	w_id	ch_r	id_chislo	id_padej	id_rod	koren	word_k
315	будут	332	2	2	0	0	буд	
316	голос	333	1	1	1	2	голос	үн
317	лица	334	1	2	1	3	лиц	жүздөр
318	солнце	335	1	1	1	3	солн	күн

1-сүрөт. Морфологиялык таблица

Мында сөздөр морфологиялык касиеттерине жараша бир канча мамычаларга бөлүнүп көрсөтүлгөн.

Морфологиялык анализатордун программасын түзүүдө орус тилиндеги көп колдонулган сөздөрдүн ар кандай морфологиялык абалынан турган 10000 ге жакын сөздөр маалыматтар базасына кийирилип, Delphi программалоо чөйрөсүнүн жардамында колдонуучу үчүн интерфейс түзүлдү.



2-сүрөт. Программанын интерфейси

Сөздөрдү жаратуунун ар кандай жолдорун техникалык жактан мүнөздөө өтө оор процесс, анткени ал көптөгөн эрежелерден жана чектөөлөрдөн турат. Бир канча уңгудан турган сөздөрдү бир тилден экинчи тилге которуу кийирилген сөзгө бир дагы окшобогон башка морфологиялык сөздүн жаралуусуна алып келет. Мисалы: ВУЗ- ЖОЖ, малоисследованный- аз изилденген ж.б.

Төмөндөгүдөй белгилөөлөрдү кийирип алабыз: w-сөз, p-префикстер, r-сөздүн уңгусу, a-аффикстердин көптүгү.

<w>::=<r>|<p><r>|<r><a>|<p><r><a>

Мисалы: игра-оюн, игрушка – оюнчук, играть-ойноо, выиграть – утуу, проиграть-утулуу.

Бул мисалдарда сөздөрдүн бир эле уңгуга морфологиялык бирдиктердин уланышынын негизинде маанилик жактан өзгөрүшү көрсөтүлдү.

Бир табигый тилден экинчи тилге машиналык которууда көп маанилүүлүктү чагылдыруу методун пайдалануу которуунун эффективдүүлүгүн жогорулатып, тексттин маанисине жараша көп маанилүүлүктөн бир мааниге өтүүгө жардам берет.

Корутунду

Негизинен машиналык которуу багытында иштөөдө төмөнкүдөй амалдар аткарылды:

- көп колдонулуучу сөздөрдүн маалыматтар базасы түзүлду;
- сөздөрдүн морфологиялык өзгөчөлүктөрү базага кийирилди;
- морфологиялык анализатордун программасы иштелип чыкты;
- жөнөкөй тексттерди сүйлөмдөргө ажыратуу жана которуу амалдары аткарылды.

Мындан сырткары көп маанилүүлүктү чагылдыруу методун программада пайдалануу маселеси каралды.

Колдонулган адабияттар

1. Ж.Ч. Джанузакова «Лингвокультурологические аспекты машинного перевода» [электрондук булак].

2. К.Х. Ким, А.П. Савинов «Синтаксический анализатор для вопросно-ответной системы»

3. А. Леонтьева, И. Кагиров «Автоматический синтаксический анализ русских текстов». Труды 10-й Всероссийской библиотеки: перспективные методы и технологии, электронные коллекции»-RCDL’2008,Дубна, Россия,2008.

УДК 004.415.2

ПРОЕКТИРОВАНИЕ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ПРОВЕРКИ ОЛИМПИАДНЫХ ЗАДАНИЙ ПО ПРОГРАММИРОВАНИЮ

Макиева Замира Джумакматовна, *доцент КГТУ им. И.Раззакова, Кыргызстан, 720044, г.Бишкек, пр. Ч.Айтматова 66,* e-mail: z.makieva@gmail.com

Целью данной работы является проектирование автоматизированной системы проведения олимпиад по программированию, главной задачей которой является автоматическая проверка программ участников. Автором сделан краткий обзор существующих систем, показана необходимость разработки, созданы функциональная модель, модель потоков данных, диаграмма деятельности и алгоритм проверки решений.

Ключевые слова: программирование, автоматическое тестирование, олимпиада, моделирование системы, диаграмма.

DEVELOPMENT OF THE AUTOMATED VERIFICATION SYSTEM FOR PROGRAMMING OLYMPIAD TASKS

Makieva Zamira Jumakmatovna, *the docent of KSTU named after I. Razzakov, Kyrgyzstan, 720044, Bishkek, C.Aitmatov Avenue 66,* e-mail: z.makieva@gmail.com

The purpose of this work is to design of the automated system of holding the Olympiad on programming, the main goal of which is automatic verification of programs of participants. The author has made the short review of the existing systems, need of development was shown, the functional model, model of data flows, activity diagram and program verification algorithm were created.

Keywords: programming, automatic testing, Olympiad, system modeling, diagram.

По программированию, как и по другим дисциплинам, проводятся соревнования среди школьников и студентов. Но в отличие, например, от соревнований по математике, здесь участники должны не просто решить задачу и представить ответ, а написать компьютерную программу, решающую целый класс однотипных задач с разными исходными данными. Соответственно, для проверки решений задач по программированию недостаточно сравнить ответы, а для каждой задачи необходимо приготовить множество тестов. Проверка вручную связана с определёнными проблемами, в первую очередь с затратами сил и времени и низким качеством ручной проверки в целом. Также приходится придумывать задачи, в которых выходных данных как можно меньше, чтобы упростить проверку. При этом отсеивается часть сложных и интересных задач. При ручной проверке оптимальность решения какой-либо задачи оценивается "на глазок", приблизительно и чаще этот критерий не учитывается за исключением явных "зависаний" проверяемого алгоритма. Наконец, после подведения итогов требуется вручную свести их в турнирную таблицу и необходимым образом её оформить в каком-либо текстовом редакторе.

В мире существует много аналогичных систем, мы рассмотрим две из них. Timus Online Judge [3] – это крупнейший в России архив задач по программированию с автоматической проверяющей системой. Основной источник задач для архива — соревнования Уральского федерального университета, Чемпионаты Урала, Уральские четвертьфиналы ACM ICPC, Петрозаводские сборы по программированию. Timus Online Judge позволяет принять участие в онлайн-версиях большинства соревнований, которые регулярно проходят в Уральском федеральном университете.[3]

Система KRSU Online Judge [2] используется для проведения кыргызстанского четвертьфинала студенческого командного чемпионата мира по программированию по правилам ACM ICPC (Association for Computing Machinery International Collegiate Programming Contest) [3, 4, 5] и для тренировок.

Все эти системы доступны для тренировок, но не для проведения соревнований. Также правила начисления баллов за решение задач в студенческих олимпиадах отличаются от олимпиад для школьников [5]. С 2014 года кафедра Программное обеспечение компьютерных систем КГТУ им. И.Раззакова проводит городскую и республиканскую олимпиады школьников по информатике, что привело к необходимости разработки собственной системы. В качестве отличий разрабатываемой системы можно отметить то, что система будет проверять решения и по правилам проведения школьных олимпиад (International Olympiad in Informatics) и по правилам проведения студенческих олимпиад (Association for Computing Machinery).

Для решения указанных проблем требуется разработать автоматизированную тестирующую систему, которая способна выполнять проверку быстро и точно, независимо от количества входных и выходных данных.

Был произведен анализ предварительных требований к системе, и в результате была разработана спецификация требований, которая представлена ниже.

Функциональные требования к системе

1) Добавление, удаление и редактирование задач и тестов к ним в архив, причем условия задач должны быть на кыргызском и русском языках.

2) Поддержка задач с множественными правильными ответами.

3) Регистрация участников самостоятельно и/или администратором.

4) Организация соревнований: составление списков задач и участников, указание времён начала и окончания.

5) Поддержка режимов тестирования по правилам ACM ICPC и IOI.

6) Приём исходного кода для проверки как в рамках соревнований, так и вне их.

7) Тестирование, включающее компиляцию исходного кода и автоматическую проверку работы получившейся программы на наборе тестов из архива, а также проверку времени работы и потребляемой памяти в процессе работы тестируемого алгоритма.

8) Вывод результатов тестирования, включая статус, количество пройденных тестов и сообщения об ошибках, возникших в процессе тестирования.

9) Составление таблицы с результатами соревнования для непосредственного просмотра и в формате PDF для скачивания и распечатки.

10) Поддержка языков программирования C++, Pascal, QBasic, возможность добавления других языков программирования.

11) Функционал резервного копирования системы в облако.

Нефункциональные требования

1) Система должна иметь возможность работать как в локальной, так и в глобальной сети (интернет).

2) Система должна разграничивать доступ к данным.

3) В случае сбоя тестирующей системы, перезапуск должен осуществляться не дольше, чем за 30 секунд.

4) Система должна быть достаточно производительной, чтобы участникам соревнований не приходилось подолгу ждать результатов. Время получения результата решения не должно превышать 2 минут в обычном режиме и 10 минут в случае сбоя системы.

При построении **концептуальной модели** можно выделить три основных типа пользователей данной системы:

1) Пользователь – зарегистрированный пользователь, который не принимает участия в соревнованиях и решает задачи из архива.

2) Участник соревнований – пользователь, который участвует в соревновании и имеет доступ только к задачам этого соревнования.

3) Администратор – пользователь, который может организовывать соревнования, добавлять в базу данных задачи, тесты и других пользователей. Администратор имеет полный доступ ко всей системе.

На рис.1 приведена Use-case диаграмма, соответствующая этой модели.

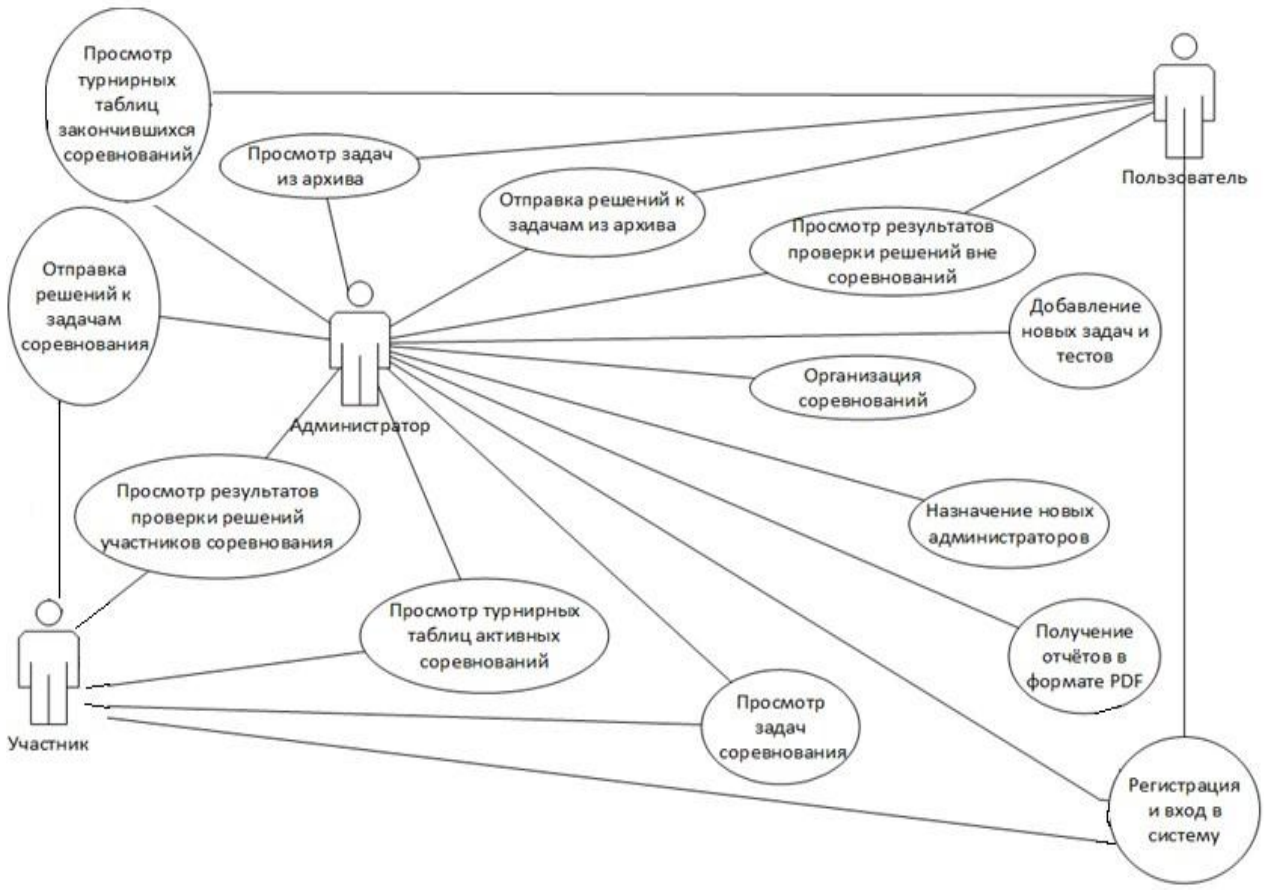


Рис. 1 Диаграмма вариантов использования

На диаграмме представлены следующие варианты использования:

1) Регистрация и вход в систему – регистрация доступная всем, вход в систему с помощью логина и пароля – зарегистрированным пользователям.

2) Просмотр задач из архива – доступен зарегистрированному пользователю, не участвующему в соревнованиях либо администратору. Задачи из соревнований в архиве не видны.

3) Отправка решений к задачам из архива – доступна простому пользователю и администратору. После отправки решение проверяется системой, и система выводит результат.

4) Просмотр результатов проверки вне соревнований – пользователи,

тренирующиеся на задачах из архива, могут просматривать статусы своих и чужих решений, но не имеют доступа к решениям в рамках соревнований.

5) Просмотр турнирных таблиц закончившихся соревнований – доступен всем, кроме тех, кто участвует в активном соревновании.

6) Просмотр задач соревнования – доступен участникам этого соревнования и администратору. После окончания соревнования задачи становятся доступны в архиве.

7) Отправка решений к задачам соревнования – доступна участникам соревнований и администратору.

8) Просмотр турнирных таблиц активных соревнований – участники соревнований могут просматривать только турнирную таблицу своего соревнования. Администратор может просматривать любые турнирные таблицы.

9) Просмотр результатов проверки решений в рамках соревнований – участники соревнований могут просматривать только статусы друг друга.

10) Добавление задач и тестов – эта функция доступна только администратору.

11) Назначение новых администраторов – администратор может назначить себе помощников из числа пользователей.

12) Организация соревнований – включает в себя составление списков задач и участников и назначение времени начала и окончания. Доступ только для администратора.

13) Получение отчётов в PDF – отчёты для распечатки и вывешивания, доступно администратору.

Модель потоков данных

Подробная модель представлена на рис. 2 в виде диаграммы потоков данных (Data Flow Diagram).

Рис. 2 Диаграмма потоков данных (DFD)

Внешними сущностями являются участники – пользователи системы и администрация – авторы задач, организаторы соревнований.

Для получения доступа к системе, участник должен зарегистрироваться и авторизоваться. Авторизованные участники могут читать условия задач и отправлять решения на проверку, указывая при этом язык программирования и задачу, которую они решали. По итогам проверки система выводит результаты проверки отдельных решений и общие итоги соревнований.

Администрация может добавлять новые тесты и задачи, организовывать соревнования и получать их итоги в формате PDF.

На рис. 3 представлена декомпозиция DFD-диаграммы тестирующей системы для демонстрации потоков данных внутри неё.

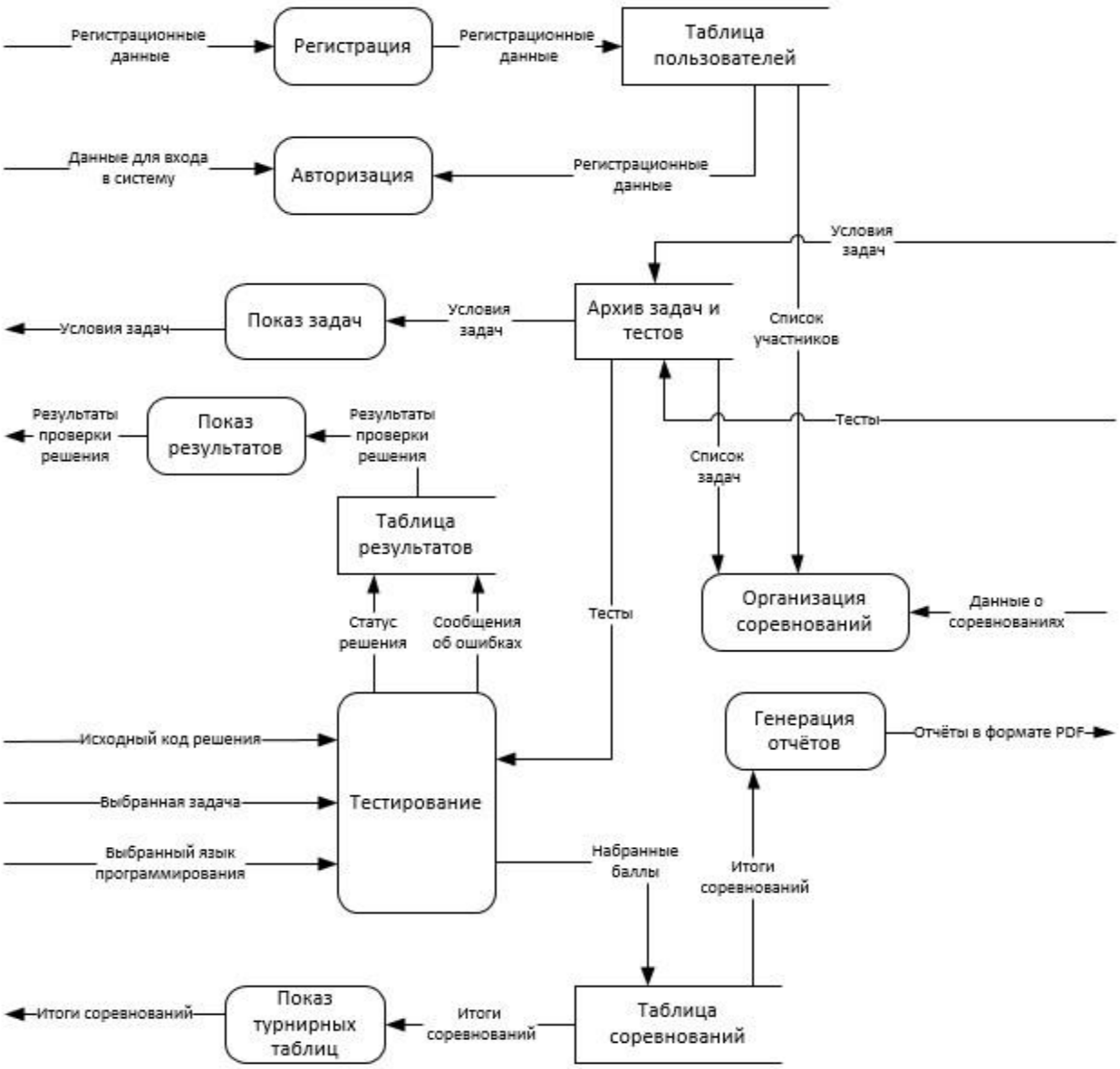


Рис. 3 Декомпозиция диаграммы потоков данных (DFD)

Здесь видно, что результаты проверки решений представляют собой статусы и сообщения об ошибках, возникших в процессе тестирования. Помимо выставления статусов и выдачи сообщений при тестировании ведётся подсчёт баллов, набранных за решение задачи, которые в дальнейшем отображаются в итогах соревнований.

Алгоритм тестирования решений к задачам

При проверке решений используется следующий алгоритм:

1) Скомпилировать исходный код. Если компиляция не удалась, вывести сообщение об ошибке компиляции и прервать тестирование.

2) Подготовить входные данные теста и запустить полученную программу.

3) Если программа работает слишком долго, прервать тестирование и выполнение программы.

4) Если программа потребляет слишком много памяти, прервать тестирование и выполнение программы.

5) Если программа завершила работу с ошибкой, прервать тестирование.

6) Если программа завершила работу успешно, то сравнить результаты работы программы и выходные данные теста. Если результат неправильный, прервать тестирование.

7) Продолжать тестирование с пункта 2, пока не кончатся тесты.

Алгоритм представлен в виде диаграммы деятельности на рис. 4.

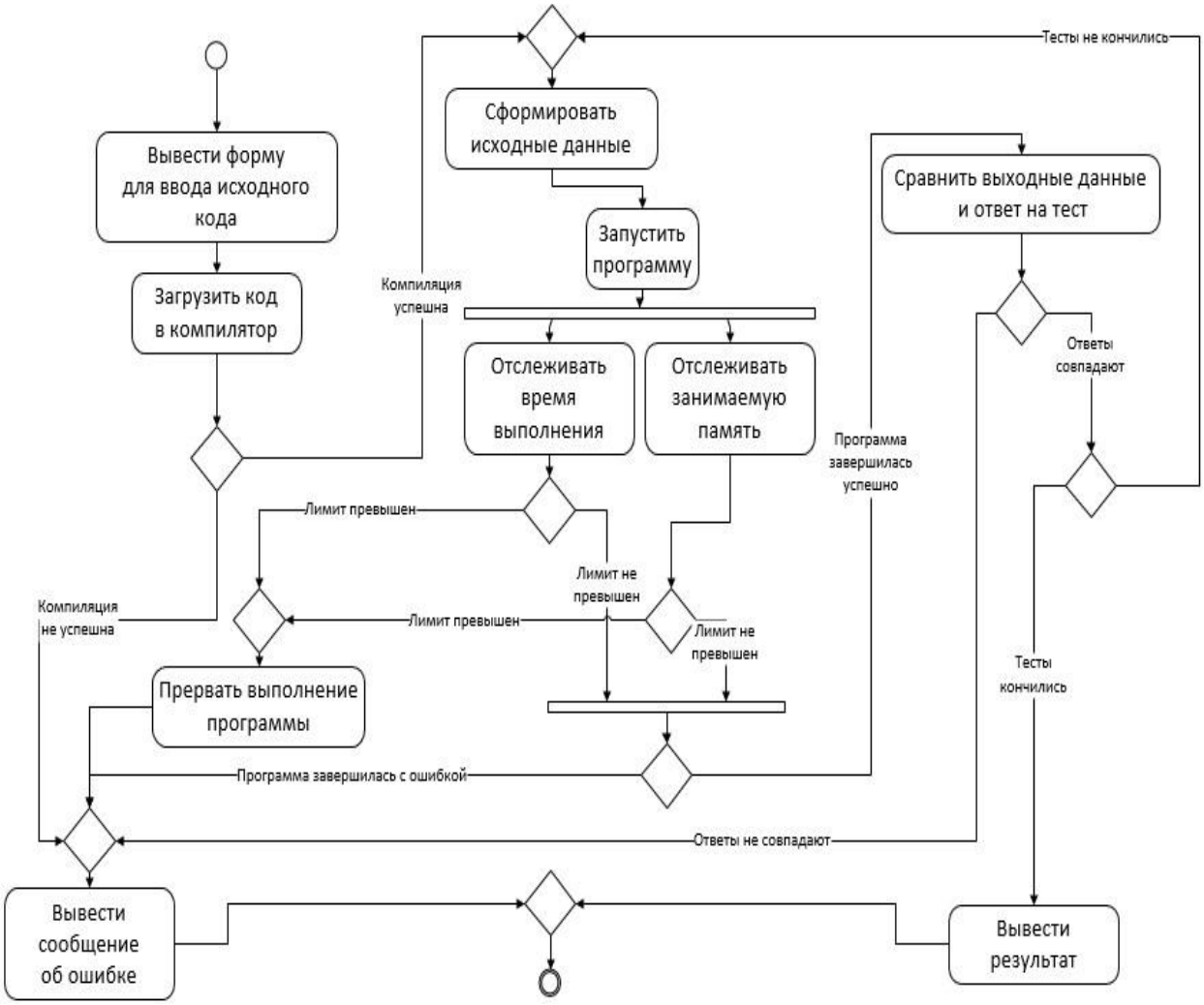


Рис. 4 Диаграмма деятельности, ядро системы

Интерфейс

Учитывая назначение системы, интерфейс должен быть простым и лаконичным. Таким образом, основной интерфейс пользователя состоит из 6 страниц:

1) Список задач, в котором представлены номера и названия задач.

2) Страница задачи, на которой представлено условие задачи и примеры тестов.

3) Форма отправки решения с тремя полями: номер задачи, выбор языка программирования из списка, текстовое поле для исходного кода.

4) Таблица результатов проверки, в которой указаны название задачи, имя пользователя, время отправки, время работы, статус и номер теста, на котором был неверный ответ, если не все тесты пройдены, сообщение об ошибке (открывается на новой странице), если есть ошибка.

5) Страница со списком прошедших соревнований.

6) Итоги соревнования в виде таблицы, в которой показано, сколько тестов прошёл каждый участник соревнования в каждой задаче.

Выводы

Проектируемая система разрабатывается на кафедре ПОКС КГТУ им.И.Раззакова, это даёт КГТУ возможность организовывать свои соревнования, в частности республиканские олимпиады по информатике для школьников и олимпиады по программированию среди студентов нашего университета, что и является основной причиной для разработки. Текущая разработка предлагает те же основные функции, что и системы-аналоги, но отличается тем, что подсчёт баллов будет вестись и по правилам ACM ICPC для студенческих олимпиад по программированию, и по правилам IOI для школьных олимпиад по информатике. Разработка этой системы выводит проведение Республиканской олимпиады школьников по информатике в Кыргызстане на уровень международных олимпиад.

Список литературы

1. Сайт проведения международных студенческих командных соревнований: https://icpc.baylor.edu
2. Сайт проверочной системы КРСУ http://olymp.krsu.edu.kg/
3. Сайт проверочной системы ТИМУС http://acm.timus.ru/
4. Корнеев Г.А., Станкевич А.С. Методы тестирования решений задач на соревнованиях по программированию. //Труды II межвузовской конференции молодых ученых. – СПб.: СПбГУ ИТМО, 2004. С.36-40.
5. Станкевич А.С. Общий подход к подведению итогов соревнований по программированию при использовании различных систем оценки. //Компьютерные инструменты в образовании. №2, 2011. С 27-38.
6. Макиева З.Дж., Каримова Г.Т., Макаева А.Д. Требования к веб-ориентированной информационной системе по секторальной квалификационной рамке Кыргызской Республики для ИКТ направлений. //Теоретический и научно-методический журнал «Вестник университета (ГУУ)» ФГБОУ ВПО «Государственный университет управления», Москва, № 19/2014. С.167-172.

UDC 025.4.03

APPLICATION OF MORPHOLOGICAL MARKUP OF KAZAKH LANGUAGE TO AUTOMATED FILLING OF THE ONTOLOGY OF FACTOGRAPHIC RETRIEVAL SYSTEM

Madina Mansurova, *al-Farabi Kazakh National University, Almaty, Kazakhstan,.*
Kairat Koibagarov, *Institute of Information and Computational Technologies, Almaty, Kazakhstan.*
Vladimir Barakhnin, *Institute of Computational Technologies SB RAS, Novosibirsk, Russia, Novosibirsk State University, Novosibirsk, Russia.*
Madina Soltangeldinova, *al-Farabi Kazakh National University, Almaty, Kazakhstan.*
Serzhan Berdibekov, *al-Farabi Kazakh National University, Almaty, Kazakhstan.*

Abstract. This work is concerned with the development of the parser for automation of morphological markup of the texts of the Kazakh National corpus. The parser includes the lexical and morphological analyzers to perform the morphological markup of the texts. The task of a lexical analyzer is to determine the boundaries of sentences, to display words, identifiers and punctuation marks. The morphological analyzer performs the search for words in the dictionary (which is a separate database) and determines their morphological parameters. At the output of the morphological analyzer, we have a list of lemmas (a normal form of the word), affixes and morphological characteristics of the word. Morphological markup of the texts is a stage of automatic

text processing, which allows to use the marked texts to solve the different problems of Natural Language processing. This paper describes the application of morphological parser of the Kazakh language to automated filling of the ontology of factographic retrieval system.

Keywords: Morphological Parser, Morphological Markup, Factogaphic Retrieval, Facts Extraction, Automated Ontology Filling.

ПРИМЕНЕНИЕ МОРФОЛОГИЧЕСКОГО АНАЛИЗАТОРА КАЗАХСКОГО ЯЗЫКА ДЛЯ АВТОМАТИЗИРОВАННОГО НАПОЛНЕНИЯ ОНТОЛОГИИ ФАКТОГРАФИЧЕСКОЙ ПОИСКОВОЙ СИСТЕМЫ

-Фараби.

Мансурова М.Е,*Казахский национальный университет имени ал*

.

Койбагаров К.Ч, *Институт информационных и вычислительных технологий МОН РК.*

Барахнин В.Б., *Институт вычислительных технологий СО РАН.*

Солтангельдинова М., *Казахский национальный университет имени ал-Фараби.* ***Бердибеков С.****,Казахский национальный университет имени ал-Фараби.*

Аннотация. Данная работа посвящена разработке анализатора для автоматизации морфологической разметки текстов корпуса казахского языка. Для осуществления морфологической разметки используются лексический и морфологический анализаторы. Задачей лексического анализатора является определение границ предложений, выделение слов, идентификаторов и пунктуационных маркеров. Морфологический анализатор выполняет поиск слов в словаре казахского языка и определяет их морфологические параметры. На выходе морфологического анализатора мы получим список лемм (нормальная форма слова), аффиксов и морфологических характеристик слова. Осуществляемая с помощью разработанного анализатора морфологическая разметка является этапом автоматической обработки текста, которая позволяет осуществлять поиск нужных пользователю слов, форм слова, лексических конструкций и т.д. В данной работе описывается применение модуля морфологического анализатора для автоматизированного наполнения онтологии фактографической поисковой системы.

Ключевые слова: морфологический анализатор, морфологическая разметка, фактографический поиск, извлечение фактов, автоматизированное наполнение онтологии.

Introduction

In Turkic philology, there are quite many investigations on this problem for cognate languages based on different conceptual approaches [1; 2; 3; 4; 5]. Analysis of the study on open publications in the field of the technology of morphological analysis of the Kazakh language word forms shows that in this field there are practically few investigations.

In the period from 1970 to 2000, the publications in the field of the Kazakh language morphology were largely theoretical. Since 2006 there were the valuable publications in international journals: Jonathan North Washington (2006) "A Novel Approach to Delineating Kazakh's Five Present Tenses: Lexical Aspect"; G. Altenbek and Wang Xiao-long (2010) "Kazakh Segmentation System of Inflectional Affixes"[6]; Zafer H.R., Tilki B., Kurt A., Kara M. “Two-level description of Kazakh morphology” (2011) [7]. Particular attention should be given to the work of Sharipbaev A.A. "Intelligent morphological analyzer based on semantic networks" (2012) [8]. All the listed works deal with some aspects in the field of morphology and syntax of the Kazakh

language and have a theoretical character of investigations. In this relation, creation of a module for morphological analysis of the Kazakh words at a high processing rate is actual.

The remaining part of the article is structured as follows: Section 2 describes of development of the parser for automation of morphological markup of the Kazakh-language texts. Section 3 provides a detailed description of the technology of automated factographic retrieval system ontology filling. This technology contains extracting keywords from corpus of texts with similar topic for following using these keywords as possible values of entity's attributes. Next, Section 4 describes the practical results. Finally, the article is completed by a relevant discussion and further research directions.

Development of the parser for automation of morphological markup of the texts of the Kazakh National corpus

Peculiarities of the Kazakh morphology

The Kazakh language refers to the class of agglutinative languages and together with Uzbek, Kyrgyz, Bashkir, Tatar, Azerbaijani, Turkish and other languages forms a Turkic linguistic family. Agglutinative languages are characterized by a consecutive addition of suffixes or ending bearing a grammatical meaning to an unchangeable root or stem having a lexical meaning.

The sequence order of affixes is strict. For example, for nouns first a suffix is added to the stem and then the ending of the plural followed by a possessive ending, then comes a case ending and finally the ending of the conjugation form (which is only added to animate nouns) [9; 10].

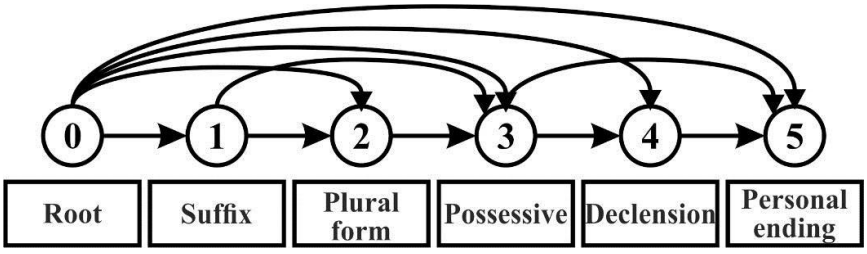


Fig. 1. Rules of affixes attachment for nouns

In the Kazakh language, there exists the law of vowel harmony: harmony between vowels and consonants of affix and the sounds of root. Vowels harmonize according to the hardness – softness principle and, as far as consonants are concerned, there is harmony between the final sound of root and the first sound of affix.

Apart from the three main rules of vowel harmony, it is necessary to take into account the following rules of exclusion:

1. The rule of removing a voiceless consonant in the affix being added if there are two voiceless consonants in the word ending. For example, *журналист + тер* (*zhurnalist + ter*) *журналистер* (*zhurnalister*), *экстремист+тер* (*ekstremist + ter*) *экстремистер* (*ekstremister*).

2. The law of vowel harmony is not observed in the following cases of the following affixes:

a) for affixes *мен*, *пен*, *бен* (*men*, *pen*, *ben*): *қаламмен* (*qalammen*); *нікі*, *дікі*, *тікі* (*niki*, *diki*, *tiki*): *баланікі* (*balaniki*);

b) for loan words with ending: : *рк*, *нк*, *кс*, *кт* (*rk*, *nk*, *ks*, *km*): – *пукнте* (*punkte*). The law of omitting vowels "i", "ы" ("i", "y") in the root, when adding a possessive affix "i", "ы" ("i", "y"). For example: *әріп – әрпі* (*arip – arpi*), *қойын – қойны* (*koiyn – koiny*).

The structure of a linguistic parser

Figure 2 presents the developed by us linguistic parser consisting of four analyzers (lexical, morphological, syntactic and semantic). The analyzers are successively arranged one after other, the output flow of one analyzer serving as an input for the following one.

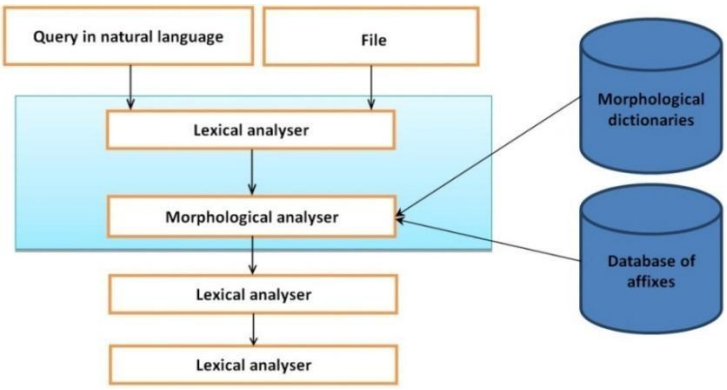


Fig. 2. The structure of linguistic parser

The task of a lexical analyzer is to determine the boundaries of sentences, to display words, identifiers and punctuation marks. The morphological analyzer performs the search for words in the dictionary (which is a separate database) and determines their morphological parameters (e.g., part of speech, number, case, etc.). A syntactic analyzer constructs a syntactic graph of a sentence. In this work, we study two analyzers – a lexical and morphological analyzer.

A lexical analyzer is a program of initial analysis of a natural text presented in the form of a chain of Unicode symbols. The output information is needed to be processed further by morphological and syntactic analyzers.

The tasks of a lexical analyzer include:

Division of the input text into words, numbers, disjunctives, etc.

Isolation of idioms not having word changing variants;

Isolation of proper names, FNP (family name, name, patronymic) when the name and patronymic are presented by initials;

Isolation of e-mail addresses and file names;

Isolation of sentences from the input text.

The procedure of isolation of words, numbers and punctuation marks is quite evident. After reading – out the next paragraph in turn, the graphematic analyzer parses tokens and assigns the corresponding graphematic characteristics to them. At this stage, tokens are isolated according to blanks and punctuation marks. However, of the greatest difficultly is determination of the beginning and the end of sentence. A lexical analyzer contains a heuristic mechanism for determination of the sentence boundaries and the result of lexical analysis is not only an array of lexemes, but also indicators of the beginning and end of the current sentence in the text.

It is not simple to find the end of a sentence either, as it may seem. An exclamation or question mark is sure to indicate the end of a sentence, but a point may be put after an abbreviation and in the middle of a decimal fraction. One must also take into account complex units of measurement (кв.м, км/час), internet – addresses (http://yandex.ru), ordinal numbers written out in figures (1917- ж.), markup names (Kassymov K.S). Let us consider the following fragment:

1917 ж. 21-26 шілдеде Покровский С.И. Орынборда болған «Бүкілқазақтық» съезде

(1917 *zh.* 21 – 26 *shildede Pokrovckiy S.I. Orynborda bolghan "Bukilqazaqtyq" syezde*).

There are four points and only the fourth points indicates the end of the sentence. Therefore, special test units are introduced into the analyzer. In particular, the simplest the simplest check: a lexeme just before the point must contain at least one vowel. This test makes it possible to choose many abbreviations used with a point in the end: *ж., т.ғ.к., тғ., ф.-м.ғ.д., қ.* **(***zh., t.g.k., f.-m.g.d., q.*).

Of course, suck a test does not guarantee the correctness of the result. On the one hand, an abbreviation or a number with a point can be at the end of a sentence. Nevertheless, the experience in the work with documents confirms the efficiency of such testing. Only after analysis of points, which are not sentence markers, we can divide the paragraph into sentences and the further analysis is performed within one sentence.

Description of the algorithm of a morphological analyzer

The work of a morphological analyzer is as follows. At its input, we have an array of words, punctuation marks and numbers marked from the input text at the stage of lexical analysis, with lexical characteristics [11]. For every word, the analyzer performs the search for words in the dictionary loaded into memory. All the stems the word being analyzed can begin with are looked for. If a stem in turn satisfies this condition, a line containing all possible affixes for this stem is extracted from the dictionary of affixes. Each affix from this line is added by turns to the stem and the result is compared with the word being analyzed. In case of their exact coincidence, a new record is introduced into the list of the search results: according to the ordinal number of the affix in the line of affixes, variable morphological parameters of the word are determined (for example, for a noun – the number and case), and by the lexical information of this stem its constant parameters (noun, verb, adj…) are determined. If as a result of such search not a single successful variant if found, the user is requested to enter a new stem into the dictionary. In case he refuses to do it, performance of the morphological analysis is stopped. If the new word is introduced into the dictionary, the procedure of searching is repeated. At the output of the morphological analyzer, we have a list of lemmas (a normal form of the word) + affix + morphological characteristics of the word (part of speech, case, number).

Thus, the results of the morphological analysis in total are presented in the form of a dynamic array. The number of its elements is equal to the number of its lexemes in the sentence. The elements of the array are other arrays each of which retains all possible interpretations of its lexemes homonyms. A dictionary of word stems, a dictionary of geographic names, a dictionary of names, a dictionary of affix conjunctions are used as initial lexical materials.

Algorithm is automated filling of the ontology of factographic retrieval system

The section describes the application of morphological parser of the Kazakh language to automated filling of the ontology of factographic retrieval system. The method of automated filling of the ontology is proposed in [12]. This algorithm allows to extract key words/ phrases from the text corpus of homogeneous subjects. The extracted key words are used further as possible meaning of attributes of matters described in the created ontology of the subject field designed for organization of factographic retrieval. The text preliminarily marked up with the help of a specially developed morphological analyzer is used as input data. To extract semantically related key words/ phrases, the method of random walks is used in the algorithm. A trained neural network with a hidden layer is used for the set of these phrases with the aim to assign a concrete phrase to the definite attribute of the matter described in the text. Owing to the neural network operation, ontology for a concrete document on the semantically related word pairs set is constructed and then, on the basis of the obtained ontology, factographic retrieval is organized.

Experiments

The biography of Kazakh poetess Fariza Ongarsynova is taken as an example of a complete cycle of the algorithm work. At the first stage, the text of biography is marked up by parts of speech. For example: *Фариза Оңғарсынқызы Оңғарсынова - қазақ ақыны, халық жазушысы, журналист. 1939 жылы 5 желтоқсанда Гурьев (қазіргі Атырау) облысы, Новобогат ауданына қарасты Манаш ауылында туған.* In this form, the document is introduced in the database MongoDB which is oriented to store collections of JSON documents.

Then, using the random walk method, key words and phrases are extracted, part of them are presented below: *Фариза Оңғарсынқызы Оңғарсынова, ақын, халық жазушы, 1939 жылы 5 желтоқсанда, Гурьев облысы, Атырау, Новобогат ауданы, Манаш ауылы, etc.*

And in the last stage, the neural network places the data on the descriptors.

"name": "*фариза оңғарсынқызы оңғарсынова*"

"position": "*ақын", "жазушы*"

"date_of_birth": *"1939 жыл 5 желтоқсан"*

"date_of_death": "2014 жылдың 23 қаңтар"

Conclusion

The work proposes the technology of automated filling of the ontology of factographic retrieval system. The proposed technology allows to markup parts of speech in the text. The authors cover the progress of the work on supplying the corpus of the Kazakh language texts with scientific framework. While making morphological marking they considered the features of the Kazakh word form change, and created the models of variable nominal and verbal word forms and the lists of their word changing affixes. In the future we plan to continue research to develop modules of syntactic and semantic analysers that expand opportunities for the filling of the ontology of factographic retrieval system.

References

1. Makhmudov M.: Systems of automatic recycling of Turkic text on lexical and morphological level, Elm, 114 p. Baku (1991) (In Russian)
2. Migalkin V.V.: Modeling of the Yakut language spelling and development a set of programs to check the spelling of Yakut texts in Windows environment, Author. diss.... Ph.D., Yakutsk (2005) (In Russian)
3. Sadyqov T.: Problems of modeling of Turkic morphology: an aspect of causing Kyrgyz nominal inflectional forms, 119 p. Publishing House of the "Ilim" (1987) (In Russian)
4. Sirazitdinov Z.A.: Modeling grammar of Bashkir language. Inflectional system. 160 p. Ufa (2006) (In Russian)
5. Sirazitdinov Z.A.: On the modeling of inflectional system agglutinative language pair combinations (for example, the Bashkir language) / Actual problems of modern Mongolian and Altaic. Proceedings of the International Scientific Conference. Elista, 2014. pp 139-143. (In Russian)
6. Altenbek G., Wang Xiao-long: Kazakh Segmentation System of Inflectional Affixes. In:

Joint Conference on Chinese Language Processing, pp.183-190 (2010)

7. Zafer H.R., Tilki B., Kurt A., Kara M.: Two-level description of Kazakh morphology. In: Proceedings of the first International Conference on Foreign Language teaching and Applied Linguistics, FLTAL 2011, Sarajevo (May 2011).
8. Sharipbaev A.A.: Intelligent morphological analyzer, based on semantic networks: Conference proceedings Open Semantic Technologies for Intelligent Systems (2012)
9. Bekmanova G.T.: Some approaches to the problems of automatic word changes and morphological analysis in the Kazakh language. In: Bulletin of the East Kazakhstan State Technical University Named by D. Serikbayev, №1, pp. 192-197, Ust-Kamenogorsk (2009) (In Russian) 10. Zhubanov A.H.: Basic principles of formalization of the Kazakh text content, 250 p. Almaty (2002) (In Russian)

УДК 81.374:811.512.161:811.512.154

PARSING AND ANNOTATION OF TURKISH-KYRGYZ DICTIONARY

Kadyr Momunaliev, *Kyrgyz-Turkish Manas University kadyr.momunaliev@gmai.com*

The case study described in this article is the first milestone on the way toward a full featured Text Encoding Initiative P5 annotation standard. The paper outlines parsing and annotating workflow to obtain initial XML-based structure of Turkish-Kyrgyz dictionary. Typography-based parsing techniques are implemented in procedural programming language environment; corresponding workflow charts are presented in form of pseudo code and block schemas. Resulted XML dictionary bases are verified and applied in desktop and web e-dictionary implementations. It is proposed that such kind of explicitly structured data representation is easier to manipulate and use as a basement for further deeper lexicographic annotations.

Keywords: dictionary data, structured data, unstructured data, XML, human-computer, typography, semantics, syntax, parsing

ПАРСИРОВАНИЕ И АННОТИРОВАНИЕ ТУРЕЦКО-КЫРГЫЗСКОГО СЛОВАРЯ

Момуналиев Кадыр Замирович, *Кыргызско-Турецкий Университет «Манас» kadyr.momunaliev@gmai.com*

Данное тематическое исследование представляет собой один из пройденных этапов на пути к достижению полноценного стандарта аннотирования Text Encoding Initiative P5.
Статья освещает методологию парсинга и аннотирования результатом которой является XML структурированные данные турецко-кыргызского словаря. Приемы парсинга основанные на типографических свойствах текста реализованы в процедурной среде прогараммирования; соответствующие алгоритмы представлены в виде блок схем и псевдо кода. Полученные XML структуры были верифицированы и применены в офф-лайн и веб приложениях типа электронного словаря. В качестве технического предложения утверждается, что аннотированные и структурированные данные легче поддаются машинной обработки и представляют собой фундамент для более углубленного лексикографического анализа.

Ключевые слова: словарные данные, структурированные данные, неструктурированные данные, XML, человек-компьютер, типография, семантика, синтаксис, парсинг

1. Introduction

Today big amounts of information accumulated in electronic media or in printed form require universal, i.e. globally accessible and efficient way of representation since the world is becoming a worldwide network of information exchange and business transactions [1]. This information from different domains of human activity needs not only physical infrastructure allowing transmission of bytes, files and documents, but should allow transmission of their semantics, i.e. their meaning. In the coming future of Semantic Web computers are going to
"understand" the contents of documents, i.e. navigate, read, access and capture relevant information. Human's part of information processing by browsing and reading is less efficient and thus cannot provide a competitive advantage.

In such a setting new standards of information representation that is purposed for computer access are being sought for. Some new technologies allowing keeping information semantics in computer-readable manner has already appeared (XML and RDF). These types of information

provide explicit semantic structure and often called machine-readable or machine-accessible formats.

The trend of data conversion to computer-readable format is not an exception for dictionaries. Many dictionaries are under their way to be converted and others are already done. There is a range of formats to encode dictionary data: binary, relational database etc. But why to choose XML –based formats, such as XDXF [2], StarDict Textual File Format, TEI P5 etc.? The main reason is that they may be used as a cross format representations, i.e. universally interchangeable data containers or structures. Many software agents are able to work with XML, since it has open and predictable structure. Even a human being is able to read and edit such files

They don't require special purpose converters to convert from one format to another, since XML query and transformation languages provide mapping standards between different XML schemas [2].

In our case study we have a formatted text representation of a dictionary that should be analyzed to obtain lexical (or editorial) XML encoded representation reflecting semantic structure of the dictionary. Parsing in this context implies splitting up textual body of a human readable document to explicitly shown lexicographical units in form of XML tree. Factually it is a conversion of human readable data to machine readable format. *Figure 1* demonstrates what data is given as input and what may be gotten as output.

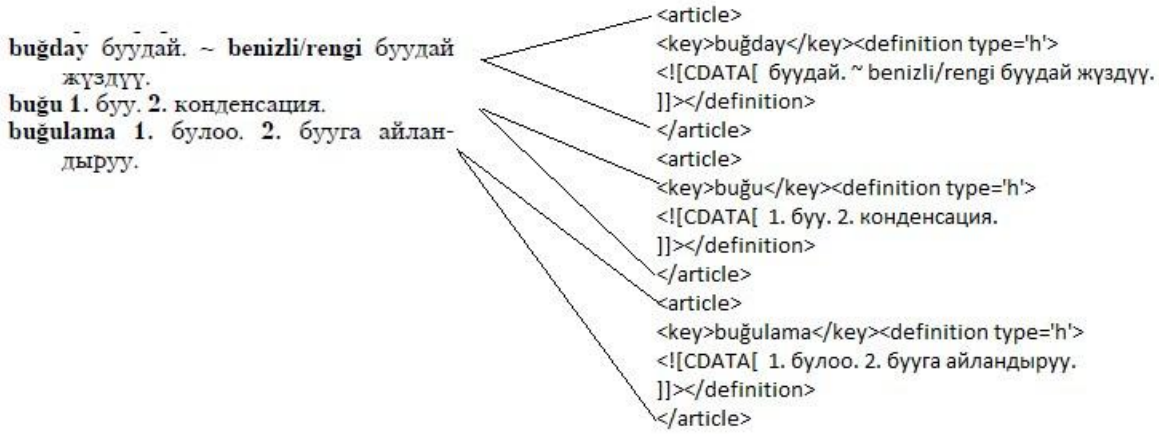


Figure 3. The sample of input(left) and output(right) data (Textual file schema) of the parser.

Why to encode dictionary data? Because if there were no machine-readable formats machines/computers would have to generate parsers for every dictionary they encounter since every dictionary may have its own specific structure and rules to interpret the content. Secondly dictionaries are rather static information sources that may be parsed once for further multiple usages.

2. Input Data

The Turkish-Kyrgyz Dictionary by Gulzura Cumakunova [3] (hereinafter Cumakunova's dictionary) has been chosen as an input data to be parsed and annotated. The dictionary is structured and formatted according to international lexicographical rules. In the input file all headings, footers, page numbers, first matter, last matter has been temporarily removed, and their annotations are not subject of this paper. The input data implies a series of dictionary entries readable by human. The

[2] XDXF (XML Dictionary eXchange Format) is a project to unite all existing open dictionaries and provide both users and developers with universal XML-based format, convertible from and to other popular formats like Mova, PtkDic, StarDict. Retrieved from www.en.wikipedia.org

[3] More information about the dictionary author could be found on: https://ky.wikipedia.org/wiki/Жумакунова,_Гулзура

main body of the dictionary data has hierarchical structure and is shown in form of a schematic tree, see *Figure 2*.

Description of *Figure 2:*

1. All dictionary entries belong to the dictionary, i.e. <dictionary/> unit is the root element
2. There must be at least one dictionary entry in the <dictionary/>, i.e. <entries/>1+
3. Each entry consists of mandatory form and definition parts: <form/> and <definition/>. Form part contains Turkish source text and definition part contains Kyrgyz target text.
4. Form block consists of two parts: 1st part contains headword (<headword/>) that may be single headword or the first word of the compound phrase, or acronym; 2nd part contains information related to <headword/> that may be special ending causing headword to inflect (<ending/>), or it may contain the compound phrase(s) in full view, or very close synonym, or acronym's expanded view. (see the tree leaves at the left)
5. Definition block is divided by two areas or blocks: sense and usage (in the tree <sense_block/> and <usage_block/>).
   a. Sense block contains one or more sense items which in turn comprise one or more options of translation equivalents very close by meaning and delimited by commas. Translation equivalents may be accompanied by sense example which consists of Turkish source text and Kyrgyz target text. Full stop character delimits translation equivalents from sense example. And finally there may be two kinds of references: 'see' reference and 'compare' reference. 'See' reference comprises all the sense item space and nothing may accompany it. 'Compare' reference is an appendix for translation(s) (as like as sense example). Both of references are preceded by reserved abbreviations in Kyrgyz('к.' abbreviated form of 'кара' – 'see', and 'сал.' abbreviated form of 'салыштыр' – 'compare').
   b. Usage block is optional and if it occurs then it may consist of one or more usage items (i.e. samples of stereotypic phrases, idioms, proverbs etc) which don't necessarily explain the certain sense of the headword, but show its usage. Usage items are not attached to certain sense item, instead they refer to the headword in general. There is no special delimiter dividing sense items group from usage items, but usage items come after sense items. Any usage item is represented by source Turkish phrase and its target Kyrgyz translation(s) delimited by round bracket enumeration (like '1)', '2)' etc). Translations of usage item are not terminated by full stop.
6. Additionally any leaf of the tree (rather sense or usage children) may contain author's explanation note placed in the round brackets and outlined by italic font. (This is not shown in the diagram.)

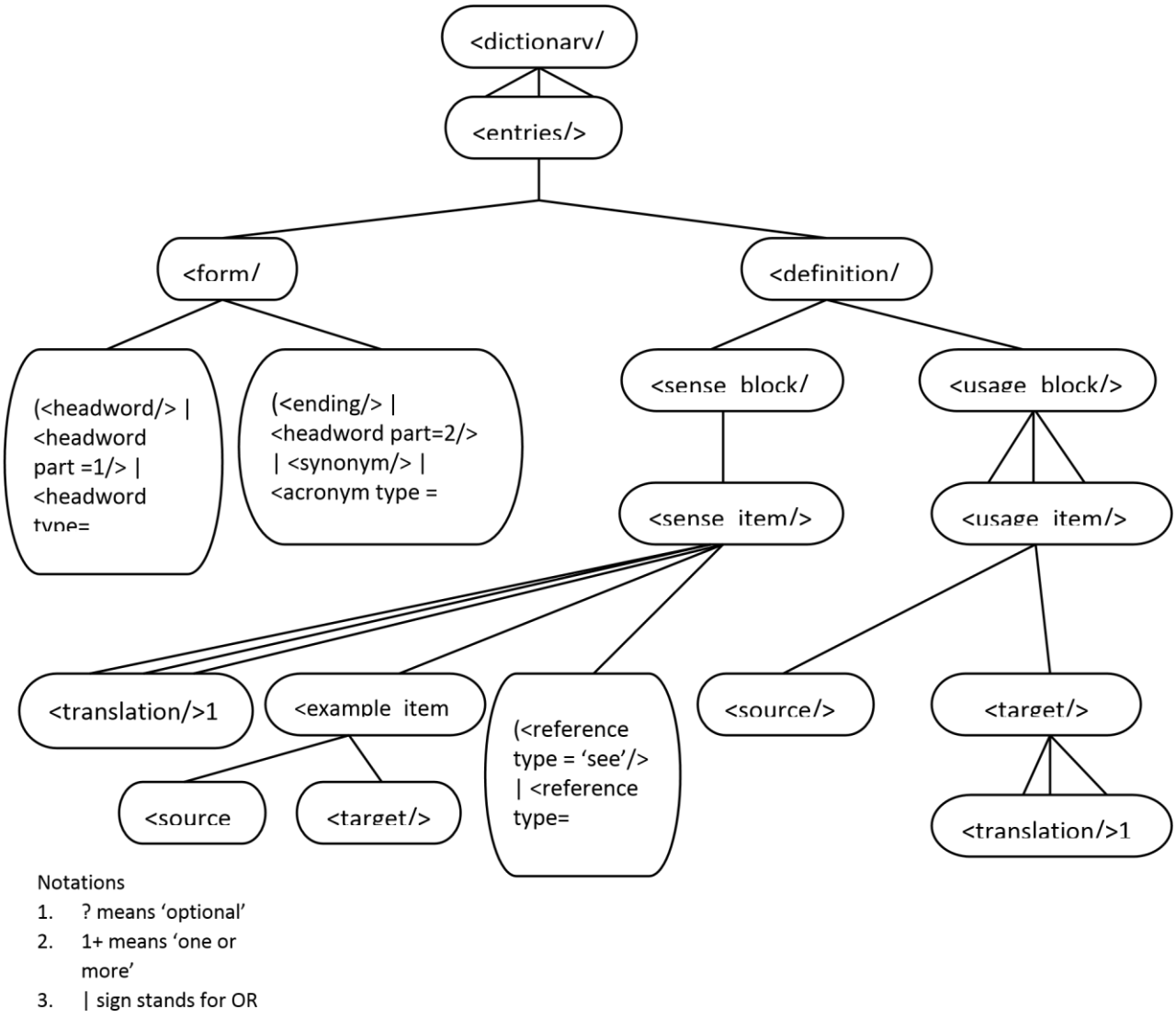


Figure 4. Lexicographical structure of the dictionary data.

The parser is intended to make use of typography features (pre-annotated, i.e. converted to XML tags), syntax and reserved constructions that described in the following tables (*Table 1*,*2*.*3*). Dictionary's typographical features and their corresponding lexicographical meanings are summarized in *Table 1*

Table 1. Typography-to-semantics interpretation schema

	Typographic Indicators			Semantics or Lexicographical meaning
Content unit	Layout	Formatting	Character set	
Text		Bold font	Latin Turkish	Turkish content (Headword, idiom, stereotypic phrases and other source matter)

Text		Normal font	Cyrillic Kyrgyz	Kyrgyz content (senses, translations and other target matter)
Text		Italic font	Latin	International names of flora and fauna species
			Cyrillic Kyrgyz	Additional explanations
			Cyrillic Kyrgyz	abbreviations[4]
Line	Out dented, i.e. hanging, i.e. shifted to the left			First line of the entry
Word	locates at the beginning of hanging line	Bold font	Latin Turkish	Headword
Text	locates at the beginning of a hanging line	Bold font	Latin Turkish	Form block[7]
Word			First letter capitalization	Proper name
Word			Full capitalization	Acronym

Syntax, i.e. punctuation meanings are shown in *Table 2.*

Table 2. Syntax of the dictionary or punctuation semantics

Punctuation name	Notation	Additional indicators	Function or meaning
Swung dash	~	bold	Headword placeholder
Slash	/		Delimits equally suitable words in a given context
Full stop	.		Sense item's end
			Usage item's end
			Example item's end
			Sense block translations' end
			Accompanies abbreviations
		bold	Accompanies sense items' enumeration numbers
Comma	,		Delimits sense block translation variants
			Comes after headword showing that it has additional information(ending or synonym)

[4] Kyrgyz abbreviations are followed by mandatory full stop and don't change, i.e. they are fixed pre-defined words. [7] There are two exceptions with 'же' (or) and acronym expanded form, see section Entry Parsing

Colon	:	Bold	Indicates that headword is not used by itself, but constitutes compound phrase that follows immediately
Round brackets	(and)		Embrace additional explanations or notes
		Only right one	Follows usage translations enumeration numbers
Dash	-	Followed by end of line symbol or tag	Line break
			Compound or united word
		Bold, preceded by space character or comma	Ending prefix in the form block

There are also reserved character constructions or/and abbreviations that also indicate or delimit some lexicographical units, see *Table 3.*

Table 3. Reserved characters and words

Character or word	Notation	Additional indicator	Construction	Function or meaning
Roman digits	I, II, II etc.	bold		Hyponym's number
Arabic digits	1, 2, 3 etc.	Bold, followed by full stop	1. | 2. | 3. etc.	Beginning of a sense item and its number
Arabic digits	1, 2, 3 etc.	Bold, followed by end round bracket	| 2) | 3) etc	Beginning of the usage item translation and its number
Kyrgyz abbreviations	See [7]	Italic, followed by full stop	[abbr].	Provide additional information or cut down repetitive text usage, see meanings in [7]

Relying on the above mentioned typographic, syntactic and reserved content features and their corresponding meanings it is possible to define some rules to delimit certain lexicographical units, see *Table 4.* This is the main principle of the parsing and annotation process which is described in the following section.

3. Method

When parsing formatted text data, there are three main features document needs to have to be parsed:

1. Layout
2. Formatting
3. Content patterns (repetitive text sequences)

Parsing is arranged relying on the one of this feature or combination of them since each of the mentioned features refers to a certain semantic unit of human readable textual data. For example, having bold word(s) in the beginning and out dented (shifted to the left) line most probably refer s to the next dictionary entry (here bold is a formatting feature and out dentation is a layout feature (see *Figure 1)).* Combinations of such kind of features give us opportunity to define rules to delimit certain lexicographic unit.

Table 4. Some rules for lexicographical units' delimitation

Delimiter	Notation	Additional feature	Delimits What
Comma + space + dash	, -		Headword and the suffix that changes headword ending
Colon	:	bold	First word from remainder part of a headword in complex headword cases
Comma	,		Semantically close meanings of a sense
Full stop	.		One sense from another sense; one translation from another translation

THE OBJECTIVE: to detect beginning and ending points of every lexicographical unit and mark it up with an appropriate XML tag.

Depending on pursued goals dictionary data may be encoded in different views. It depends on what kind of information is going to be captured via encoding. There are three main views (or information aspects) among others dealing with complexity of both typography and information structure.

- (a) the typographic view — the two-dimensional printed page, including information about line and page breaks and other features of layout
- (b) the editorial view — the one-dimensional sequence of tokens which can be seen as the input to the typesetting process; the wording and punctuation of the text and the sequencing of items are visible in this view, but specifics of the typographic realization are not
- (c) the lexical view—this view includes the underlying information represented in a dictionary, without concern for its exact textual form [3].

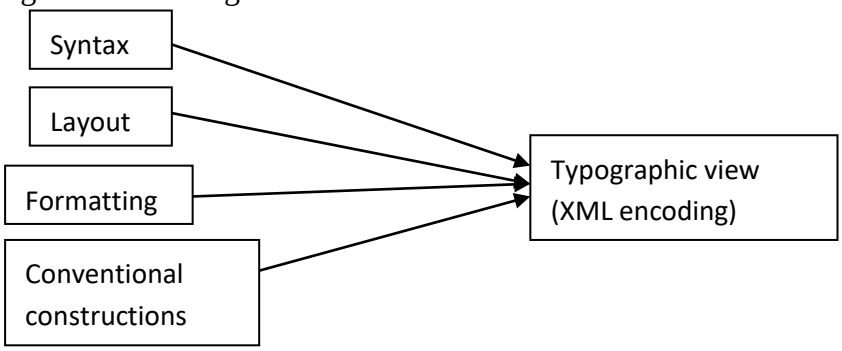
In this context our objective is to obtain editorial view of the dictionary. But before this, typographical view may simplify the whole task of annotation processes.

Pre-annotation Phase

Parser is intended to accept pre-annotated typographical view of the dictionary. Preannotation implies converting visually detectable typographical features to XML format. At the beginning it is easier to obtain typographic (not editorial or lexical) view in XML representation from formatted source text, because this procedure doesn't require any delimitation or parsing techniques, only converting layout, formatting and special purpose content (content indicators) into form of xml tags. For example, "some string in bold font" would be represented as "<bold> some string in bold font</bold>"; or to encode the order of lines every line might have its number: <line number = 10>the content of tenth line</line>; and content information such as abbreviations should be marked up by <abr/> tag etc. It is like an html code of a document, since html was originally intended to reflect the document structure (but not semantics). We can use names of tags and attributes to encode any information we need. All these pre-annotated XML elements will be useful in main annotation, i.e. parsing and marking up processes. Pre-annotation allows reduce the number

of different data types into one data type: linear XML (see *Fig.2.*). This process should preserve typographic information but in different representation.

Figure 5. Reducing different information media number to one linear XML.[5]



As a result of this simplification the code and logic of the parser becomes clearer, i.e. without excessive complexities of formatted text editors' inner data representation mechanisms. Having performed this reduction the programmer will need only XML processor (XSLT or other XML supporting languages) and a means to perform parsing of bare (without formatting) textual content. Data in the XML representation can be edited even in a simple text editor.

Parsing Technique Based on Delimitation & Markup

Entries Delimitation

The main issue of parsing was that not every dictionary entry captured exactly one paragraph. So another delimiter or delimitation mechanism had to be defined. The task of delimitation has various solutions. For example the most obvious solution would be to find every occurrence of line that is shifted to the left. This solution exploits layout feature of the typographical view, see *Table 1 and Figure 2.*

Figure 6. The main issue in entries delimitation task.

Another solution is to implement content-based parsing scenario where new delimiter (string of characters), i.e. text that certainly refers to the two articles meeting spot should be defined. Whatever parser's principle is the workflow of entries delimitation and basic structure delineation would be as following (see *Listing 2*): (In fact we chose the latter one, which defines new delimiter string)

Listing 1. Dictionary entries delimitation workflow pseudo code

```
1   Begin
2   For each line in Dictionary Do
3   Select line
4   If line is out dented Then
```

[5] Conversion of punctuation characters to XML tags is rather optional

```
5       (first line of the entry found)
6       Set cursor to the beginning of the line
7       put end tag of previous entry (e.g. </article>)
8       put start tag of next entry (e.g. <article>)
9       Else
10      Proceed to the next iteration
11      End IF
12      End For
13      Set cursor to the very beginning of Document
14      Cut the first occurrence of close entry tag(</article>)
15      Set cursor to the very end of Document
16      Paste close tag cut in 14
17      End
```

Note: In this algorithm the very first entry will be preceded by excessive end tag (</article>) and the very last entry will lack ending tag (</article>), that is why lines 13-16 are added.

Entry Parsing

Form Block Parsing

Now that until this moment delimitation is performed, i.e. every dictionary entry is delimited and marked up with <article > tags we can access each of them one by one. This may be realized either by means of XPath or by means of procedural programming language with or without XML support. Choice of solution depends on programmer's preferences. But whatever the tool is, headword of an entry should be recognized as the first word in the form block, and form block is the sequence of Turkish characters in bold font starting from the very beginning of the entries content. By traversing every entry and finding the word satisfying this condition it is possible to implement form block markup function, see *Listing 2*.

Listing 2. Pseudo code of the form block parsing and marking

```
1       Begin
2       For each entry in Dictionary Do
3       Select content
4       Find text which is bold & Turkish
5       If word is found & it is at the beginning of content
6       Then
7       Mark it up as form block(Form block's found)
8       Select first word
9       Mark it up as headword
10      Select remainder part of content
11      If remainder part is not empty
12      Mark remainder part as headword tail
13      Parse headword tail (see Listing 3.)
14      End If
15      Else
16      Throw an exception (entry syntax is wrong)
17      End If
18      End For
19      End
```

Not every headword consists of one word or it may have specific cases when it changes or it may have very close synonyms. So that the headword variations (possible syntaxes) are:

1. headword
2. headword, -ending
3. headword: ~ headword tail
4. headword, Synonym
5. HEADWORD (Acronym's Expanded Form'текст кыск.)

Description: 1. Simple headword; 2. Headword with ending; 3. Complex headword: one meaning but several words; 4. two or more words very close by meaning and by spelling; 5. Abbreviated word with the meaning of each capital letter and text in Kyrgyz containing reserved contraction 'кыск.' (contraction of 'abbreviation' like 'abbr.').

Actually this information may be found in form block and it should be parsed (parser is described in *Listing 3.*)

Listing 3. Headword tail parsing and markup (location: dictionary/entry/form/)

```
BEGIN
CAPTURE <tail>
GET content of <tail> as Content
IF Content is empty THEN
EXIT
ELSE_IF Content starts with ', -' THEN
MARKUP text after ', -'as </ending>
UNMARK <tail/>
ELSE_IF Content starts with ':' THEN
MARKUP text after ': 'as <headword part=2/>
CAPTURE <headword>
RENAME <headword> to <headword part=1>
CAPTURE <tail/>
UNMARK <tail/>
ELSE_IF Content starts with ',' THEN
MARKUP text behind ', 'as <synonym/>
UNMARK <tail/>
ELSE_IF Content starts with ' (' THEN
MARKUP text behind ' ('as <acronym/>
UNMARK <tail/>
ELSE
Throw an exception (entry syntax is wrong)
END_IF
END
```

Definition Block Parsing (see Results)
Sense Block Parsing (see Results)
Usage Block Parsing (see Results)

4. Discussion

Input File Problem

Initially dictionary was available in pdf format. But since it was difficult to process text data in pdf file directly (or maybe we lack knowledge on how to work with it) it was decided to convert it to more text accessible formats like txt, html or doc. That is why AdobeAcrobat's (version 11.0.2) pdf converter was used, which in fact offers various types of files to save as, e.g. eps, html, docx, doc, pptx, xlsx, xml table, txt etc. Of course the most interesting format was xml, since we were

going to obtain such an encoding of data that would provide us with explicitly structured semantic units of the dictionary. As turned out AdobeAcrobat's xml file didn't preserve all the information we needed for parsing. That is why the most close to origin file doc format was chosen. Here beneath in the *Table 1*, we showed criteria which we took as a guide to choose a file type to start up with:

Table 5. AdobeAcrobat's output files comparison with regard to dictionary data completeness

File type	Formatting	Layout	Content
Doc	Preserved	Preserved	Preserved
Xml	Lost	Lost	Preserved

The only shortcoming of doc file was that it contained some misspellings or precisely saying character mis-encodings when Turkish character occurred in Kyrgyz word or vice versa. Probably this happened as a result of OCR reading when it attempted to read word with one encoding but the word was actually in different, e.g. in Turkish(Latin character set) word "aba" first character is in Cyrillic. This issue was fixed programmatically as part of the input file normalization. Another task was to remove all data except dictionary entries: introduction, usage notes, headings, page numbers etc.

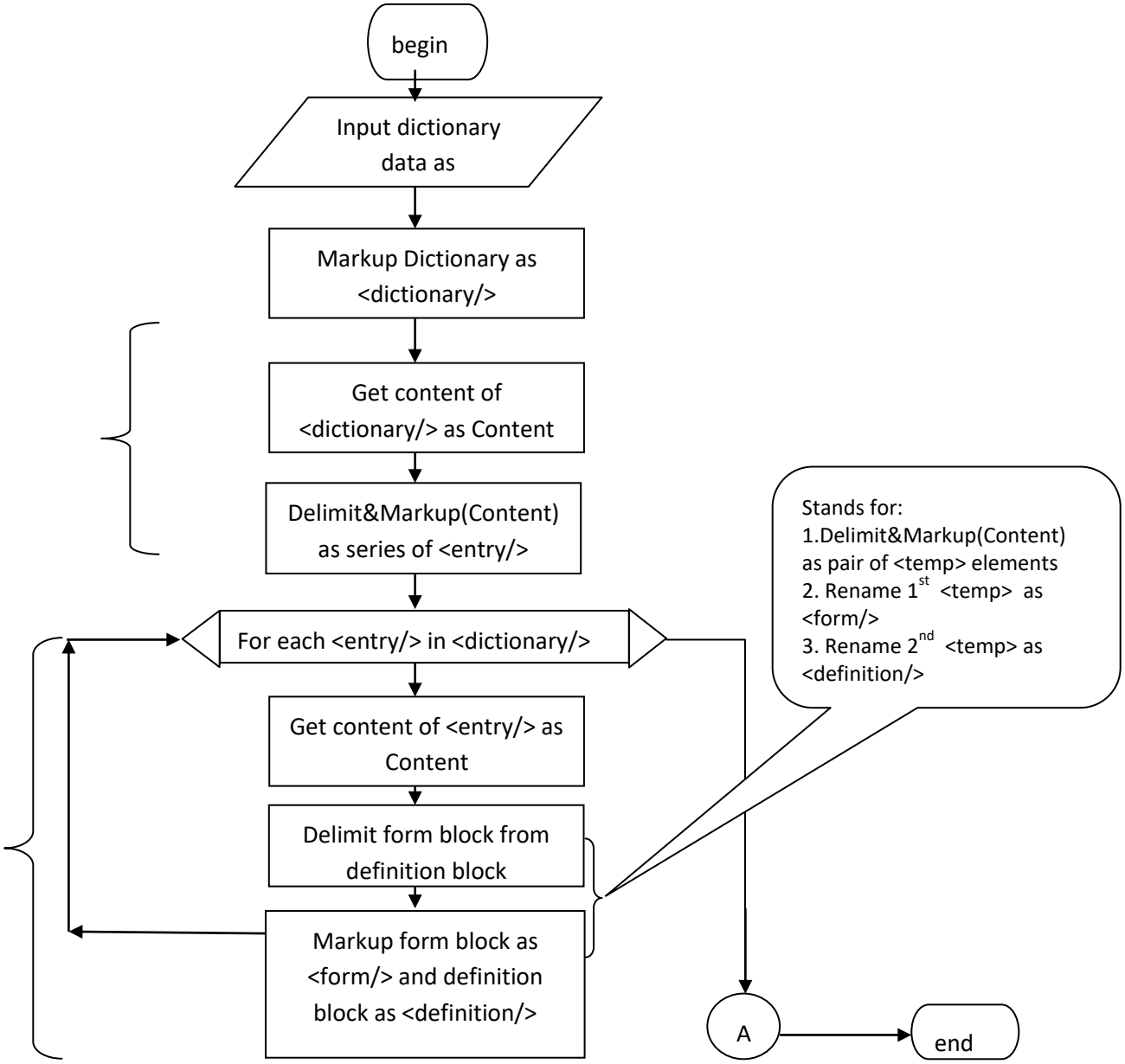
Ambiguity Problem

Any sense item including the last one theoretically may have a sense example. If so and if example and usage item are delimited in similar way there is no way to identify the item coming after the last sense weather it is example or usage item.

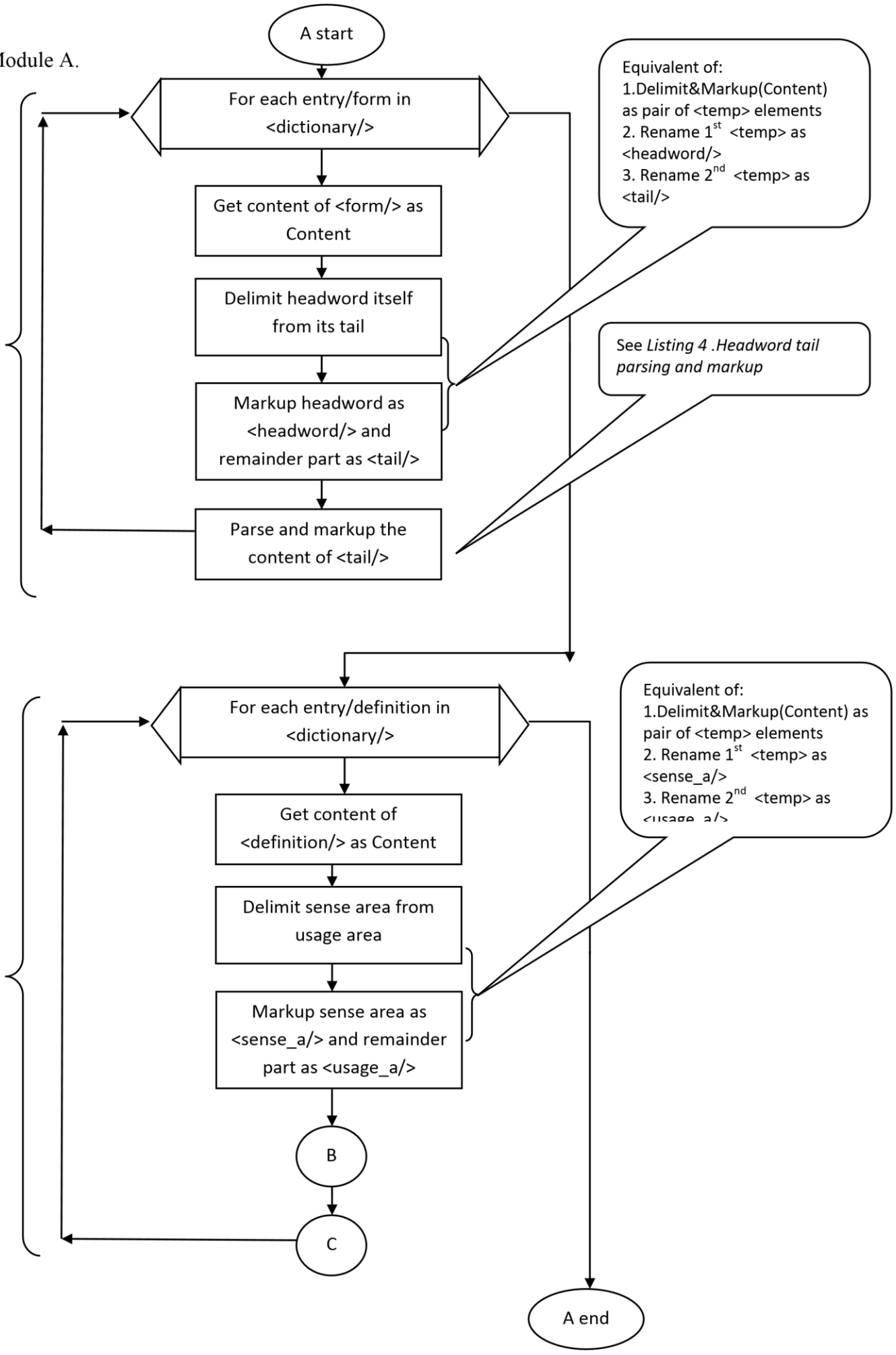
5. Results:

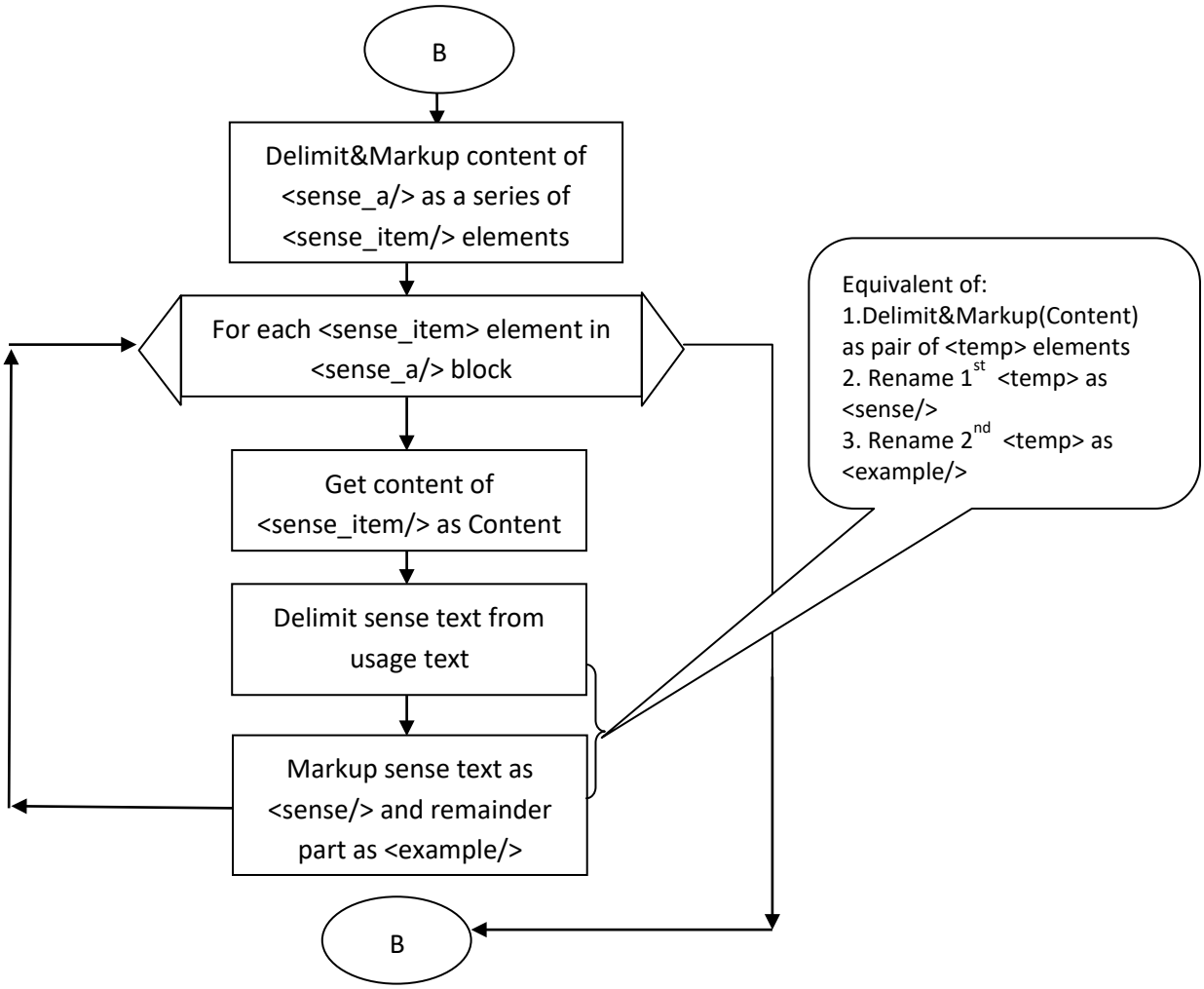
An Algorithm of the Parser. The main result of the work is the workflow of the Cumakunova's dictionary parser, see *Figure 5*. Node A is expanded in Module A, and nodes B and C in corresponding Modules B and C

Figure 7. The workflow of the parser

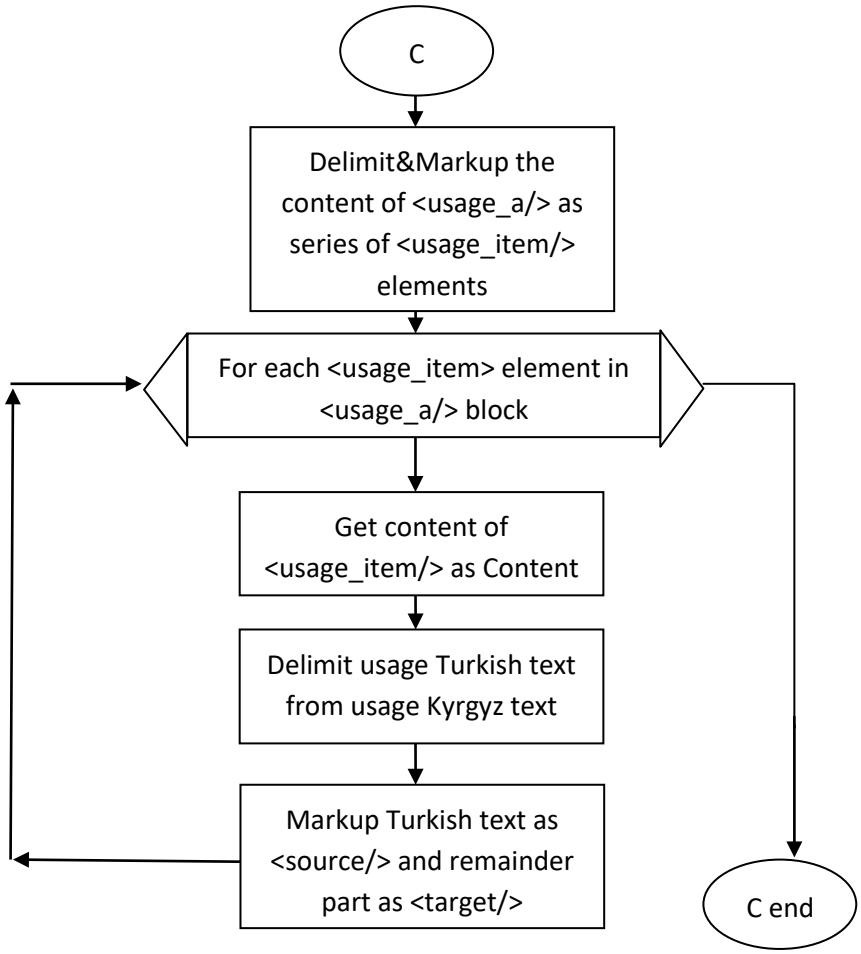

begin
Input dictionary data as
Markup Dictionary as <dictionary/>
Get content of <dictionary/> as Content
Delimit&Markup(Content) as series of <entry/>
For each <entry/> in <dictionary/>
Get content of <entry/> as Content
Delimit form block from definition block
Markup form block as <form/> and definition block as <definition/>
Stands for:
1.Delimit&Markup(Content) as pair of <temp> elements
2. Rename 1st <temp> as <form/>
3. Rename 2nd <temp> as <definition/>
A
end

Module A.


A start
For each entry/form in <dictionary/>
Get content of <form/> as Content
Delimit headword itself from its tail
Markup headword as <headword/> and remainder part as <tail/>
Parse and markup the content of <tail/>
Equivalent of:
1.Delimit&Markup(Content) as pair of <temp> elements
2. Rename 1st <temp> as <headword/>
3. Rename 2nd <temp> as <tail/>
See Listing 4 .Headword tail parsing and markup
For each entry/definition in <dictionary/>
Get content of <definition/> as Content
Delimit sense area from usage area
Markup sense area as <sense_a/> and remainder part as <usage_a/>
Equivalent of:
1.Delimit&Markup(Content) as pair of <temp> elements
2. Rename 1st <temp> as <sense_a/>
3. Rename 2nd <temp> as
B
C
A end


B
Delimit&Markup content of <sense_a/> as a series of <sense_item/> elements
For each <sense_item> element in <sense_a/> block
Get content of <sense_item/> as Content
Delimit sense text from usage text
Markup sense text as <sense/> and remainder part as <example/>
B
Equivalent of:
1.Delimit&Markup(Content) as pair of <temp> elements
2. Rename 1st <temp> as <sense/>
3. Rename 2nd <temp> as <example/>

Module C


C
Delimit&Markup the content of <usage_a/> as series of <usage_item/> elements
For each <usage_item> element in <usage_a/> block
Get content of <usage_item/> as Content
Delimit usage Turkish text from usage Kyrgyz text
Markup Turkish text as <source/> and remainder part as <target/>
C end

Module B.

Applications Parsed data has been applied to obtain StarDict's Textual file dictionary bases and to create PhpMyAdmin (MySQL administration tool written in PHP) xml import file. But it should be stated that this is only an intermediate stage whence primary goal of the parental research is TEI standard which allows for creating reliable ontologies.

StarDict Dictionary. Stardict has its Textual file dictionary format that uses RELAX NG schema. According to StarDict creator, Hu Zheng, Textual file format was designed to reflect structure of a dictionary [4]. This textual representation of a dictionary has several advantages against earlier created binary opaque format that hides source data and doesn't allow editing dictionaries.

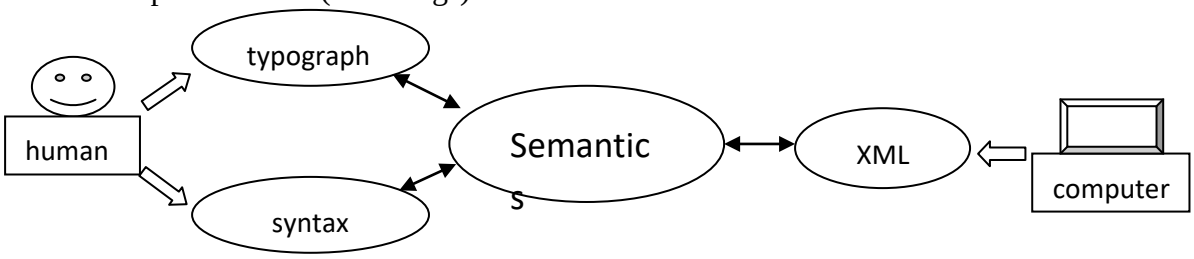
Textual file format may be used to:

1. examine dictionary content
2. make changes to a dictionary
3. create a new dictionary from scratch

It should be said that binary files such as images, audio, video are not stored in Textual file directly, instead their links refer to the resource location. The order of articles (or dictionary entries) in this format doesn't matter.

6. Conclusion

This study of data structuring has shown that semantic structure of a dictionary as of any other abstraction is nothing but different encodings weather by means of typography(layout, formatting, text) and syntax or by means of computer purposed data representation language such as XML (see *Figure 4*). In the first case human beings are able to decode the essential meaning of the data since they are aware of typography-to-semantics interpretation rules. For example in case of Cumakunova's dictionary, when they see bold text they get the knowledge that this text is in Turkish and when they see normal text it must be in Kyrgyz; or when they see out dented line they understand that this is the beginning of a new dictionary entry etc. To enable machines/computers to operate with the same semantics the human encodings must be converted to machine-readable format, i.e. information about any semantic unit must be provided explicitly. Computer must be provided by names, attributes and boundaries of units in order to know their semantics and structure. XML technology provides this meta-information enabling computers to access, process and operate this kind of structured information at the semantic level not on physical. Figure 8. Different representations(encodings) of semantics and



Efficient global information exchange is impossible without commonly understood and shared standards and concepts. Big amounts of information have to be presented as much as possible in unambiguous and precise manner. This goal was pursued in efforts of parsing and annotation of Turkish-Kyrgyz dictionary, since large text corpora from different epochs have to be parsed and annotated for performing further analysis, such as building a meta– lemma list for the project interdependencies between language and genomes [3].

References

1. Dieter Fensel: Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. Springer-Verlag Berlin Heidelberg GmbH, 2004.

2. Dieter Fensel: Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce (Draft version). Springer-Verlag Berlin Heidelberg GmbH, 2004. Page 50. Retrieved from: http://software.ucv.ro/~cbadica/didactic/sbc/documente/silverbullet.pdf, accessed 7/28/2016

3. TEI CONSORTIUM, eds.: TEI P5: Guidelines for Electronic Text Encoding and Interchange. http://www.tei-c.org/Guidelines/P5/

4. Hu Zheng,: Textual Dictionary File Format. StarDict project on GitHub, https://github.com/huzheng001/stardict-3/blob/master/dict/doc/TextualDictionaryFileFormat, accessed 8/6/2016

5. Гүлзура Жумакунова: Түркчө-Кыргызча сөздүк, 50 000 сөз. Кыргыз-Түрк «Манас» Университетинин басылмалары, Бишкек, 2005.

6. Гүлзура Жумакунова: Түркчө-Кыргызча сөздүк, 50 000 сөз. Кыргыз-Түрк «Манас» Университетинин басылмалары, Бишкек, 2005. pages 28-29

УДК 510.5, 519.768.2

МЕТОДОЛОГИЯ АВТОМАТИЗИРОВАННОГО ПОПОЛНЕНИЯ СЛОВАРЯ СИСТЕМЫ МАШИННОГО ПЕРЕВОДА ДЛЯ КАЗАХСКО-РУССКОЙ И КАЗАХСКО-АНГЛИЙСКОЙ ЯЗЫКОВОЙ ПАРЫ

У.А. Тукеев, *Казахский Национальный Университет имени аль Фараби, Алматы, Казахстан.* *ualsher.tukeyev@gmail.com* ***Д.Р. Рахимова,*** *Казахский Национальный Университет имени аль Фараби, Алматы, Казахстан.*di.diva@mail.ru

Ж.М. Жуманов, *Казахский Национальный Университет имени аль Фараби, Алматы, Казахстан.* *z.zhake@gmail.com*

Аннотация: В статье описывается методология автоматизированного пополнения словаря системы машинного перевода Apertium для казахско-русской и казахско-английской языковой пары. Цель данной методологии состоит в оказании помощи пользователю обнаружить лучшую морфологическую парадигму в одноязычном морфологическом словаре Apertium. Приведены практические результаты.

Ключевые слова: заполнение словарей, Apertium, казахский, русский и английский язык.

METHODOLOGY OF THE AUTOMATED ENRICHMENT OF MACHINE TRANSLATION SYSTEM DICTIONARIES FOR KAZAKH-RUSSIAN AND KAZAKHENGLISH LANGUAGE PAIR

U.A. Tukeyev, *Al Farabi Kazakh National University, Almaty, Kazakhstan.* *ualsher.tukeyev@gmail.com*

D.R.Rakhimova, *Al Farabi Kazakh National University, Almaty, Kazakhstan.* *di.diva@mail.ru*

Zh.M. Zhumanov *Al Farabi Kazakh National University, Almaty, Kazakhstan.* *z.zhake@gmail.com*

Abstract: This paper describes the methodology of the automated enrichment dictionary of the machine translation system Apertium for the Kazakh-Russian and Kazakh-English language pair. The purpose of this methodology consists in assistance to the user to find the best morphological paradigm in the monolingual morphological Apertium dictionary. Practical results are presented.

Keywords: filling of dictionaries, Apertium, Kazakh, Russian and English.

Введение

На данный момент существует много различных словарей, как печатные, так и электронные. Словари, используемые в машинном переводе (МП) может содержать переводы на различные языки сотен тысяч слов и фраз, а также предоставить пользователям дополнительные возможности. Такие, как давая пользователю возможность выбрать языки и направление перевода, обеспечить быстрый поиск слов, возможность ввода фразы и т.д.

На сегодняшний день существует множество методов расширения словарей. Мы применили метод, реализованный Микелям Эспла-Гомис(Esplà-Gomis M) и др. [2], и адаптировали для казахского языка. Мы использовали данный инструмент для заполнения словарей казахско-русского и казахско-английского языка в свободной / с открытым исходным кодом системы Apertium машинного перевода.

1.Казахско-русский и казахско-английский словарь

Apertium представляет собой систему машинного перевода поверхностнотрансферного типа. Следовательно, в основном он имеет дело со словарями и правилами поверхностного трансфера. На практике поверхностный трансфер отличается от глубокого тем, что при нём не выполняется полный синтаксический разбор предложений [4]. А правила, в отличии от операций на дереве синтаксического разбора, представляют собой операции с группами лексических единиц. В Apertium используются три типа словарей: два монолингва словарей и один двуязычный словарь для каждой языковой пары. Так мы имеем для казахско-русского и казахско-английского языковых пар следущие:

apertium-rus.rus.dix: морфологический словарь для русского языка: он, в свою очередь, содержит информацию о словоизменении на русском языке.

apertium-eng-kaz.eng.dix: одноязычный словарь английского языка.

apertium-kaz.kaz.lexc: морфологический словарь для казахского языка: он содержит информацию о словоизменении на казахском языке.

apertium-eng-kaz.eng-kaz.dix: двуязычный словарь для англо-казахской пары.

apertium-kaz-rus.kaz-rus.dix: двуязычный словарь, содержит переводные соответствия слов и символов двух языков. В языковой паре любой из языков, составляющих эту пару, может быть как входным, так и выходным языком, т. е. эти термины употребляются условно.

Словари системы Apertium имеют формат XML, каждое слово имеет тег, показывающие ее лексические свойства (часть речи, число, склонение и др.). В контексте системы Apertium парадигма является примером склонения/спряжения определённой группы слов. В морфологическом словаре леммы ссылаются на парадигмы, что позволяет нам показать все словоформы этих лемм без необходимости записи всех возможных окончаний.

Примером использования парадигмы может служить следующее. Допустим, мы хотим добавить в словарь существительные свойство и старшинство. Вместо записи одинаковых окончаний, мы можем записать окончания форм слова свойство, а потом сказать "старшинство изменяется как свойство". Или рассмотрим пример на английском языке, мы хотим добавить в словарь прилагательные happy и lazy. Вместо записи одинаковых окончаний: happy, happ (y, ier, iest) lazy, laz (y, ier, iest)

Мы можем записать окончания форм слова happy, а потом сказать "lazy изменяется как happy", или "shy изменяется как happy", "naughty изменяется как happy", "friendly изменяется как happy" и т. д. В этом примере happy и свойство и будет парадигмой, моделью изменения всех остальных. Точное описание определения парадигм будет дано позже. Парадигмы определяются в тэгах <pardef> и используются в тэгах <par> [3].

Наполнение двуязычных словарей. Таким образом, теперь у нас есть два морфологических словаря и далее мы перейдём к двуязычному словарю. Двуязычный словарь описывает соответствия слов. Все словари имеют один и тот же формат (которой описан в DTD, dix.dtd). В этом словаре парадигмы создаются для каждой части речи

отдельно, например, для глаголов и для прилагательных есть различные парадигмы. В дальнейшем в словаре будут описаны следующие:

```
<!-- numerals -->

<e><p><l>нөл<s n="num"/></l><r>ноль<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>бір<s n="num"/></l><r>один<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>екі<s n="num"/></l><r>два<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>үш<s n="num"/></l><r>три<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>төрт<s n="num"/></l><r>четыре<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>бес<s n="num"/></l><r>пять<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>алты<s n="num"/></l><r>шесть<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>жеті<s n="num"/></l><r>семь<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>сегіз<s n="num"/></l><r>восемь<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>тоғыз<s n="num"/></l><r>девять<s n="num"/></r></p><par n="__num_gender"/></e>
<e><p><l>он<s n="num"/></l><r>десять<s n="num"/></r></p><par n="__num_gender"/></e>
```

Рисунок 1. - Пример заполнения имени числительного для двуязычного казахскорусского словаря.

Как уже видно, заполнение самого словаря требует определенные знания и затрачивается не мало времени. Связи с этим появилось востребованность в автоматизированной системе заполнения словарей.

2. Описание метода автоматизированного пополнения словаря системы машинного

Этот метод состоит из двух этапов:

На первом этапе, обобщенное дерево суффиксов [1], в котором все суффиксы всех парадигм склонения и спряжения в словаре хранятся эффективным способом с комментариями, какие парадигмы генерируют каждый из этих суффиксов. Это дерево позже используется для анализа неизвестного слова и определения способов, которыми оно может быть разбито на пары основа/суффикс.

На втором этапе, метод использует различные стратегии для того, чтобы показать пользователю различные поверхности форм в результате комбинации корней и парадигм, полученных на первом этапе. Затем пользователь может решить, какие из этих словоформ правильны или нет, помогая системе в конечном счете найти наиболее подходящую пару основа/парадигма для слова, которая вставляется в словарь. Были опробованы две стратегии определения слов, которые спрашиваются у пользователя, с целью как можно большего уменьшения взаимодействий с ним: один - эвристический, второй - основанный на использовании деревьев решений ID3 [2].

В этом разделе мы описываем эксперименты, проводимые, чтобы попытаться адаптировать методику, ограничения наших подходов и возможные шаги улучшения этого метода для языков.

2.1 Описание проблемы

Эта методология, описанная в [2], основана на ряде предположений, которые определяют возможность использования метода. Эти предположения:

Изменения слов в языке происходит с использованием суффиксов, т.е. способ, которым язык производит формы слов, заключается в добавлении суффиксов к статической основе (это условие выполняется для казахского и русского языков); все суффиксы, генерируемые парадигмами в словаре могут быть получены заранее; две парадигмы не должны генерировать один и тот же набор суффиксов.

Предыдущие работы [2] показывают, что третье условие не всегда выполняется (Например, парадигмы для топонимов и антропонимов, как правило, создают те же самые суффиксы (в случае английского языка, это был бы пустой суффикс)), и в этих случаях метод полезен только, чтобы уменьшить количество возможных основ/парадигм. Оставив этот факт в стороне, эта методология была успешно использована для каталонского, испанского,

английского, баскского, мальтийского и хорватского языков. Обширные эксперименты, проведенные на хорватском языке, позволяют хорошо оценить возможность использования метода для русского языка, учитывая, что оба они - славянские языки с очень похожим системами склонения и спряжения. Метод был протестирован со словарем русского языка в Apertium (apertium-rus.rus.dix) и неизвестными русскими словами. Он работает, как ожидалось, но обширные эксперименты не проводились. В следующем разделе содержится набор использованных неизвестных русских слов и команды, использованные для тестирования. Однако, казахский язык является гораздо более сложной проблемой. Основные трудности адаптации метода для казахском языка связаны с:

форматом словаря: в то время как английский и русский языки используют, основанный на XML формат dix, изначально используемый Apertium для хранения одноязычных словарей, казахский язык основан на формализме HFST для конечных автоматов.

двухуровневыми морфологическими правилами: важной проблемой, с которой сталкиваются при работе с системой казахского языка на Apertium, является то, что он использует двухуровневые морфологические правила, которые изменяют результат конкатенации основ и суффиксов. Это означает, что поверхностная форма не производится непосредственно словарем и, таким образом, невозможно извлечь прямым путем основы и суффиксы из формы. Это делает невозможным использование метода, описанного в [2] напрямую.

Таким образом, должно быть определено лучшее адаптированное решение для того, чтобы приспособить данный метод для казахского языка, заменив первый шаг процесса (обнаружение пар кандидатов основа/парадигма с помощью обобщенного дерева суффикса) более специфичным решением.

2.2 Подход решения проблемы адаптации к казахскому языку

Была изучена возможность параллельного подхода, который многократно использует как можно больше идей, воплощенных в методологии, используемой для русского языка.

Обратите внимание, что, как и в других тюркских языках, когда основа, такая как «мектеп» (школа), получает дополнительные морфемы в свои суффиксы, например, для генерации формы в 3-ем лице притяжательной формы и в творительном падеже, последние буквы основы иногда изменяются, в результате добавления суффикса, но суффикс также может изменяться, в зависимости от последних букв основы: «мектебінен» (с его школы), «мектеп» изменился на «мектеб», но окончание «інен» отличается от суффикса («сыдан»), который относится к «қалта» (карман): қалтасынан (с его кармана). Мы подсчитали, что, в худшем случае, могут измениться три буквы основы.

В этих условиях, чтобы быть в состоянии помочь пользователю в добавлении неизвестных слов в словари, мы разработали пакет скриптов, которые позволяют создавать и компилировать два модифицированных словаря из существующих apertium–kaz.kaz.lexc и одного вспомогательного файла.

Был разработан новый класс в нашем первоначальном пакете java, который читает выходные данные этого процесса и использует тот же интерфейс, чтобы спросить пользователя о правильности поверхностной формы слова и вставляет полученных кандидатов в словарь. Стоит отметить, что первые два действия не нужно повторять до тех пор, пока определение парадигм в словаре не изменится каким-либо образом. Поэтому, все ресурсы, необходимые для применения этого метода, могут быть получены только один раз и затем применены на лету к новым словам. Вывод всего этого процесса затем может быть подан в DictionaryAnalyser, написанный на Java, который задает вопрос.

3.Практические результаты

В этом конечном результате мы проверили возможность использования нашего подхода для помощи неспециалистам в задаче добавления новых слов в морфологические словари казахского и русского языка в Apertium. Мы сделали предварительные испытания,

чтобы подтвердить, что наш оригинальный подход подходит для русского языка без изменений, но не подходит для казахского языка, который использует совершенно другой формат и стратегию для построения морфологических одноязычных словарей. Мы разработали коллекцию скриптов, которые позволяет нам адаптировать оригинальные идеи нашего метода для казахского языка. С помощью этих сценариев, можно создать файл, содержащий все возможные основы/парадигмы и коллекцию поверхностных форм из списка неизвестных слов, которые затем предлагаются пользователю. В заключении, мы адаптировали оригинальный интерфейс для того, чтобы адаптировать его для чтения этого файла, что позволяет использовать метод двоичных вопросов для вставки новых слов в словарь казахского языка.

После запуска инструмента пользователь может выбрать один из различной комбинации кандидатских стеблями и парадигм, отвечая на вопросы, заданные системой. Когда пользователь подтверждает, что слова были обнаружены правильно, они перемещаются в соответствующий словарь раздел. В случае, когда система обнаруживает более одного решения для слова, все возможные варианты записываются в словарь наряду с количеством найденных возможных вариантов.

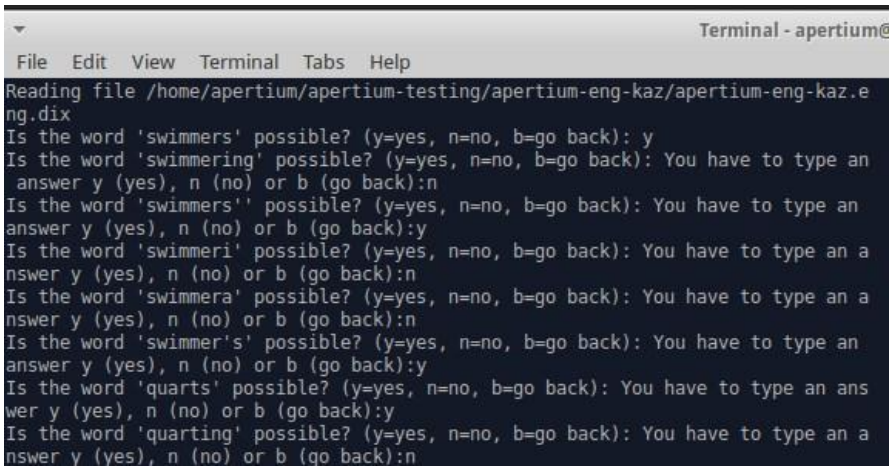


Рисунок 2.- Пример применения метода в заполнении словаря.

```
  </section>
  <section id="Guessed" type="Temporal">
<e r="" lm="swimmer" a="QueringUser" c="1 possible solutions; choose one and confirm it"><i>swimmer</i><par n="staff__n"/></e>
<e r="" lm="quart" a="QueringUser" c="2 possible solutions; choose one and confirm it"><i>quart</i><par n="house__n"/></e>
<e r="" lm="quart" a="QueringUser" c="2 possible solutions; choose one and confirm it"><i>quart</i><par n="Smith__np"/></e>
<e r="" lm="geographer" a="QueringUser" c="1 possible solutions; choose one and confirm it"><i>geographer</i><par n="staff__n"/></e>
<e r="" lm="purse" a="QueringUser" c="1 possible solutions; choose one and confirm it"><i>purse</i><par n="staff__n"/></e>
  </section>
```

Рисунок 3.- Генерация словарных статей

Заключение

Данный метод был впервые применен для сложных языковых пар, как казахскорусский и казахско-английский. Система была адаптирована для казахского языка: был создан словарь apertium-kaz-kaz-suffixgen.lexc, в котором поддельная, зависимая от парадигмы, основа был связана с каждой парадигмой открытого класса и разработано 113 лексических форм. Мы используем эту методологию, чтобы расширить количество слов в словарях, которые в основном эффекты к качеству машинного перевода. Технология позволяет пользователям, которые не обладают глубокими знаниями в области вычислительной представления морфологии, но понимают язык и активно участвуют в построении словарей. Это означает, что все больше людей могут добавлять словарные статьи, создающие большие словари в более короткие сроки.

Список литературы

1. McCreight E.M. A Space-Economical Suffix Tree Construction Algorithm. //Journal of the ACM. – 1976. - 23 (2). – P.262–272.

2. Esplà-Gomis M., Sánchez-Cartagena V.M., Pérez-Ortiz J.A., Sánchez-Martínez F., Forcada M.L., Carrasco R.C. An efficient method to assist non-expert users in extending dictionaries by assigning stems and inflectional paradigms to unknown words. //Proceedings of EAMT 2014, The Seventeenth Annual Conference of the European Association for Machine Translation (Dubrovnik, June 16–18, 2014). – 2014. – P.19–26.

3. Abduali B., Sundetova A., Zhanbussunov N., Musabekova Zh., Study of the problem of creating structural transfer rules for the Kazakh-English and Kazakh-Russian machine translation systems on Apertium platform. //Вестник Казну им. Аль-Фараби, том 20 (Messenger of AlFarabi KazNU, vol. 20) by proceedings of the International Conference "Computational and Informational Technologies in Science, Engineering and Education" (CiTech-2015, 24-27 September, 2015). - Almaty: Al-Farabi Kazakh National University Press, 2015. - P.77-82

4. Рахимова Д.Р., Қалдашбеков Е.Е., Мұсабекова Ж.Ғ., Абақан М., Кызырканова С., Жамалиева А., Абдуали Б. Apertium платформасы негізінде орыс тілінен қазақ тіліне аудару жүйесін құру. //Материалы международной научной конференции студентов и молодых ученых "Фараби Әлемі», Алматы, Казахстан, 13-16 апреля, 2015 г. - Алматы, 2015. - С.175.

УДК 621.3

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ИМЕНИ СУЩЕСТВИТЕЛЬНОГО ДЛЯ СИСТЕМЫ КАЗАХСКО- КИТАЙСКОГО МАШИННОГО ПЕРЕВОДА

Унзила Каманур *Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан Unzila.88@mail.ru*

Алтынбек Шарипбай *Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан*

Гүлмира Бекманова *Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан*

Лена Жеткенбай *Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан*

Цель статьи - В этой работе будет рассмотрен создание онтологической модели имен существительных казахского, русского языков, предназначенных для системы машинного перевода. Эта модель даст нам возможность сравнить сходства и различие двух языков. Также можно применить для разработки информационного поиска, машинного перевода и автореферирования, диалоговых и других систем.

Ключевые слова: агглютинативные языки, морфология, онтологии.

ONTOLOGICAL MODEL OF NOUNS SYSTEM FOR KAZAKH-CHINESE MACHINE TRANSLATION

Unzila Kamanur *L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan Unzila.88@mail.ru*

Altynbek Sharipbay *L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan*

Gulmira Bekmanova *L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan*

Lena Zhetkenbay *L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan*

Abstract. In this work, the machine translation for jüyecine Kazakh, Chinese nouns establishment of the ontological model. This model allows us to compare the similarities and differences between the two languages. Information retrieval, machine translation and avtoreferattaw, and other systems can be used to create a dialogue.

Keywords: agglutinative languages, morphology, ontology.

1 . Введение

В настоящей работе строятся онтологические модели имен существительных казахского и китайского языков для создания системы машинного перевода между ними.

Казхский язык входит в кыпчакскую группу [1], а Китайская язык - ветвь синотибетской языковой семьи.

Китайский язык очень древний язык, который считается одним из самых древних языков. Кроме того, язык постоянно развивается и является одним из совершенствующих себя языков. Китайский язык входит в ряд хрупких языков. В качестве любых грамматических показателей используются символы иероглифы. В различных частях речи отдельные слова тоже пишутся иероглифами. Важным фактором непрерывного развития этого языка является использование иероглифов в качестве языкового инструмента. Китайский язык является одним из 6 рабочих и официальных языков Организации Объединенных наций. Если взглянуть с исторической стороны, то этот язык является языком национальности хань, который входит в состав национальностей КНР.

Из-за того что китайский язык является непроизводной формой, при формировании слов в структуре не происходит большие изменения. Казахский язык богат окончаниями и суффиксами, а в китайском языке более 70 % слов составляет двуслоговые слова. Также в китайском языке один иероглиф является одним слогом, вследствие этого большинство морфем будет одинарным. Например: 山，海，江，我，人. Вместе с этим они бывают двуслоговые и многослоговые, например: 蜘蛛，蝴蝶，仿佛，巧克力, 阿司匹林. Каждый слог многослоговых слов не дает никакого значения, только каждый слог соединяясь с другим дает какой-нибудь смысл, при произношении не имеет значений. Например: 巧 иероглиф имеет единственное значение, а 巧克力 не имеет никакого смысла, и поэтому не морфема. Соединение три иероглифов дает одну морфему. Слова, состоящие из двух или нескольких морфем называются слитными словами. Например: 英雄, 劳动，哈萨克, 文化局.

Слова на китайском языке во многих случаях по внешнему виду не отражает морфологических признаков. Только в составе предложения можно точно определить, какой части речи относиться слова. По типу связей двух слов делается морфологический анализ. Эта отличительная особенность предложений написанных этими иероглифами и показывают гибкость использования иероглифов. [5-8]

В настоящее время существуют различные виды систем машинного перевода. Применение тех или иных систем машинного перевода может быть продиктовано сложностью формализации того или иного естественного языка и наличием национальных языковых корпусов естественного языка.

Проблеме машинного перевода с китайского языка на другие тюркские языки посвящены [9-15]. Проблемы формализации грамматических правил казахского языка решены в работах . Эти результаты существенно облегчают создание системы казахско-китайского машинного перевода.

2 Онтологическая модель имени существительного для системы машинного перевода

В настоящее время онтология является мощным и распространенным инструментом моделирования отношений между объектами различных предметных областей. Принято классифицировать онтологии по степени зависимости от задач или прикладной области, по

модели представления онтологических знаний и его выразительным возможностям и другим параметрам [Gruber,1993][Gruber1995].

Основным смыслом использования онтологии в области технической науки – показать концептуальным чертежом обобщенные и отдельные формализации охватывающие все множества данных по одной определенной образовательной сфере. В основе концептуального чертежа дается множества понятий и данные о понятий (особенность, связь, ограничения, аксиомы и закрепление понятий, вся эта информация необходима для описания процесса решения задачи по выбранной предметной области)

Почему возникает необходимость в создании онтологии? Обычно онтология используется в зависимости от следующих причин:

- Для общего понимания структуры информации людьми и программными агентами;
- Для возможности повторного использования знаний в предметной области;
- Для четких предположений в предметной области;
- Для разделения знаний предметной области от оперативных знаний;
- Для проведения анализа знаний в предметной области;

Использование общего понимания структуры информации людьми и программными агентами вместе являются одной из общих целей создания онтологии.

Для создания онтологии необходимо определить его предметную область и масштабы. Итак ответим на несколько основных вопросов:

1. Какую область будет охватывать онтология? Ответ: Имена существительные.

2. Для чего нам необходима онтология? Ответ: нужна для создания сравнительной онтологической модели по именам существительным казахского и китайского языков.

3. На какие виды вопросов должна отвечать информация в онтологии? Ответ: для определения спряжения и видов окончания имени существительного по содержанию и значению.

4. Кто будет использовать и поддерживать онтологию? Ответ: Лингвисты и программисты.

В соответствии с этими ответами сравнительная онтологическая модель имени существительного будет такой: O(X, R, I), X – это наименования, входящие в структуру имени существительного, R – связи между наименованиями, а I – множества наименований структур и отношений.

Для создания онтологической модели используется языки описания онтологии и редакторы. Разные онтологические языки дает возможность использовать разные возможности. Сравнительная онтологическая модель имени существительного была создана одной из последних обработок стандартных языков описывающих онтологию - Protege (http://protege.stanford.edu). Protege OWL дает возможность описывать не только понятия, но и также конкретные объекты. Наименования и знаки использованные в сравнительной онтологической модели имени существительного казахского и китайского языков описаны в 1-таблице

Таблица 1 – имена и знаки использованные в сравнительной онтологической модели имен существительных из казахского и китайского языков

N	**На казахском**		**На китайском**	**English**	**UNIFIED**
	Существительные			Noun	
1.	**По составу**			**Structure**	
1.1		Единственное		Simple	

1.2		Сложное		Complex	
1.2.1		Объединенное		Compound words	
1.2.2		Парное слово		Reduplicates	
1.2.3		Словосочетание		Compound words	
1.2.4		Сокращенное слово		Abbreviations	
2.	**По форме**				
2.1		Основное		Derıvations	DERIVNS
2.2		Производное		Derivatives	DERIV

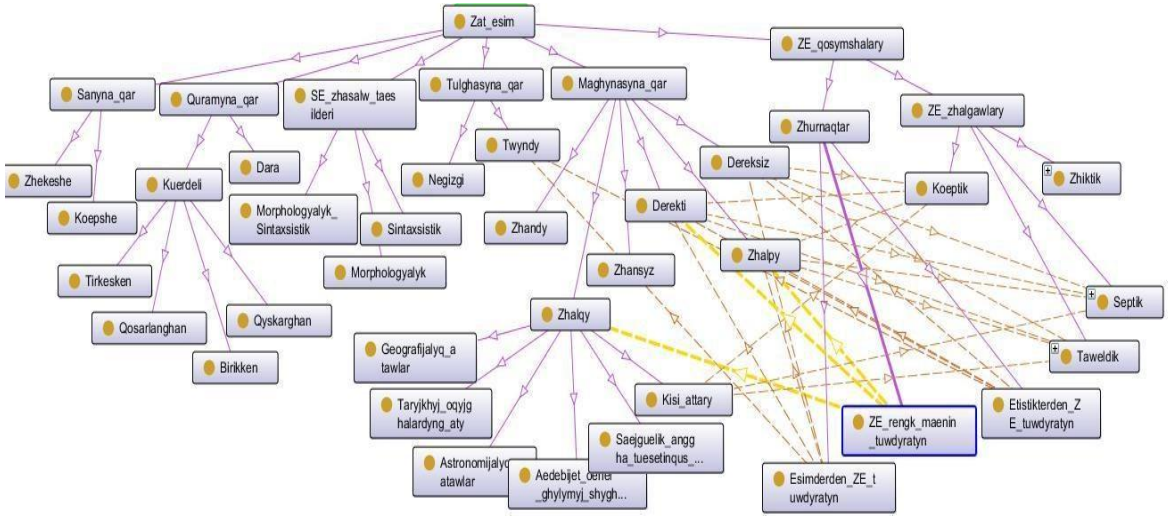
2.3					
3.	**По значению**		**Виды по значению** 名词的种类	**According to the meaning**	
3.1		Одушевленное	表示人的 (Biǎoshì rén de) имена существительные означающие имена людей	Animate	ANIM
3.2		Неодушевленное	表示事物的 (Biǎoshì shìwù de)： Имена существительные означающие предметы и явления	Inanimate	INANIM
3.3		Общее	表示时间： (Biǎoshì shíjiān) Имена существительные означающие период	Common	COM
3.4		Собственное	表示方位的： (Biǎoshì fāngwèi de) Имена существительные означающие место и расположение	Proper	PROP

4.			表示处所的 (Biǎoshì chùsuǒ de)：Имена существительные означающие место		
		Конкретное		Concrete	CON
		Абстрактное		Abstract	ABS
5.	По количеству				
5.1		Единственное		Singular	SG
5.2		Множественное		Plural	PL
5.3				Collectıve	
6.	**По окончанию**				

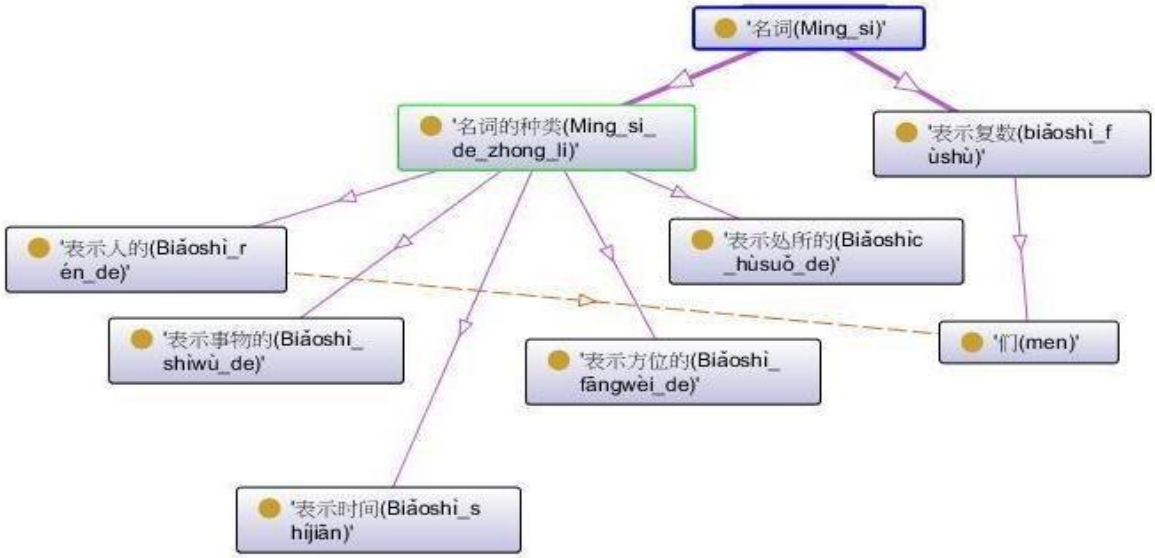
6.1	*По падежному окончанию*			*Caseending*	*Cases*
6.1.1		Именительное		Nominativecase	NOM
6.1.2		Родительный		Genitivecase	
6.1.3		Дательный		Direction-dativecase	DAT
6.1.4		Винительный		Accusativecase	ACC
6.1.5		Местный Жатыс		Locativecase	LOC
6.1.6		Исходный Шығыс		Ablativecase	ABL
6.1.7		Творительный		Instrumentalcase	
6.2	*По множественному числу*		表示复数 biǎoshì fùshù 表示 ： имен суще ль ным ающим а люд й единя ание men получестве н ачени		Plural

6.2.1				Singular	SG
6.2.2				Plural	PL
6.3	*По личному окончанию*			Personalending	Pers_end
		1 лицо единственное		1 personalsingular	PERS.1SG
		2 лицо единственное		2 personalsingular	PERS.2SG
		3 лицо единственное		3 personalsingular	PERS.3SG
		2 лицо единственное вежливое		2 personalsingularformal	
		1 лицо множественное		1 personalplural	PERS.1PL
		2 лицо множественное		2 personalplural	PERS.2PL
		3 лицо множественное		3 personalplural	PERS.3PL
		2 лицо множественное вежливое		2 personalpluralformal	
6.4	*По притяжательной форме*			Possesiveending	Poss_end
		1 лицо		1 Possesivesingular	POSS.1SG
		единственное			
		2 лицо единственное		2 Possesivesingular	POSS.2SG
		3 лицо единственное		3 Possesivesingular	POSS.3SG
		2 лицо единственное вежливое		2 Possesivesingularformal	
		1 лицо множественное		1 Possesiveplural	POSS.1PL
		2 лицо множественное		2 Possesiveplural	POSS.2PL
		3 лицо множественное		3 Possesiveplural	POSS.3PL
		2 лицо множественное вежливое		2 Possesivepluralformal	

Созданная сравнительная онтологическая модель имени существительного казахского и китайского языков в среде Protege с помощью этих значений показан в 1-рисунке



1-рисунок. Онтологическая модель имени существительного казахского языка



2-рисунок. Онтологическая модель имени существительного китайского языка

Таким образом, сравнительная онтологическая модель имени существительного для системы машинного перевода охватывает все компоненты и множества морфологического анализа, а именно в казахском языке по составу имени существительного делятся на единственное и множественное, по значению – общее, собственное, конкретное, абстрактное, одушевленное и неодушевленное, а в китайском – не предусмотрено такое разделение, по значению рассмотрены 表示人的: имена стуществительные означающие имена людей, 表示事物的: имена существительные означающие предметы и явления, 表示方位的: имя существительное означающее место и направление, 表示时间：имена существительные означающие период, 表示处所的: имена существительные означающие место. В казахском языке слова означающие период и расположение рассматривается как наречие, а в китайском языке как имя существительное. В китайском языке имена существительные не спрягаются, только одно множественное окончание «men», а также его круг пользования узок. Его специфичность в том, что этот предмет делится на группы по виду, форме, внешнему виду, качественному свойству. Например: предмет широкий, длинный, в комочек и т.д. Вот такие имена предметов разделенные на группы обязательно используется с именем числительным сочетающимся с местоимением, а между ними ставятся слова означающие величину.

Предметы и явления окружающей среды воспринимаютя и узнаются по разному. В китайском языке задается вопрос кто? именам существительным человеку и его деятельности, профессий и работе. В этом языке очень много слов определяющее

наименование предметов, отвечающее на вопросы кто? что? В общем эти все являются наименованиями предметов, явлений, профессий, и каждый предмет имеет смысловой, объяснительный, понятийный и другие свойства. Обычно “不，很” имена существительные не согласуются напрямую с наречиями, а также с именем числительным, между ними без слов понадобится слово дающее значение величины. Имена существительные в предложений часто бывают подлежащим и дополнением. [4]

3 Выводы: Созданная онтологическая модель для компьютерной обработки казахского и русского языков является важнейшим шагом при сравнительном исследовании двух языков. И поэтому исследование структуры и значений имен существительных казахского и китайского языков и их результаты сравнений безусловно влияет на систему машинного перевода и создание системы обработки естественного языка.

Список литературы

1. Казахская грамматика. Фонетика, словообразование, морфология, синтаксис. – Астана,2002.

2. Нұрсаидов М.А. «Қытай тілінің өзіндік ерекшеліктеріне қысқаша сипаттама» Алматы, 2002 ж.

3. 卢福波《对外汉语教学实用语法》北京语言文化大学出版社 2002 年

4. Дүкен Мәсімханұлы “Қытай филологиясына кіріспе” (оқулық, 2004 ж., Астана).

5. Бекманова Г.Т., Махимов А.К. Графематический анализ казахского неструктурированного текста // Информатизация общества: труды 3-ой международной научно-практической конференции.– Астана, 2012.– С. 509-511

6. Бекманова Г.Т., Шарипбаев А. А. Формализация морфологических правил казахского языка с помощью семантической нейронной сети // Доклады Национальной академии наук Республики Казахстан. – 2009.– №4. – С. 11-16

7. Bekmanova G., Sharipbaev A. The synthesis of word forms of Turkic language using semantic neural networks // Modern problems of applied mathematics and information technologies: abstracts – Al Khorezmy, 2009.–P.145.

8. G.T.Bekmanova, A.A.Sharipbaeyv, S.K.Buribayeva Formalization of morphological rules of inflection in the Kazakh language // Вестник. Астана: Евразийский национальный университет им. Л.Н.Гумилева, 2012. – Специальный выпуск.– С.18-26

9. Methods and models of Методы и алгоритмы распознавания слов казахского языка: монография // Астана: ЕНУ им. Л.Н.Гумилева, 2010. 132с.

10. A.Sharipbayev, G.Bekmanova, B.Yergesh, A.Buribayeva, and M.K.Karabalayeva. Intellectual morphological analyzer based on semantic networks. Processings of the OSTIS-2012, 2012, pp.397-400.

УДК 004.5:811.512.145

РЕЧЕВОЙ ЧЕЛОВЕКО-МАШИННЫЙ ИНТЕРФЕЙС НА ТАТАРСКОМ ЯЗЫКЕ

А.Ф. Хусаинов, *Институт прикладной семиотики Академии наук Республики Татарстан*
Казанский (Приволжский) федеральный университет, Казань, Россия
khusainov.aidar@gmail.com

Аннотация

В работе приводятся результаты исследований по созданию речевого человекомашинного интерфейса на татарском языке. Он включает в себя два основных

элемента: систему автоматического распознавания речи и систему синтеза речи. Планируется использовать возможности речевого ввода и вывода информации на татарском языке при создании множества диалоговых систем: машинного переводчика, интеллектуального помощника и т.д.

Ключевые слова: распознавание речи, синтез речи, речевой интерфейс, татарский язык

SPEECH HUMAN-MACHINE INTERFACE FOR THE TATAR LANGUAGE

A.F. Khusainov, *Institute of Applied Semiotics of the Tatarstan Academy of Sciences Kazan (Volga region) federal university, Kazan, Russia,* khusainov.aidar@gmail.com

In this paper, we describe our recent work of creation speech human-machine interface for the Tatar language. Our work consists of two main elements: speech recognition system and speech synthesizer. These systems will be used in mobile and desktop applications, for instance, machine translation system, smart assistant.

Key words: speech recognition, speech synthesis, speech interface, the Tatar language

1 Введение

Увеличение числа информационных систем и условий их работы диктует развитие новых средств взаимодействия с ними. Одним из актуальных способов взаимодействия является использование произнесённых команд и сообщений на естественном языке. Речевой интерфейс получает всё большее распространение благодаря нескольким факторам: естественности использования речи для человека, удобству использования в определенных условиях, а также развитию технологий анализа, синтеза речи и понимания текста.

Данная работа описывает последние результаты, полученные при создании систем автоматического распознавания и синтеза речи в контексте использования татарского языка. Для решения обеих задач используется корпусный подход: система распознавания обучается на базе создаваемого многодикторного корпуса слитной речи, синтезатор речи – на основе студийных записей профессиональных дикторов.

2 Система распознавания слитной татарской речи

Система распознавания речи состоит из 4 основных элементов:

Акустические модели: модели произношения акустических единиц языка (фонем, дифонов и т.д.);

Модели фонетических транскрипций слов (лексический уровень): словарь используемых слов, фонетические транскрипции слов;

Уровень языковой модели: описываются правила употребления слов языка;

Декодер – производит анализ входного речевого сигнала на основе моделей. **2.1.**

Акустические модели

Акустические модели создаются для условий относительно качественных записей: 16 бит в секунду, 16 кГц. Их можно будет использовать для распознавания речи, например, в офисных помещениях, перед компьютером, для анализа выступлений в не очень шумных помещениях. В дальнейшем планируется создавать отдельные корпусы для речи, передаваемой через телефонный канал связи, а также эфирной речи (телевидение и радио).

Крупнейшие разработчики систем распознавания речи используют для обучения речевые корпусы общей продолжительностью в тысячи часов речи. Однако экспериментально установлено, что результаты, устойчивые к смене пола, возраста и особенностей произношения диктора, можно достичь, начиная с 50 часов записей.

Общие характеристики речевого корпуса для распознавания речи, созданные в рамках исследования, приведены в Табл. 1. Все записи речевого корпуса имеют метаописания: сохраняется информация о дикторах (пол, возраст, родной язык, наличие диалекта), условиях записи (звуковое оборудование, шумовые условия).

Таблица 1. Характеристики татарского речевого корпуса

Параметр	Значение
Количество дикторов корпуса читаемой речи	341
Общая продолжительность корпуса читаемой речи	44:59:51
Средняя продолжительность записи	7:55
Продолжительность спонтанной речи	5:19:33
Количество дикторов корпуса спонтанной речи	168

2.2. Модель фонетических транскрипций слов

На основе созданных акустических моделей можно построить систему распознавания фонем татарского языка. Для перехода с уровня отдельных фонем на уровень слов необходимо создание алгоритма построения фонетических транскрипций слов. Данный алгоритм работает на основе правил графем-фонемных преобразований.

Алфавит татарского языка состоит из 39 символов: к буквам русского алфавита в нём добавлены татарские буквы Ә-ә, Ө-ө, Ү-ү, Җ-ж, Ң-ң, Һ-һ. Вопрос о составе фонемного алфавита татарского языка не имеет однозначного ответа, поэтому при его формировании учитывались как имеющиеся на данный момент результаты фонетических исследований татарского языка, так и особенности работы программных средств распознавания речи, заключающиеся в необходимости выделения основных звуков языка, которые влияют на смысл произнесенного, группируя при этом схожие по звучанию звуки. В результате проведенного исследования был сформирован алфавит из 57 фонем татарского языка. На основе определенного алфавита фонем были сформулированы акустические закономерности в татарском языке. В конечном итоге для случая татарского языка было выделено 37 правил фонетической транскрипции [1].

2.3. Языковая модель татарского языка

Задача создания языковых моделей возникает при решении множества задач, от проверки орфографии и до систем машинного перевода. Во всех случаях языковая модель призвана описывать существующие в языке закономерности и на их основе уметь оценивать вероятности произнесения конкретных последовательностей слов.

Татарский язык относится к группе агглютинативных языков и имеет богатую морфологию. При построении стандартных статистических языковых моделей для таких языков возникает проблема с большим числом словоформ, которые необходимо включать в словарь. Большое количество различных аффиксальных цепочек, которые могут следовать за основой слова, делает невозможным построение словаря адекватных размеров с небольшим уровнем OOV (out of vocabulary) слов. Решением этих проблем является уменьшение базовой моделируемой единицы до элемента, меньшего чем слово. В качестве базовых подходов в текущем исследовании были выбраны следующие: слова, морфемы, основы плюс аффиксальная цепочки, статистически выделенные морфы, слоги, буквы.

Исходной информацией для обучения языковой модели выступил текстовый корпус татарского языка [2]. Полученный для работы фрагмент после процедуры фильтрации и разделения на обучающую и тестовую части имеет следующие характеристики, Табл. 2.

Таблица 2. Характеристики текстового корпуса

Параметр	Значение
Количество файлов	217 294
Количество слов	69 810 033
Количество слогов	186 014 478 (2,66 / слово)
Количество морфем	110 280 448 (1,58 / слово)
Количество морфов	93 458 542 (1,34 / слово)
Количество основ и афф. цепочек	97 461 218 (1,4 / слово)
Количество букв	434 636 548 (6,23 / слово)
Размер	901 МБ

С учётом отсутствия до настоящего момента публикаций на тему построения и сравнения различных видов статистических языковых моделей для татарского языка, схема эксперимента была составлена таким образом, чтобы собрать максимально полную оценку влияния факторов на качество итоговой языковой модели. Так, были построены и проанализированы отдельные статистические модели для всех комбинаций из следующий категорий:

1. Тип элемента – 6 типов: слово, слог, морфема, морф, основа и аффиксальная цепочка, буква;

1. Размерность n-грамм: биграммы, триграммы, 4-граммы (5-граммы для модели на основе букв);

2. Алгоритм сглаживания модели – 5 типов: абсолютное сглаживание, GoodTuring, Kneser-Ney, Witten-Bell, модифицированный алгоритм Kneser-Ney.

Качество построенной модели оценивалось по таким показателям, как логарифм вероятности для тестового подкорпуса, perplexity (степень уверенности модели при анализе тестовых данных), OOV (количество элементов тестовой выборки, не вошедших в словарь) и размеру модели (по числу используемых n-грамм).

По результатам построения моделей был сделан вывод о том, что с точки зрения алгоритма сглаживания наилучшие результаты показали основной и модифицированный алгоритмы Kneser-Ney. Среди 95 построенных моделей наилучшее качество показала пословная модель, далее следуют модели на основе морфем и основ с аффиксальной цепочкой, морфов, слогов и букв, Табл. 3.

Таблица 3. Сравнение языковых моделей

Базовый элемент	Log вероятности, тыс.
Слово (4-грамм)	-12 209,0
Основа+цепочка (4-грамм)	-12 386,7
Морфема (4-грамм)	-12 638,7
Морф (4-грамм)	-12 772,4
Слог (4-грамм)	-14 282
Буква (5-грамм)	-20 741,5

В заключительном эксперименте была построена биграммная модель на основе классов слов для словаря в 20 тысяч элементов. Для выделения классов слов использовался алгоритм Брауна [3]. Построенная модель отличается нулевым значением внесловарных слов, небольшим размером, однако уступает стандартным пословным моделям в качестве

описания тестового подкорпуса. Интерес представляет результат разбиения словаря из 20 тысяч слов на классы: автоматически выделенные классы объединяют слова со схожими значениями. Например, в отдельные классы выделены названия населённых пунктов, чисел, годов, фамилий, названий стран, профессий и т.д.

Для использования при распознавании была использована пословная 3-граммная модель со словарём в 100 тысяч наиболее частотных слов татарского языка.

2.4. Распознавание слитной татарской речи

В качестве декодера был использован инструмент Julius [4]. Дикторонезависимая система распознавания слитной татарской речи была построена на основе пословных моделей. Система реализована в двух вариантах: в виде консольного приложения, предоставляющего также и служебную информацию, и в виде оконного приложения, демонстрирующего распознанный текст. Для удобства пользователей был реализован алгоритм автоматического определения границ фраз.

Разработанная система распознавания речи будет адаптирована для использования в конкретных приложениях (например, словарях, машинном переводчике, интеллектуальном помощнике), в течение 2016 года станет доступна для использования на сайте [5].

3 Система автоматического синтеза татарской речи

Задача синтеза речи состоит в формировании аудиосигнала на основе фразы, представленной в текстовом виде. Большая часть подходов к синтезу речи основывается на конкатенативном подходе (дифонный синтез [6], Unit selection [7]). Исходной информацией в данных подходах служат выделенные из речевых записей акустические единицы. Начиная с 2002 года, набирает популярность параметрический подход, в котором базовыми элементами являются статистические модели звуков языка.

При разработке синтезатора татарской речи используется параметрический подход на основе скрытых Марковских моделей (HMM-based speech synthesis, HTS [8]). Для обучения скрытых Марковских моделей необходимо наличие аннотированного речевого корпуса.

Особенностью создания корпуса для синтеза речи являются требования к качеству используемой аппаратуры и условиям записи. Для создании корпуса татарского языка был задействован профессиональный диктор, записи были сделаны в звукозаписывающей студии, в звуконепроницаемом помещении с использованием профессионального оборудования.

В записанных аудиофайлах экспертами были вручную размечены все интонационные группы, отмечены заимствованные и акцентные слова, после чего для всего прочитанного текста была построена фонетическая транскрипция.

Далее, была реализована модель разметки корпуса, основные элементы которой можно представить следующим образом:

1. Уровень фонем: текущая фонема, две предшествующие, две последующие фонемы.
2. Уровень слогов: тип слога (V, VC, CV, CVC, VCC, CVCC); позиция фонемы в слоге; количество фонем в предыдущем, текущем, последующем слоге; номер текущего слога в слове; гласная в текущем слоге.
3. Уровень слов: часть речи, количество слогов для предыдущего, текущего, следующего слова; количество предшествующих и последующих слов во фразе.
4. Уровень фразы: количество слов/слогов в предыдущей, текущей, последующей фразе.

4. Заключение

Разработанные системы автоматического распознавания слитной речи и её синтеза позволяют начать работы по внедрению речевого человеко-машинного интерфейса на татарском языке.

Дальнейшее развитие речевых технологий предусматривает совместное использование результатов исследований в области семантического анализа текста на

татарском языке, что позволит создавать интеллектуальные системы. Планируется разработка мобильных приложений, предоставляющих возможности машинного перевода, работы со словарями, помощи слабовидящим, диктовки и т.д.

Список литературы

1. Хусаинов, А.Ф. Система автоматического распознавания фонем татарского языка // Компьютерная обработка тюркских языков: труды первой международной конференции. (Астана, 3-4 октября, 2013). Астана: ЕНУ им. Л. Н. Гумилева, 2013. С. 211–217.

2. Dz. Suleymanov, O.A. Nevzorova, and B. Khakimov. National Corpus of the Tatar Language “Tugan Tel”: Structure and Features of Grammatical Annotation. Proc. International Conference Georgian Language and modern Technology. (Tbilisi, 2013). P. 107-108.

3. P. F. Brown, V. J. Della Pietra, P. V. deSouza, J. C. Lai and R. L. Mercer. Class-Based n-gram Models of Natural Language, Computational Linguistics 18(4), 1992. P. 467-479.

4. Open-Source Large Vocabulary Continuous Speech Recognition Engine [Электронный ресурс]. URL: https://github.com/julius-speech/julius.

5. Программные продукты, локализованные на татарский язык [Электронный ресурс]. URL: http://tatsoft.tatar.

6. Moulines, E., Charpentier, F. “Pitch-sinchronous waveform processing techniques for text-to-speech synthesis using diphones”. In Speech Communication, 9 (5/6), 1990, pp. 453–467.

7. Sagisaka, Y. ATR v-talk speech synthesis system. In Proc. ICSLP-92, 1992, Banff, Canada.

8. Yoshimura, T., Tokuda, K.. Masuko, T., Kobayashi, T., Kitamura, T. “Simultaneous modeling of spectrum, pitch and duration in HMM-basedspeech synthesis”. In Proc. Eurospeech, 1999, pp. 2347–2350.

УДК 161.112:811.512.1-027.271

SIMILARITIES AND DIFFERENCES OF TURKIC LANGUAGES

Adalı, Eşref, *PhD (Engineering), Full Professor, Istanbul Technical University, Turkey e-mail : adali@itu.edu.tr*

Although the origin of the Turkic languages area the same, by the time they have been changed. In this paper we will show the similarities and differences of Turkic languages. We take six Turkic languages in our study which are Turkish, Azeri, Turkmen, Uzbek, Uighur, Kazakh and Tatar languages. In this context, the alphabets, phonology, morphology and syntax of Turkic language will be shown.

Keywords: Turkic Languages, alphabets, phonology, morphology, syntax

СХОДСТВА И ОТЛИЧИЯ ТЮРКСКИХ ЯЗЫКОВ

Адалы Эшреф, *Phd (Инженер), Профессор, Стамбульский Технический Университет, Турция, e-mail : adali@itu.edu.tr*

В происхождении областей Тюркских языков по прежнему, со временем они были изменены. В этой статье мы покажем сходства и отличия Тюркских языков. В нашем изучении мы берем шесть Тюркских языков, которые Турецкий, азербайджанский, туркмены, узбек, уйгур, казахский язык и татарский язык. В этом контексте будут показаны алфавит, фонологии, морфологии и синтаксис Тюркского языка.

Ключевые слова: Тюркские Языки, алфавит, фонология, морфология, синтаксис

1. Introduction

Original Turkic language is a very ancient language going back 5500 to 8500 years. The earliest written texts for the Turkic languages are the Old Turkic runic inscriptions of the Orkhon and Yenisey valleys (north central Mongolia) dating from 700 to 800. Turkic language has a phonetic, morphological and syntactic structure, and at the same time it possesses a rich vocabulary. By the time this language changed and some new languages were formed which are Turkish, Azeri, Turkmen, Uzbek, Uighur, Kazakh and Tatar languages.

In this chapter we will study the similarities and differences of Turkic languages those are Turkish, Azeri, Turkmen, Uzbek, Uighur, Kazakh and Tatar. In this context, the alphabets, phonology, morphology and syntax of Turkic language will be shown.

The fundamental and common features of Turkic languages are:

- Vowel harmony,
- The absence of gender,
- Agglutination,
- Adjectives precede nouns,
- Verbs come at the end of the sentence.

2. Phonology and Alphabet

The oldest alphabet of Turks as known Gokturk alphabet. We can see this alphabet on the monuments of Orhon, Yenisev and Talas which are presently in Mongolia. These monuments were erected in 8th century. After the waning of the Gokturk state, the Uighurs produced a new alphabet named Uighur. By the time Turks adopt Arabic, Krill and Latin alphabets depends on religious or political reason. In the Table-1, current situation is depicted.

As far as phonology is concern we can see that Turkic languages have reach vowels. The number of vowel is about 8 to 14 (some of them are presented by an accent). The position of vowel of Turkish language is shown on vowel quadrilateral, in Figure-1; the red letters are Turkish [1]. In Table-2, Turkish vowel is shown from different viewpoints.

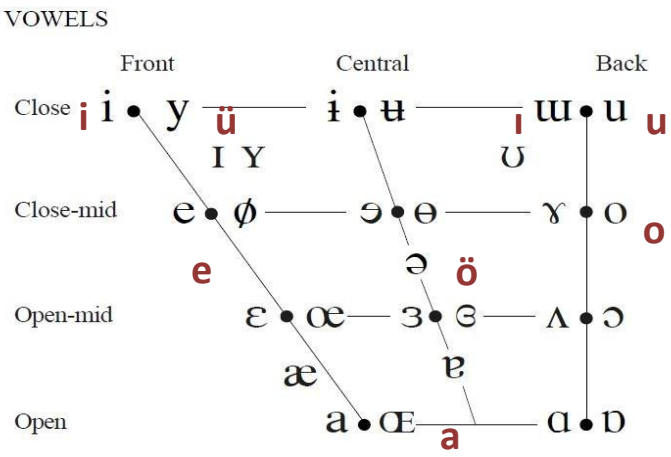


Figure 9: Turkish letters on vowel quadrilateral

Table 1: The Alphabet of Turkic Language, today

Turkish		Turkmen		Azeri		Uzbek		Uighur			Kazakh				Tatar				IPA
La	La	La	La	La	La	La	La	La	La	Ar	La	La	Kr	Kr	La	La	Kr	Kr	
A	a	A	a	A	a	A	a	A	A	ا	A	a	A	a	A	a	A	a	/a/
		Ä	ä	E	e			Ë	ë	ەٸ/	Ä	Ä	Ə	ə	Ä	ä	Ə	ə	/æ/
B	b	B	b	B	b	B	b	B	B	ب	B	B	Б	б	B	b	Б	б	/b/
C	c	J	j	C	c	J	j	J	J	ج					C	c	Д ж	д ж	/dʒ/
Ç	ç	Ç	ç	Ç	ç	Ch	ch	Ch	Ch	چ	Ç	Ç	Ч	ч	Ç	ç	Ч	ч	/tʃ/
D	d	D	d	D	d	D	d	D	D	د	D	D	Д	д	D	d	Д	д	/d/
E	e	E	e	Ə	ə	E	e	E	E	ى/	E	e	E	e	E	e	E	e	/e/, /æ/

				E	e						É	é	Э	э					/e/
											Yo	yo	Ё	ё	Yo	yo	Ё	ё	/jo/
F	f	F	F	F	f	F	f	F	F	ف	F	f	Ф	ф	F	f	Ф	ф	/f/
G	g	G	G	G	g	G	g	G	G	گ	G	g	Г	г	G	g	Г	г	/g/, /ɟ/
Ğ	ğ			Ğ	ğ	G’	g’	Gh	Gh	غ	Ğ	ğ	Ғ	ғ	Ğ	ğ	Г	г	/ɰ/
H	h	H	H	H	h	H	h	H	H	ھ	H	h	Һ	һ	H	h	Һ	һ	/h/
				X	x	X	X	X	X	خ	X	x	X	x	X	x	X	x	/x/
I	ı	Y	y	I	ı						I	ı	ӏ	ы	I	ı	ӏ	ы	/ɯ/
İ	i	İ	İ	İ	i	İ	İ	İ	I	ی/	Ї	ї	Ӏ	и	İ	i	Ӏ	и	/i/
J	j	Ž	ž	J	j			Zh	Zh	ژ	J	j	Ж	ж	J	j	Ж	ж	/ʒ/
K	k	K	k	K	k	K	K	K	K	ك	K	k	К	k	K	k	K	k	/k/, /c/
L	l	L	l	L	l	L	L	L	l	ل	L	l	Л	л	L	l	Л	л	/ɬ/, /l/
M	m	M	m	M	m	M	M	M	m	م	M	m	М	м	M	m	M	м	/m/
N	n	N	n	N	n	N	N	N	n	ن	N	n	Н	н	N	n	Н	н	/n/, /ŋ/
		Ň	ň			Ng	ng	Ng	ng	ڭ	Ñ	ñ	Ң	ң	Ñ	ñ	Ң	ң	Nasal n
O	o	O	o	O	o	O	O	O	o	و/	O	o	O	o	O	o	O	o	/o/
Ö	ö	Ö	ö	Ö	ö	O’	o’	Ö	ö	ۇ/	Ö	ö	Ө	ө	Ö	ö	Ө	ө	/œ/
P	p	P	p	P	p	P	P	P	p	پ	P	p	П	п	P	p	П	п	/p/
R	r	R	r	R	r	R	R	R	r	ر	R	r	Р	р	R	r	R	r	/r/
				Q	q	Q	Q	Q	q	ق	Q	q	Қ	қ	Q	q			/q/
S	s	S	s	S	s	S	S	S	s	س	S	s	С	с	S	s	С	с	/s/
Ş	ş	Ş	ş	Ş	ş	Sh	sh	Sh	sh	ش	Ş	ş	Ш	ш	Ş	ş	Ш	ш	/ʃ/
											Şş	şş	Щ	щ	Şş	şş	Щ	щ	/ʃtʃ/, /ʃː/
T	t	T	t	T	t	T	T	T	t	ت	T	t	Т	т	T	t	T	т	/t/
U	u	U	u	U	u	U	U	U	u	ۋ/	W	w	У	у	W	w	У	у	/u/
											U	u	ұ	ұ	U	u	У	у	/ʊ/
Ü	ü	Ü	ü	Ü	ü			Ü	ü	ۈ/	Ü	ü	Ү	ү	Ü	ü	Ү	ү	/y/
V	v	W	w	V	v	V	V	W	w	ۋ	V	v	В	В	W	w	В	В	/v/
Y	y	Ý	ý	Y	y	Y	Y	Y	y	ې	Y	y	Й	й	Y	y	Й	й	/j/
Z	z	Z	z	Z	z	Z	z	Z	z	ز	Z	z	З	з	Z	z	З	з	/z/
											Yw	yw	Ю	ю	Yw	yw	Ю	ю	/ju/, /jy/
											Ya	ya	Я	я	Ya	ya	Я	я	/ja/, /jɑ/

As we can see on Table-1, some nation use Latin (La) based alphabet some use Krill (Kr) and one of them use Arabic (Ar) alphabet. In order to make understandable, we will give all examples by Latin equivalent characters.

Table 2: The Turkish vowels

	Front		Central	Back	
	Un-round	Round		Un-round	Round
Close	**i**	**ü**		**ı**	**u**
Mid	**e**	**ö**			**o**
Open			**a**		

Table-3: Vowels of Turkic Languages

Language	Number of vowels	vowels
Turkish	8	a, e, ı, i, o, ö, u, ü
Azeri	9	a, e, é, ı, i, o, ö, u, ü
Turkmen	9	a, ä, e, y, i, o, ö, u, ü
Uzbek	6	a, o, o', u, e, i
Uighur	8	a, e, é, i, o, ö, u, ü
Kazakh	9	a, e, ä, ı, i, o, ö, u, ü
Kırgız	8	a, e, ı, i, o, o, u, u
Tatar	9	a, e, ä, ı, i, o, ö, u, ü

All Turkic languages have similar vowel harmony. The vowel and consonants harmonies of Turkish are drown in Figure-2 and Figure-3 respectively. Almost all Turkic languages obey this vowel and consonant rules. The vowel harmony create the sound of Turkish. There are no diphthongs in Turkish. Therefore if a suffix beginning with a vowel is attached to a stem ending in a vowel, either the initial vowel of the suffix is deleted, or the consonant 'y' is added. As a result, suffixes are divided into two groups: those which can lose their initial vowel and those which can acquire the buffer consonant 'y'.

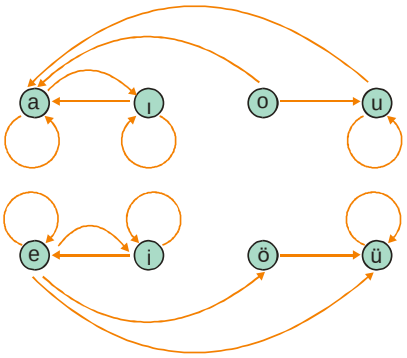


Figure 10 : Vowel harmony of Turkish

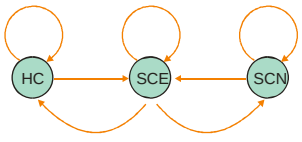


Figure 3 : Consonant harmony of Turkish Hard consonants (HC)

Soft consonants do not have hard equivalence consonants (SCE)

Soft consonants have hard equivalence consonants (SCN)

3. Morphology

Turkic languages are agglutinative languages that have productive inflectional and derivational suffixes. Therefore, we will study the morphology of Turkic language in detailed; singular and plural, pronouns, cases and verbs.

3.1 Singular and Plurals

The basic plural suffixes of Turkic language are –lAr. A will be either "e" or "a" depends on previous vowel due to vowel harmony. Some Turkic language have differences. The plural suffixes of Turkic languages are given in Table-4. If a numeral adjective before a noun, plural suffix will not be used. Eg;

Okul (school – singular), Okullar (schools – plural), Beş okul (Five Schools)

Table 4: Plural Suffixes of Turkic Languages

Language	**Last vowel letter**				**Last letter**	**Soft consonant**				**Hard consonant**				**Sonorant**			
Last vowel	**e,i**	**ö, ü**	**a,ı**	**o, u**	**r, w, y**	**e,i**	**ö, ü**	**a,ı**	**o, u**	**e,i**	**ö, ü**	**a,ı**	**o, u**	**e,i**	**ö, ü**	**a,ı**	**o, u**
Turkish	-ler	-ler	-lar	-lar													
Azeri	-ler	-ler	-lar	-lar													
Turkmen	-ler	-ler	-lar	-lar													
Uzbek	-lar	-lar	-lar	-lar													
Kirghiz	-ler	-lör	-lar	-lor		der	dör	dar	dor	-ter	-tör	-tar	-tor				
Kazakh	-ler	-ler	-lar	-lar	lAr	der	der	dar	dar	-ter	-ter	-tar	-tar				
Uighur	-ler	-ler	-lar	-lar													
Tatar	-ler	-ler	-lar	-lar										när	när	nar	nar

3.2 Pronouns

The basic pronouns of Turkic languages are the same, but pronunciation varies a little bit. Essentially there are six personal pronoun; first, second and third singular and first, second and third plural. The second singular may have three forms; informal, formal and respectful. Uzbek, Kirgiz and Kazakh languages have two and Uyghur language has three forms of second plural pronoun. As can be seen, there is no gender for third singular person.

Table 5: Pronouns of Turkic Languages

	Turkish	**Azeri**	**Turkmen**	**Uzbek**	**Kirghiz**	**Kazakh**	**Uyghur**	**Tatar**
1. single	ben	mən	men	men	men	men	men	min
2. single (informal)	sen	sən	sen	sen	sen	sen	sen	sin
2. single (formal)	siz			siz	siz	siz	siz	

2. single (respectful)							sili	
3. single	o	O	ol	u	al	ol	u	ul
1. plural	biz	Biz	biz	biz	biz	biz	biz	bez
2. plural (informal)	siz	Siz	siz	siz	siler	sender	senler	sez
2. plural (formal)				sizlar	sizder	sizder	siler	
2. plural (respectful)							sizler	
3. plural	onlar	Onlar	olar	ular	alar	olar	ular	alar

3.3 Cases

The cases are very important in Turkic languages. They play the roles of prepositions and postpositions like *at, in, to, from* and *on, under, in, with* etc. In general, there are 5 or 6 noun case: simple case, accusative (-i), dative (-e), locative (-de), ablative (-den), genitive (-in) and instrumental (-le). When a case is added to noun, the rules of vowel and consonant harmonies are applied and also added a buffer letter if the last letter is a vowel. All possible cases of Turkic languages are given in Table-6. [2], [8]

Table 6 : The Cases of Turkic Languages

	Last cons.	Last vow.	**Turkish**	**Azeri**	**Turkmen**	**Uzbek**	**Kirghiz**	**Kazakh**	**Uyghur**	**Tatar**
Accusative (I)		e,i	-(y) i *	-(n) I	-(n)i	-ni	-ni	-ni	-ni	-ne
		ö,ü	-(y) ü	-(n) ü	-(n)ü		-nü	-di	-ni	-ne
		a,ı	-(y) ı	-(n) ı	-(n)y	-nı	-nı	-nı	-ni	-nı
		o,u	-(y) u	-(n) u	-(n)y	-ni	-n*ü*	-dı	-ni	-nı
	SC						-di	-di		
							-dü	-di		
							-dı	-dı		
							-du	-dı		
	HC						-ti	-ti		
							-tü	-ti		
							-tı	-tı		
							-tu	-tı		
		e,i	-(y) e	-(y) *ə*	-(n) e	-ga	-ga, a-e	-ge	-ge	-gä
		ö,ü	-(y) e	-(y) *ə*	-(n) e	-ga	-ge	-ge	-ge	-gä
		a,ı	-(y) a	-(y) a	-(n) a ň)a (	-ga	-ğa	-ğa	-ğa/-	-ga
		o,u	-(y) a	-(y) a	-(n) a	-ga	-ğö	-ğa	-ğa/-	-ga
	HC					-ke	-ke	-ke	-ke	-ke
						-ke	-ke	-ke	-ke	-ke
						-qa	-ka	-qa	-qa	
										-qa
						-qa	-kö	-qa	-qa	
Locative (de)		e,i	-de	-d*ə*	-de, -nde	-da	-de	-de	-de	--qadä
		ö,ü	-de	-	-de, -nde	-da	-de	-de	-de	-dä

				dɑ						
		a,ı	-da	-da	-da, -nda	-da	-da	-da	-da	-da
		o,u	-da	-da	-da, -nda	-da	-dö	-da	-da	-da
	HC		-te	-də			-ta	-te	-te	-tä
			-te	-dɑ			-te	-te	-te	-tä
			-ta	-da			-to	-ta	-ta	-ta
			-ta	-da			-tö	-ta	-ta	-ta
Ablative (den)		e,i	-den	-dɑn	-den	-dän	-dan	-den	-din	-dän
		ö,ü	-den	-dɑn	-den	-dän	-den	-den	-din	-dän
		a,ı	-dan	-dan	-dan	-dän	-don	-dan	-din	-dan
		o,u	-dan	-dan	-dan	-dän	-dön	-dan	-din	-dan
	HC		-ten	-dɑn			-tan	-ten	-tin	-tän
			-ten	-dɑn			-ten	-ten	-tin	-tän
			-tan	-dan			-ton	-tan	-tin	-tan
			-tan	-dan			-tön	-tan	-tin	-tan
Genitive (in)		e,i	-(n) in	-(n) in	-(n) iň	-niň	-nın	-ni ň	-niň	-ne ň
		ö,ü	-(n) ün	-(n) ün	-(n) ü ı	-nü ň	-nün		-niň	-ne ň
		a,ı	-(n) ın	-(n) ın	-(n) yň	-nıň	-nin	-nı ň	-niň	-nıň
		o,u	-(n) un	-(n) un	-u ı (n)	-nu ň	-n nu		-niň	-nıň
	SC						-dir	-di ň		
							-n dü			
							-dır	-dı ň		
							-n du			
	DC						-tin	-tiň		
							-tür			
							-tın	-tıň		
							-tur			

* (y) will be y or n

3.4 Verbs

Although Turkic languages have five basic and many sophisticated tenses, we will examine just basics tenses. Examples are given for Turkish.

- Simple (Aorist) 'Wide' Tense : *gelirim; I come, I'll come* ○ Present Continuous Simple Tense : *geliyorum; I am coming* ○ Future, Simple : *geleceğim; I will come*
- Past, Definite Tense (-di) (seen past tense) : *geldim; I came, I did come* ○ Past Dubitative, Simple (-miş) (heard/perceived/reported past tense) : *It's said that I came* ○ Present Dubitative Continuous Compound Tense : *geliyormuşum; It's said that I'm coming* ○ Present Continuous, Compound Conditional Tense : *geliyorsam ; If I'm coming* ○ Present Dubitative Simple Compound Tense : *gelirmişim; It's said that I would come* ○ Present Simple Compound Conditional Tense : *gelirsem; If I come, if I should/would come* ○ Present Perfect Continuous Tense : *gelmekteyim; I have been coming* ○ Past Continuous Compound Narrative Tense : *geliyordum; I was coming* ○ Present Perfect Tense : *geldim; I have come*
- (Timeless) Past, Simple Compound Narrative Tense : *gelirdim; I would come, I used to come*
- Colloquial Past, Compound Narrative : *I had come, I came, I have come* ○ Past Definite, Compound Conditional : *geldiysem; If I came, if I have come* ○ Past Perfect, Compound Narrative : *gelmiştim; I had come*
- Past Perfect, Continuous : *gelmekteydim (gelmekte idim); I had been coming* ○ Past, Dubitative Compound (a tense for sarcasm) : *gelmişmişim*; *It's said that I had come* ○ Past Dubitative, Compound Conditional : *gelmişsem; If I have come* ○ Future Continuous : *geliyor olacağım*; *I shall be coming* ○ Future Perfect Continuous : *gelmekte olacağım*; *I shall have been coming* ○ Future Perfect (Past in the Future) : *gelmiş olacağım; I shall/will have come* ○ Future in the past : *gelecektim*; *I was going to come*
- Future, Dubitative Compound : *gelecekmişim; It's said that I'm going to come* ○ Future, Compound Conditional : *geleceksem; If I'm going to come*

The form of a verb consists of stem (infinive), tense and person information even though personal pronoun is missing. For example; *Geleceğim* : I will come. The structure of a verb is shown in Figure-4. According to morphological structure of Turkic language a verb composed by consecutive suffixes; stem (infinitive form of verb), tens and personal suffix. Some exception will be given related table.

Figure 4: The structure of a Turkish verb

In order to study of verb structures of Turkic languages, two verbs are taken; gelmek (to come, stem is gel) and okumak (to read, stem is oku). These two example will show us harmony of vowel harmony and consonant. The following abbreviations will be used. In order to indicate similarities and differences between languages, the similar ones are grouped in tables.

○ A for a, ä, e, ə ○ H for ı, i, u, ü, o, u

3.4.1 Tenses

The stem of will be the same for all tenses. The basic structure of a Turkic verbs is shown in Table-8. [1], [2], [3], [4], [5], [6], [7], [8], [11]

Table-8 : The basic Structure of a Turkish verb

Stem/ infinitive	Suffix			
	Negative	Tense	Interrogative (*)	Personal (*)

(*) The position of interrogative and personal suffix may switch their position

Table-9 : Personal Suffix of verbs

	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl
TUR	-Hm	-Hn	-Ø	-rHz	-sHnHz	-lAr
AZE	-Am	-sAn	-Ø	-ıx, -ik, -ux, -ük	-sHnHz, SHz	-lAr
TKM	-(H)n	sHň	-Ø	-(H)s	-sHňHz	-lAr
KAZ	-mHn, -pHn	-sHň, -sHz	-Ø	-bHz, -pHz	-sHňdAr, sHzdAr, sHz	-Ø,
KIR	-mHn, -m	-sHň	-Ø	-bHz	-sHňAr	-ş
TAT	-mın,-men, m	-sın, siň	-Ø	-bHz,	-sHz	-lAr
UYG	-men	-sin	-Ø	-miz	-siler	-Ø, -lAr, -ş-
UZB	-män	kel-ä--säň	-Ø , -di, -ti	-miz	-siz, -lär	-Ø, -lär, -(I)ş

The Examples of five basic tenses of these two verbs are in the following tables:

Simple Present Tense							
		Affir: Stem-(H)r-ps		Neg: Stem-mA-ps		Interr: Stem-(H)r mH-ps? (*)	
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
TUR	gel-ir-im	gel-ir-sin	gel-ir	gel-ir-riz	gel-ir-siniz	gel-ir-ler	gel-me-m
	oku-r-um	oku-r-sun	oku-r	oku-r-uz	oku-r-sunuz	oku-r-lar	oku-ma-m
		Affir: Stem-(H)r-ps		Neg: Stem-mA-sl		Interr: Stem-(H)r-ps mH?	
AZE	gəl-ər-əm	gəl-ər-sən	gəl-ər	gəl-ər-ik	gəl-ər-siniz	gəl-ər-lər	gəl-mər-əm
	oxu-yar-am	oxu-yar-san	oxu-yar	oxu-yar-ıq	oxu-yarsınız	oxu-yar-lar	oxu-mar-am
TKM	gel-er-in	gel-er-siň	gel-er	gel-er-is	gel-er-siňiz	gel-er-ler	gel-me-r-in
	oka-r-yn	oka-r-syň	oka-r	oka-r-ys	oka-r-syňyz	oka-r-lar	oka-ma-r-syň
KAZ	kel-er-min	kel-er- siň kel-er- siz	kel-er	kel-er-miz	kel-er-siňder kel-er-sizder	kel-er	kel-mes-pin
	oqı-r-mın	oqı-r-sıň oqı-r-sız	oqı-r	oqı-r-mız	oqı-r-sıňdar oqı-r-sızdar	oqı-r	oqı-mas-pın
TAT	kil-er-men	kil-er-seň	kil-er	kil-er-bez	kil-er-sez	kil-er-lär	kil-me-men
	ukı-r-mın	ukı-r-siň	ukı-r	ukı-r-bız	ukı-r-sız	ukı-r-lar	ukı-ma-mın
		Affir: Stem-(ä,y)-ps		Neg: Stem-mäy-ps		Interr: Stem-(ä,y)-ps mH?	
UZB	kel-a-man	kel-a--san	kel-a-di	kel-a-miz	kel-a-siz	kel-a-di-lar	kel-may-man
	oqi-y-man	oqi-y-san	oqi-y-di	oqi-y-miz	oqi-y-siz	oqi-y-di-lar	oqi-may-man

KIR	kel-e-min	kel-e-sin	kel-et	kel-e-biz	kel-e-sıňer	kel-i-şet	kel-bey-min
	oqu-y-mun	oqu-y-sun	oqu-y-t	oqu-y-buz	oqu-y-suňar	oqu-şat	oqu-bay-mın
UYG	kél-i-men	kél -i-sen	kél -i-du	kél -i-miz	kél -i-siler	kél -i-du	kél -mey-men
	oqu-y-men	oqu-y-sen	oqu-y-du	oqu-y-miz	oqu-y-siler	oqu-y-du	oqu-may-men

(*) exception in Turkish gel-lir-ler mi?

Present Continuous Tense							
		Affir: Stem-(H)yor-ps		Neg: Stem-mH-(H)yor-ps		Interr: Stem-(H)yor-ps mH? (*)	
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
TUR	gel-i-yor-um	gel-i-yor-sun	gel-i-yor	gel-i-yor-uz	gel-i-yorsunuz	gel-i-yor-lar	gel-mi-yor-um
	oku-yor-um	oku-yor-sun	oku-yor	oku-yor-uz	oku-yorsunuz	oku-yor-lar	oku-mu-yor-um
AZE	gəl-ir-əm	gəl-ir-sən	gəl-ir	gəl-ir-ik	gəl-ir-siniz	gəl-ir-lər	gəl-mir-əm
	oxu-yur-am	oxu-yur-san	oxu-yur	oxu-yur-uq	oxu-yursunuz	oxu-yur-lar	oxu-mur-am
TKM	gel-ýär-in	gel-ýär-siň	gel-ýär	gel-ýär-is	gel-ýär-siňiz	gel-ýär-ler	gel-me-ýär-in
	oka-ýar-yn	oka-ýar-syň	oka-ýar	oka-ýar-ys	oka-ýarsyňyz	oka-ýar-lar	oka-ma-ýar-yn
		Affir: Stem-vati-ps		Neg: Stem-mAy-vati-ps		Interr: Stem-vati-ps mH?	
UYG	kel-vati-men	kel-vati-sin	kel-vati-du	kel-vati-miz	kel-vati-siler	kel-vati-du	kel-mey-vatimän
	oqu-vatimen	oqu-vati-sen	oqu-vati-du	oqu-vati-miz	oqu-vatisiler	oqu-vati-du	oqu-may-vatimen
		Affir: Stem-yap-ps		Neg: Stem-mA-yap-ps		Interr: Stem-yap-ps mH?	
UZB	kel-yap-man	kel-yap-san	kel-yapti	kel-yap-miz	kel-yap-siz	kel-yapti-lar	kel-ma-yap-man
	oqi-yap-man	oqi-yap-san	oqi-yapti	oqi-yap-miz	oqı-yap-siz	oqi-yapti-lar	oqi-ma-yap-man
		Affir: Stem-Hvde, UUdo-ps		Neg: Stem-Hvde,UUdoemes-ps		Interr: Stem-Hvde, UUdo-ps mH?	
KIR	kel-üüdömün	kel-üüdö-siň	kel-üüdö	kel-üüdö-büz	kel-üüdösüňö	kel-üüdö	kel-böödö-mün
	oquuda-mın	oquu-da-sin	oquu-da	oquu-da-bız	oquu-dasıňar	oquu-da	oquu-baadomun
		Affir: Stem-(e,y)-ps		Neg: Stem-mH-ps		Interr: Stem-(e,y)-ps mH?	
TAT	kil-ä-m	kil-ä-seň	kil-ä	kil-ä-bez	kil-ä-sez	kil-ä-lär	kil-mi-m
	ukı -y-m	ukı-y-sıň	ukı-y	ukı-y-bız	ukı-y-sız	ukı-y-lar	ukı-mı-y-m
		Affir: Stem-(a, -e,-y, -ıp,-ip,-p) *otır, tur, jatır, jür,* -ps		Neg: Stem -g(q)An-joq		Interr: Stem-(a, -e,-y, -ıp,-ip,-p) *otır, tur, jatır, jür,* -ps ba?	
KAZ	kel-e jatırmın	kel-e jatır-sıň kel-e jatır-sız	kel-e jatır	kel-e jatırmız	kel-e jatırsıňdar kel-e jatırsızdar	kel-e jatır	kele jat-qan joq kel-e jatır-mın ba?

	okı-p otyrmın	okı-p otyrsıň okı-p otyr-sız	okı-p otır	okı-p otyrmız	okı-p otyrsızdar okı-p otyrsızdar	okı-p otır	okı-p otyr-gan joq? okı-p otyr-mın ba?

(*) Exeption in Turkish gel-liyor-mu yum?

Future Tense							
		Affir: Stem-(y)AcAk-ps		Neg: Stem-mA-(y)AcAk-ps		Interr: Stem-(y)AcAk-ps mH(*)	
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
TUR	Gel-eceğ-im	Gel-ecek-sin	gel-ecek	gel-eceğ-iz	gel-eceksiniz	gel-ecek-ler	gel-me-yeceğ-im
	oku-yacağım	oku-yacak-sın	oku-yacak	oku-yacağ-ız	oku-yacak-sınız	oku-yacak-lar	oku-ma-yacağım
AZE	gəl-əcəy-əm	gəl-əcək-sən	gəl-əcək	gəl-əcəy-ik	gəl-əcəksiniz	gəl-əcək-lər	gəl-mə-yəcəyəm
	oxu-yacağam	oxu-yacaksan	oxu-yacak	oxu-yacağ-ıq	oxu-yacaq-sınız	oxu-yacaq-lar	oxu-ma-yacağam
TAT	kil- äçäkmen	kil-äçäk-seň	kil-äçäk	kil-äçäk-bez	kil-äçäk-sez	kil-äçäk-ler	kil-mä-yäçäkmen
	ukı-yaçakmın	ukı-yaçak-sıň	ukı-yaçak	ukı-yaçakbız	ukı-yaçak-sız	ukı-yaçak-lar	ukı-ma-yaçakmın
		Affir: ps-Stem-jAk		Neg: ps-Stem-jAk dääl		Interr: ps-Stem-jAk mH?	
TKM	men gel-jek	sen gel-jeksiň	ol gel-jek	biz gel-jek	siz gel-jeksiňiz	olar gel-jekler	men gel-jek däl
	men oka-jak	sen oka-jaksyň	ol oka-jak	biz oka-jak	siz oka-jaksyňyz	olar oka-jaklar	men oka-jak däl
		Affir: Stem-(H,A)-ps		Neg: Stem-(m,p)Ay-ps		Interr: Stem-(H,A)-ps mH?	
KAZ	kel-e-min	kel-e-siň	kel-e-t	kel-e-biz	kel-e-siňer	kel-i-şet	kel-bey-siň
	oqu-y-mun	oqu-y-suň	oqu-y-t	oqu-y-buz	oqu-y-suňar	oqu-şat	oqu-bay-sıň
KIR	kel-e-mın	kel-e-siň	kel-e-t	kel-e-bız	kel-e-siňar	kel-e-şat	kel-pey-siň
	oqı-y-mın	oqı-y-siň	oqı-y-t	oqı-y-bız	oqı-y-siňer	oqı-şet	oqı-pay-siň
UZB	kel-a-man	kel-a-san	kel-a-di	kel-a-miz	kel-a-siz	kel-a-dilar	kel-ma-y-man
	oqi-y-man	oqi-y-san	oqi-ydi	oqi-y-miz	oqi-y-siz	oqi-y-dilar	oqi-ma-y-man
		Affir: Stem-(H)-ps		Neg: Stem-mAy-ps		Interr: Stem-(H)-ps mH?	
UYG	kél-i-men	kél-i-sen	kél-i-du	kél-i-miz	kél-i-siler	kél-i-du	kél-mey-men
	oqu-idiğanmen	oqu-ydiğansen	oqu-ydiğandu	oqu-ydiğanmiz	oqu-ydiğan-siler	oqu-ydiğan	oqu-mayydiğan-men

Past Definite Tense (-dili)							
		Affir: Stem-dH-ps		Neg: Stem-mA-dH-ps		Interr: Stem-dH-ps mH	
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
TUR	gel-di-m	gel-di-n	gel-di	gel-di-k	gel-di-niz	gel-di-ler	gel-me-di-m

	oku-du-m	oku-du-n	oku-du	oku-du-k	oku-du-nuz	oku-du-lar	oku-ma-du-m
AZE	gəl-di-m	gəl-di-n	gəl-di	gəl-di-k	gəl-di-niz	gəl-di-lər	gəl-mə-di-m
	oxu-du-m	oxu-du-n	oxu-du	oxu-du-q	oxu-du-nuz	oxu-du-lar	oxu-ma-dı-m
TKM	gel-di-m	gel-di-ň	gel-di	gel-di-k	gel-di-ňiz	gel-di-ler	gel-me-di-m
	oka-dy-m	oka-dy-ň	oka-dy	oka-dy-k	oka-dy-ňyz	oka-dy-lar	oka-ma-dy-m
UZB	kel-di-m	kel-di-ng	kel-di	kel-di-k	kel-di-ngiz	kel-ishdi-lar	kel-ma-di-m
	oqi-di-m	oqi-di-ng	oqi-di	oqi-di-k	oqi-di-ngiz	oqi-shdi-lar	oqi-ma-di-m
KAZ	gel-di-m	kel-di-ň kel-di-ňiz	kel-di	kel-di-k	kel-di-ňder kel-di-ňizder	kel-di	kel-me-di-m
	oqı-dı-m	oqı- dı-ň oqı- dı-ňız	oqı-dı	oqı-dı-k	oqı-dı-ňdar oqı-dı-ňızdar	oqı-dı	oqı-ma-dı-m
KIR	kel-di-m	kel-di-ň	kel-di	kel-di-k	kel-diňiz(der)	kel-iş-ti	kel-be-di-m
	oqu-du-m	oqu-du-ň	oqu-du	oqu-du-k	oqu-duňuz(dar)	oqu-ş-tu	oqu-ba-dı-m
UYG	kel-di-m	kel-di-ň	kel-di	kel-di-k	kel-di-ňlar	kel-di	kel-mi-di-m
	oqu-dı-m	oqu-dı-ň	oqu-dı	oqu-dı-q	oqu-dı-ňlar	oqu-dı	oqı-mi-dy-m
TAT	kil-de-m	kil-de-ň	kil-de	kil-de-k	kil-de-gez	kil-de-lär	kil-mä-de-m
	ukı-dı-m	ukı-dı-ň	ukı-dı	ukı-dı-k	ukı-dı-gız	ukı-dı-lar	ukı-ma-dı-m

Past Tense (-miş)							
		Affir: Stem-mHş-ps		Neg: Stem-mA-mHş-ps		Interr: Stem-mHş-ps mH	
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
TUR	gel-miş-im	gel-miş-sin	gel-miş	gel-miş-iz	gel-miş-siniz	gel-miş-ler	gel-me-miş-im
	oku-muş-um	oku-muş-un	oku-muş	oku-mu-şuz	oku-muşsunuz	oku-muş-lar	oku-ma-mış-ım
AZE	gəl-miş-əm	gəl-miş-sən	gəl-miş-dir	gel-miş-ik	gəl-miş-siniz	gəl-miş-lər	gəl-mə-miş-əm
	oxu-muş-am	oxu-muş-san	oxu-muş-dur	oxu-muş-uq	oxu-muşsunuz	oxu-muş-lar	oxu-ma-mış-am
		Affir: Stem-ip(iptir)-ps		Neg: Stem-mA(pA)-ip(iptir)ps		Interr: Stem-ip(iptir)-ps mH	
TKM	gel-ipdir-in	gel-ip-siň	gel-ipdir	gel-ipdir-is	gel-ip-siňiz	gel-ipdir-ler	gel-mä-ndir-in
	oka-pdyr-yn	oka-p-syň	oka-pdyr	oka-pdyr-ys	oka-p-syňyz	oka-pdyr-lar	oka-ma-ndyr-yn
KIR	kel-iptir-min	kel-iptir-siň	kel-iptir	kel-iptir-biz	kel-iptirsiňer	kel-işiptir	kel-be-ptir-min
	oqu-pturmun	oqu-ptur-suň	oqu-ptur	oqu-ptur-buz	oqu-ptursuňar	oqu-şuptur	oqu-ba-ptır-mın
UZB	kel-ib-man	kel-ib-san	kel-ib-di	kel-ib-miz	kel-ib-siz	kel-ib-di-lar	kel-mab-man
	oqi-b-man	oqi-b -san	oqi-b-di	oqi-b-miz	oqi-b-siz	oqi-b-di-lar	oqi-mab-man
UYG	kél-ip-ti-men	kél-ip-sen	kél-ip-tu	kél-ip-siz	kél-ip-siler	kél-ip-ti	kél-me-p-ti-men
	oqu-p-ti-men	oqu-p-sen	oqu-p-tu	oqu-p-siz	oqu-p-siler	oqu-p-tu	oqu-ma-p-timen
KAZ (I)	kel-ip-pin	kel-ip-siň kel-ip-siz	kel-ip-ti	kel-ip-piz	kel-ip-siňder kel-ip-sizder	kel-ip-ti	kel-me-p-pin
	oqı-p-pın	oqı-p-sıň oqı-p-sız	oqı-p-tı	oqı-p-pız	oqı-p-sıňdar oqı-p-sızdar	oqı-p-tı	oqı-ma-p-pın

		Affir: Stem-g(q)An-ps		Neg: Stem-mA-g(q)An-ps		Interr: Stem-g(q)An-ps mH	
TAT	kil-gän-men	kil-gän-seň	kil-gän	kil-gän-bez	kil-gän-sez	kil-gän-när	kil-mä-gän-men
	ukı-gan-mın	ukı-gan-sıň	ukı-gan	ukı-gan-bız	ukı-gan-sız	ukı-gan-nar	ukı-ma-gan-mın
KAZ (II)	kel-gen-min	kel-gen-siň kel-gen-siz	kel-gen	kel-gen-biz	kel-gensiňder kel-gensizder	kel-gen	kel-me-gen-min
	okı-gan-mın	okı-gan-sıň okı-gan-sız	okı-gan	okı-gan-bız	okı-gansıňdar okı-gansızdar	okı-gan	okı-ma-gan-mın

Conditional¤								
¤	1Sg¤	2Sg¤	3Sg¤	1Pl¤	2Pl¤	3Pl¤	Negative¤	Interrog.¤
	Affirmative:·Stem-sA-ps¤		Negative:·Stem-mA-sA-ps¤			Interrogative:·Stem-sA-ps·mH?¤		
TUR¤	gel-se-m¤	gel-se-n¤	gel-se¤	gel-se-k¤	gel-se-niz¤	gel-se-ler¤	gel-me-sem¤	gel-sem·mi?¤
	oku-sa-m¤	oku-san¤	oku-sa¤	oku-sa-k¤	oku-sa-nız¤	oku-sa-lar¤	oku-ma-sam¤	oku-sam·mı?¤
ZE¤	gəl-sə-m¤	gəl-sə-n¤	gəl-sə¤	gəl-səк¤	gəl-sə-niz¤	gəl-sə-lər¤	gəl-mə-səm¤	gəl-səm·mi?¤
	oxu-sa-m¤	oxu-san¤	oxu-sa¤	oxu-sa-q¤	oxu-sa-nız¤	oxu-sa-lar¤	oxu-ma-sam¤	oxu-sam·mı?¤
TKM¤	gel-se-m¤	gel-se-ň¤	gel-se¤	gel-se-k¤	gel-se-ňiz¤	gel-se-ler¤	gel-me-sem¤	gel-semmi?¤
	oka-sa-m¤	oka-saň¤	oka-sa¤	oka-sa-k¤	oka-sa-ňız¤	oka-sa-lar¤	oka-ma-sam¤	oka-sammy?¤
UYG¤	kel-se-m¤	kel-se-ň¤	kel-se¤	kel-se-k¤	kel-se-ňler¤	kel-se¤	kel-mi-sem¤	kel-sem·mi?¤
	oqu-sa-m¤	oqu-saň¤	oqu-sa¤	oqu-sa-k¤	oqu-saňlar¤	oqu-sa¤	oqu-mi-sam¤	oqu-sam·mı?¤
TAT¤	kil-sä-m¤	kil-sä-ň¤	kil-sä¤	kil-sä-k¤	kil-sä-gez¤	kil-sä-lär¤	kil-mä-säm¤	kil-säm-me?¤
	ukı-sa-m¤	ukı-saň¤	ukı-sa¤	ukı-sa-k¤	ukı-sa-gız¤	ukı-sa-lar¤	ukı-ma-sam¤	ukı-sam-mı?¤
KAZ¤	kel-se-m¤	kel-se-ň¶ kel-se-ňiz¤	kel-se¤	kel-se-k¤	kel-se-ňder¶ kel-se-ňizder¤	kel-se¤	kel-me-sem¤	kel-sem·be?¤
	oqı-sa-m¤	oqı-sa-ň¶ oqı-sa-ňız¤	oqı-sa¤	oqı-sa-k¤	oqı-sa-ňdar¶ oqı-sa-ňızdar¤	oqı-sa¤	oqı-ma-sam¤	oqı-sam·ba?¤
KIR¤	kel-se-m¤	kel-se-ň¤	kel-se¤	kel-se-k¤	kel-se-ňer¤	kel-iş-se¤	kel-be-sem¤	kel-sembi?¤
	oqu-sa-m¤	oqu-saň¤	oqu-sa¤	oqu-sa-k¤	oqu-sa-··ňar¤	oqu-ş-sa¤	oqı-ba-sam¤	oqı-sambı?¤
UZB¤	kel-sa-m¤	kel-sa-ng¤	kel-sa¤	kel-sa-k¤	kel-sa-ngiz¤	kel-sa-lar¤	kel-ma-sam¤	kel-sammi?¤
	oqi-sa-m¤	oqi-sang¤	oqi-sa¤	oqi-sa-k¤	oqi-sa-ngiz¤	oqi-sa-lar¤	oqi-ma-sam¤	oqi-sammi?¤

Obligational							
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
		Affir: Stem-mAlH-ps		Neg: Stem-mA-mAlH-ps		Interr: Stem-mAlH-ps mH?	
TUR	gel-meli-yim	gel-meli-sin	gel-meli	gel-meli-yiz	gel-melisiniz	gel-meli-ler	gel-me-meli-yim
	oku-malıyım	oku-malı-sın	oku-malı	oku-malı-yız	oku-malısınız	oku-malı-lar	oku-ma-malıyım
AZE	gəl-məli-yəm	gəl-məli-sən	gəl-məli	gəl-məli-yik	gəl-məlisiniz	gəl-məli-lər	gəl-mə-məliyəm
	oxu-malıyam	oxu-malı-san	oxu-malı	oxu-malı-yıq	oxu-malısınız	oxu-malı-lar	oxu-ma-malıyam
		Affir: ps-Stem-mAlH		Neg: ps-Stem-mAlH dääl		Interr: ps-Stem-mAlH mH?	
TKM	men gel-meli	sen gel-meli	ol gel-meli	gel- meli	siz gel-meli	olar gel-meli	men gel-meli däl
	men okamalı	sen oka-malı	ol oka-malı	biz oka-malı	siz oka-malı	olar okumalı	men oka-maly däl

		Affir: Stem-e-ps kHrek		Neg: Stem-mA-ps kHrek		Interr: Stem-e-ps kHrek mH	
TAT	miňa kil-ergä kiräk	siňa kil-ergä kiräk	aňa kil-ergä kiräk	bezgä kilergä kiräk	sezgä kilergä kiräk	alarga kilergä kiräk	miňa kil-mäskä kiräk
	miňa ukı-rga kiräk	siňa ukı-rga kiräk	aňa ukı-rga kiräk	bezgä ukırga kiräk	sezgä ukı-rga kiräk	alarga ukırga kiräk	miňa ukı-mäskä kiräk

		Affir: Stem-V-ps kHrek		Neg: Stem-V- ps kHrek emes		Interr: Stem-e-ps kHrek mH	
KAZ	kel-üv-im kerek	kel-üv-in kerek	kel-üvi kerek	kel-üv-imiz kerek	kel-üv-i larıň kerek	kel-üvi kerek	kel-üv-im kerek emes
	oqı-v-ım kerek	oqı-v-ıň kerek	oqı-v-ı kerek	oqı-v-ımız kerek	oqı-v-larıň kerek	oqı-v-ı kerek	oqı-v-ım kerek emes
KIR	kel-üü-m kerek	kel-üü-n kerek	kel-üü-sü kerek	kel-üü-büz kerek	kel-üü-ňör kerek	kel-üü-sü kerek	kel-be-ş-im kerek
	Oqu-ş-um kerek	oqu-ş-uň kerek	Oqu-şu kerek	Oqu-şu-buz kerek	Oqu-şu-ňar kerek	oqu-şu kerek	Oqu-ba-ş-ım kerek emes
UZB	kel-ish-im kerak	kel-ish-ing kerak	kel-ish-i kerak	kel-ish-imiz kerak	kel-ish-ingiz kerak	kel-ish-(lari) kerak	kel-mas-lgim kerak
	oqi-sh-im kerak	oqi-sh-ing kerak	oqi-sh-i kerak	oqi-sh-imiz kerak	oqi-sh-ingiz kerak	oqi-sh-(lari) kerak	oqi-mas-ligim kerak
UYG	kél-iş-im kérek	kél-iş-iň kérek	kél-iş-i kérek	kél-iş-imiz kérek	kél-iş-iň-lar kérek	kél-iş-i kérek	kél-mes-lik-im kérek
	oqu-şum kérek	oqu-şuň kérek	oqu-şi kérek	oqu-şum-miz kérek	oqu-şuň-lar kérek	oqu-şi kérek	oqu-mas-liq-im kérek

Imperetives							
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
		Affir: Stem-special		Neg: Stem-mA-special		Interr: Stem-special mH?	
TUR	gel-e-yim	gel	gel-sin	gel-el-imz	gel-in	gel-sin-ler	gel-me-yeyim
	oku-yayım	oku	okusun	oku-ya-lım	oku-yun	oku-sun-lar	oku-ya-yım
AZE	gəl-im	gəl	gəl-sin	gəl-ək	gəl-iniz	gəl-sin-lər	gəl-mə-səm
	oxu-y-um	oxu	oxu-sun	oxu-y-ag	oxu-yun	oxu-sun-lar	oxu-ma-yam
TKM	gel-e- ýin	gel	gel-sin	gel-e-liň	gel-li-ň	gel-sin-ler	gel-me- ýin
	oka- ýyn	oka	oka-syn	oka-lyň	oka-ň	oka-syn-lar	oka-ma- ýyn
UYG	kel-ey Kéley	Kel (-gin)	kel-sun	kel-eyli kéleyli	kel-iňlar kéliňlar	kel-sun	kel-m-ey
	oqu-y	Oqu (-ghin)	oqu-sun	oqu-ylı	oqu-uňlar	oqu-sun	oqu-m-ay
TAT	kil-im	kil	kil-sen	kil-ik	kil-egez	kil-sen-när	kil-mi-m
	ukı-ym	ukı	ukı-sın	ukı-yk	ukı-gız	ukı-sın-nar	ukı-mı-ym
KAZ	kel-e-yin	kel kel-iňiz	kel-sin	kel-e-yik	kel-iňder kel-iňizder	kel-sin	kel-me-yin
	oqı-y-ın	oqı oqı-ňız	oqı-sın	oqı-yıq	oqı-ňdar oqı-ňızdar	oqı-sın	oqı-ma-yın
KIR	kel-e-yin	kel	kel-sin	kel-e-li	kel-gile	kel-iş-sin	kel-be-yin
	oqu-y-un	oqu	oqu-sun	oqu-ylu	oqu-gula	oqu-ş-sun	oqu-ba-yın
UZB	kel-ay	kel-(gin)	kel-sin	kel-aylik	kel-inglar	kel-sinlar	kel-ma-y
	oqi-y	oqi -(gin)	oqi -sin	oqi -ylik	oqi -nglar	oqi -sinlar	oqi -ma-y

Wish							
	1Sg	2Sg	3Sg	1Pl	2Pl	3Pl	Negative
		Affir: Stem-(I,y)A-ps		Neg: Stem-mA(I,y)A-ps		Interr: Stem-(I,y)A-ps mH?	
TUR	gel-le-yim	gel-le-sin	gel-le	gel-le-lim	gel-le-siniz	gel-le-ler	gel-me-yei-yim
	oku-ya-yım	oku-ya-sın	oku-ya	oku-ya-lım	oku-ya-sınız	oku-ya-lar	oku-ma-ya-yım
AZE	gəl-im	gəl-ə-sən	gəl-sə	gəl-sək	gəl-sə-niz	gəl-ə-lər	gəl-mə-səm
	oxu-yam	oxu-ya-san	oxu-sa	oxu-saq	oxu-sa-nız	oxu-sa-lar	oxu-ma-yam
		Affir: ps-Stem-mAkçH		Neg: ps-Stem-mAkçH dääl		Interr: ps-Stem-mAkçH mH?	
TKM	men gel-	sen gel-	ol gel-mekçi	biz gel-mekçi	siz gel-mekçi	olar gel-	men gel-mekçi
	mekçi	mekçi				mekçi	dääl
	men okamakçı	sen okamakçı	ol oka-makçı	biz okamakçı	siz okamakçı	olar okamakçı	men oka-makçı dääl
		Affir: Stem-gAy-ps		Neg: Stem-mA-gAy-ps		Interr: Stem-e-ps kHrek mH	
UYG	kel-gey-men	kel-gey-sen	kel-gey	kel-gey-miz	kel-gey-siz (ler)	kel-gey	kel-mi-gey-men
	oqu-ğay-men	oqu-ğay-sen	Oqu-ğay	oqu-ğay-miz	oqı-ğay-siz (ler)	oqu-ğay	oqu-mi-ğay-men
KAZ (I)	kel-ğey-min	kel-ğey-siň	kel-ğey	kel-ğey-miz	kel-ğeysıňder	kel-ğey	kel-me-ğey-min
	oqı-ğay-mın	oqı-ğay-sıň	oqı-ğay	oqı-ğay-mız	oqı-ğaysıňdar	oqı-ğay	oqı-ma-ğay-mın
UZB	kel-gäy-män	kel-gäy-sän	kel-gäy	kel-gäy-miz	kel-gäy-siz	kel-gäy-(lär)	kel-mä-gäy-män
	oqı-gäy-män	oqı-gäy-sän	oqı-gäy	oqı-gäy-miz	oqı-gäy-siz	oqı-gäy-(lär)	oqı-mä-gäy-män
		Affir: Stem-mAk-ps		Neg: Stem-mAk-ps emes		Interr: Stem-mAk-ps mH	
KIR	kel-mek-min	kel- mek –siň	kel- mek	kel- mek-piz	kel-meksıňer	kel- iş-mek	kel-mek emesmin
	oqu-makmın	oqu- mak-sıň	oqu-mak	oqu- mak-pız	oqu-maksıňar	oqu- ş-mak	oqu-mak emesmin
		Affir: Stem-s-ps kile		Neg: Stem-me-s-ps kile		Interr: Stem-s-ps kile mH	
TAT	kil-äsem kilä	kil-äseň kilä	kil-äse kilä	kil-äse-biz kilä	kil-äse-giz kilä	kil-äse-leri kilä	kil-me-sem kilä
	ukı-ysım kilä	ukı-ysıň kilä	ukı-ysı kilä	ukı-äsa-bız kilä	ukı-ysı-gız kilä	ukı-sa-ları kilä	ukı-ma-sam kilä
		Affir: Stem-g(q)H-ps keldi, keledi		Neg: Stem-me-		Interr: Stem-g(q)Ay-ps mH	
KAZ (II)	Meniň kel-gim keledi	Seniň kel-giň keledi Sizdiň kel-giňiz keledi	Onıň kel-gi-si keledi	Bizdiň kel-gimiz keledi	Senderdiň kel-gi-leriň keledi Sizderdiň kel-gi-leriňiz keledi	Olardıň kelgi-ler-i keledi	Meniň kel-gi-m kel-me-y-di, Meniň kel-gi-m keledi me?

	Meniň okı-gım keledi	Seniň okı-gıň keledi Sizdiň okı-gıňız keledi	Onıň okı-gısı keledi	Bizdiň okı-gımız keledi	Senderdiň okı-gı-larıň keledi Sizderdiň okı-gı-larıňız keledi	Olardıň okıgı-lar-ı keledi	Meniň okı-gı-m kel-me-y-di, Meniň okı-gı-m keledi me?

4. Syntax

By definition, the general word order of Turkic languages is S-O-V (subject-object-verb), therefore the verb is usually at the end of the sentence. But Turkic languages are also very flexible, in other word they have free syntax. A general rule of Turkish word order is that the modifier precedes the modified:

- Adjective (used attributively) precedes noun;
- Adverb precedes verb;
- Object of postposition precedes postposition.

Some sentence example are given in Table-10. As seen in this table, Word order of Turkic languages are the same. [2], [8], [9], [10] , [12], [13]

Table-10: Two Sentences Example of Turkic Languages

TUR	Ağır	kazan	geç	kaynar
AZE	Ağır	qazan	gec	qaynayar
TKM	Agyr	gazan	giç	gaýnar
KAZ	Awur	qazan	keş	qaynaydı
KIR	Oor	kazan	keç	kaynayt
TAT	Avır	kazan	ozak	kaynıy
UYG	Eghir	qazan	waqche	qaynaydu
UZB	Teran	daryo	tinch	oqar

TUR	Dağ	dağa	kavuşmaz,	insan	insana	kavuşur
AZE	Dağ	dağa	Qovuşmaz,	insan	insana	qovuşar
TKM	Dag	daga	Duşmaz,	adama	adama	duşar
KAZ	Taw	tawğa	Qosılmas,	adam	adamğa	qosıladı
			Menen		adam	Adam
KIR	Too	too		adam		
TAT	Taw	tawga	körüşpöyt, Kilmäs,	adäm	adämgä	körüşöt Oçrar.
		tagh			Insan	tipishar
UYG	Tagh	bilen	tipishalmas,	insan	bilen	
UZB	Tog'ning	ko'rki	tosh bilan,	odamning	ko'rki	bosh bilan

Taw tawga kilmäs, mägär adäm adämgä oçrar.

Acknowledgement

I would like to express my special thanks of gratitude to L. Garaylı, A. Gatiatullin,

A. Anarbaeva, M. Orhun, L. Zhetkenbay, S. Hazratov who review their native language part of this paper.

References

1. Banguoğlu, T.; *Türkçenin Grameri*, TDK, 1986 2nd Print

2. Tantuğ, A. C.; *Akraba Ve Bitişken Diller Arasinda Bilgisayarli Çeviri İçin Karma Bir Model*, 2007, pHD thesis, İTÜ

3. Oztopçuu, K ; *A Comparison of Modern Azeri with Modern Turkish*, Berkeley/Ucla, 1993

4. Üşenmez, E. ; *Modern Özbek Türkçesi*, Akademik Kitaplar, 2012

5. Straughn, C. A.; *Evidentiality in Uzbek And Kazakh*, Chicago, Illinois, Dec. 2011

6. *Özbekçe Dil Bilgisi*, Vikikitap

7. Ata, A. Tulum, M. M.; *Uygur Türkçesi*, Anadolu Üniv. 2011

8. Orhun, M. ; *Uygurcadan Türkçeye Bilgisayarli Çeviri*, 2010, pHD thesis, İTÜ

9. Krippes, K. A. *Kazakh Grammatical Sketch With Affix List*, Hieroglyphic Press, Columbia, Maryland, 1993

10. Gökgöz, E., Kurt, A. Kara, M.; *Contemporary Turkish Dialects And LiteraturestwoLevel Qazan Tatar Morphology*, 1st Int. Conf. on Foreign Language Teaching And Applied Linguistics, May 5-7 2011 Sarajevo

УДК 81’32

COMBINED MORPHOLOGICAL AND SYNTACTIC DISAMBIGUATION FOR CROSSLINGUAL DEPENDENCY PARSYNG

Ageeva Ekaterina Higher School of Economics Russia E-mail: yekaterina.ageeva@gmail.com
Tyers, Francis UiT Norgga arktalas universitehta Norway
E-mail: francis.tyers@uit.no

Abstract

This paper describes a system for combined morphological disambiguation and dependency parsing and applies it to cross-lingual parsing of two under-resourced Turkic languages, Crimean Tatar and Tuvan. The system is based on finite-state morphological analysis followed by greedy transition-based dependency parsing. We show that it is possible to parse a related Turkic language using only a
Treebank designed with another Turkic language in mind.

Keywords: syntactic analysis, dependency grammar, machine learning, banks syntactic trees, cross-language analysis.

КОМБИНИРОВАННЫЕ МОРФОЛОГИЧЕСКИЕ И СИНТАКСИЧЕСКИЕ НЕОДНОЗНАЧНОСТИ ДЛЯ СИНТАКСИЧЕСКОГО АНАЛИЗА ЗАВИСИМОСТЕЙ КРОСС-ЯЗЫКОВ

Агеева Екатерина Высшая школа экономики, Россия, E-mail: yekaterina.ageeva@gmail.com
Тайерс Френсис Норвегия , E-mail: francis.tyers@uit.no

Аннотация

В данной статье описана система для комбинированной морфологической неоднозначности и синтаксического анализа зависимостей

и применяет его к кросс-язычной разборе двух стран с ограниченными ресурсами тюркских языков, крымскотатарских и тувинских. Система

основана на конечном состоянии морфологического анализа с последующими жадными в разборе перехода на основе зависимостей. Покажем, что можно разобрать, связанную с использованием тюркского

языка только Treebank разработан в виду с другим тюркскими языками.

Ключевые слова: синтактический анализ, грамматика зависимостей, машинное обучение, банки синтактических деревьев, межъязыковой анализ.

1 Introduction

Morphological and syntactic analysis are the stepping stones for more complex language processing applications, such as machine translation, information retrieval, question-answering, and many others. We also explore the applicability of the joint method to cross-lingual parsing. Cross-lingual techniques are applied to different tasks, such as sentiment analysis (Wan, 2009), word sense disambiguation (Lefever et al., 2010), and others. The principal idea behind crosslingual language processing is to apply the resources (e.g. corpora, treebanks, analysers) of one language to process a different language, which is usually underresourced. Although cross-lingual dependency parsing has been performed before,

by e.g. Xiao et al. (2014) and Tiedemann (2015), it did not benefit from combined

morphosyntactic disambiguation. This work presents a free/open-source tool for combined morphological and syntactic parsing, and uses it to experiment with applying a parsing and disambiguation model trained on a Kazakh treebank to

parse two related languages, Crimean Tatar and Tuvan.

	Treebank						Morphology	
	Train	Dev	Test	*P*	*M*	*D*	Coverage	Ambiguity
Kazakh	2,849	717	950	29	234	29	98.4	1.28
Tuvan	—	—	855	19	125	26	99.6	1.19
Crimean Tatar	—	—	639	19	103	26	99.5	1.77

Table1: Statisticsforthecorpora, *P* isthesetofpartofspeechtags, *M* isthesetofuniquemorphological analyses and *D* is the number of dependency relations used.

2 Related work

Joint syntactic and morphological parsing (also called joint disambiguation) has emerged as an effort to improve parsing accuracy for languages with rich morphology, for which the standard parsing techniques perform poorly. One of the first experiments discussing the benefits of joint processing was carried out by Tsarfaty (2006). Tsarfaty explored the effects of joint morphological segmentation and part-of-speech tagging on parsing quality for Hebrew: the model that performed segmentation and tagging jointly had an advantage over the pipeline approach. Cohen et al. (2007) make the next step in joint parsing and include syntactic relations into their model. They trained two analysers, the first of which includes segmentation and partof-speech tagging modules, and the second performs constituency parsing. The analysers are combined using the "product of experts" learning technique, which takes the product of independent probability functions to produce the final result. This work is also concerned with Modern Hebrew. Goldberg and Tsarfaty (2008) have developed the first model that incorporated morphology and syntax as a single classifier, as opposed to the two separate classifiers of Cohen et al. (2007) and achieve better results. Further experiments with joint morphological and syntactic disambiguation have explored its effects on parsing both morphologically rich languages and languages with high ambiguity of word forms, such as Chinese and English. Li et al. (2011) have developed several joint parsing models for Chinese, which incorporate different features and use various pruning strategies to reduce search space. These models have shown improvement over the pipeline models for Chinese. Similar work has been done by Bohnet and Nivre (2012), who also proposed to use the joint technique for dependency parsing, as opposed to constituency parsing in the works previous to this. Bohnet and Nivre develop a parser and experiment with Czech, German, English and Chinese, achieving state-of-the-art accuracy. Çetinoğlu et al. (2013) use this parser to process Turkish, and also report an improvement in parsing accuracy. Bohnet, Nivre, et al. (2013) further expand dependency parsing models by adding more sophisticated morphological information, and using word clusters to incorporate lexical information into the model. In addition, joint models may also deal with sub-word segments, and a number of works for Chinese word segmentation demonstrate improvement over the pipeline models (see e.g. Jiang et al. (2008), Kruengkrai et al. (2009), Sun (2011), and Zhang et al. (2008)).

3 Data sets and resources

Kazakh For training and testing we use the treebank developed by Tyers and Washington (2015). This consists of 402 sentences from different domains: learners' books, folk tales, legal texts and Wikipedia articles. The morphological analyser used was by Washington et al. (2014) and for pipeline disambiguation we used the hybrid tool developed by Assylbekov et al. (2016), consisting of approximately 150 hand-written rules in constraint grammar and a statistical model.

Language	Gold tags	Pipeline	Oracle
Kazakh	71.7	61.4	61.8
Tuvan	71.9	50.4	51.0
Crimean Tatar	79.0	64.0	73.8

Table 2: Reference results (labelled attachment score, LAS) for the three languages in question. Gold tags means that the input to the parser was the part-of-speech tag and morphological information from the test corpus; Oracle is the result of parsing all the possible paths and selecting the one with highest LAS; Pipeline is the result of applying the statistical disambiguator described in Assylbekov et al. (2016) trained on the Kazakh treebank. Note that the Oracle may be lower than using the Gold tags as the morphological analyser may not cover all forms or return all valid analyses.

Tuvan For testing we used 115 grammar-book sentences from Anderson et al. (1999) annotated for universal dependencies (Nivre et al., 2016) distributed with the morphological analyser described in Tyers, Washington, et al. (2016).

Crimean Tatar For testing we use a treebank consisting of 150 grammar-book sentences from Kavitskaya (2010) annotated for universal dependencies. The morphological analyser is available from the `apertium-crh` package [6] under development at the Apertium project.

4 Reference system

In order to assess the capabilities of a combined parsing model, we have performed experiments with a non-combined reference parser. This parser is purely syntactic, and the partof-speech and morphological information is pre-disambiguated. The non-combined parser is a basic greedy transition-based implementation as described by Kübler et al. (2009). We used the decision tree classifier and its internal tuning and cross-validation algorithms that have been implemented as part of the `scikit-learn` module for Python (Pedregosa et al., 2011)

In order to choose the best feature set for the model we performed experiments with different combinations of features, testing the performance of each set against the

[6] `https://svn.code.sf.net/p/apertium/svn/incubator/apertium-crh/`

development set for Kazakh. The combination of features which resulted in the highest labelled-attachment score was chosen. This was part-of-speech, lemma and morphological information for the word at the top of the stack and the first four words at the front of the buffer.

Using the model described above, we conducted three experiments to determine the boundaries of what the combined model can achieve given our data. Each experiment has been run on the Kazakh corpus. In addition, we performed two cross-lingual experiments: the Crimean Tatar and Tuvan test corpora were using the model trained on Kazakh data. The experimental results are summarised in Table 2.

Gold The model was given the unambiguous input of the gold part-of-speech and morphological analysis from the testing section of the corpus.[7] This gives an upper bound on performance of the pipeline model. If our disambiguator had 100% accuracy with respect to the corpus, this is the parsing performance we could expect to achieve.

Pipeline This model can be considered to be the *state of the art* for each language. The output of the morphological analyser is first disambiguated (by a hybrid disambiguator, see Assylbekov et al. (2016)) and the best output of that disambiguation is given to the parsing model.

Oracle With this model, each path from the lattice output (see Figure 1) by the morphological analyser is expanded and parsed. The resulting output is scored with labelled-attachment score, and for each sentence, the best score is taken. This can be considered to be the upper bound of performance for the combined model.

4.1 Formats and metrics

As the Kazakh treebank takes advantage of the new tokenisation standards in the CoNLL-U format,[3] and the parser only supports CoNLL-X, certain transformations were needed to perform the experiments. The corpus was flattened with conjoined tokens receiving a dummy surface form. The converted data is available alongside the original.

Furthermore it was necessary to come up with a new format for expressing ambiguous analyses in a format similar to the CoNLL series of formats. The format is identical to CoNLL-U, but allows for each ID to be repeated with a different analysis.

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	REL	DEPREL	MISC
1	Осында	осында	ADV	adv	_	_	_	_	_
1	Осында	осы	PRON	prn	_	_	_	_	_
2	орыс	орыс	NOUN	n	_	_	_	_	_
3	тілінде	тіл	NOUN	n	_	_	_	_	_
4	сөйлейтін	сөйле	VERB	n	_	_	_	_	_
4	сөйлейтін	сөйле	VERB	n	_	_	_	_	_
5	адам	адам	NOUN	n	_	_	_	_	_

[7] This is equivalent to taking the first six columns of CoNLL-U format and feeding them to the parser.

[3] http://universaldependencies.org/format.html

6-8	бар ма	_	_	_	_	_	_	_	_
6	бар ма	бар	ADJ	n	_	_	_	_	_
7	бар ма	е	VERB	cop	_	_	_	_	_
8	бар ма	ма	PART	qst	_	_	_	_	_
9	?	?	PUNCT	sent	_	_	_	_	_

5 Combined model

5.1 Preprocessing

As both Crimean Tatar and Tuvan lack an annotated corpus with which to train a part-of-speech tagger or morphological disambiguator, so the input to the parser is a lattice (see Figure 1) rep-

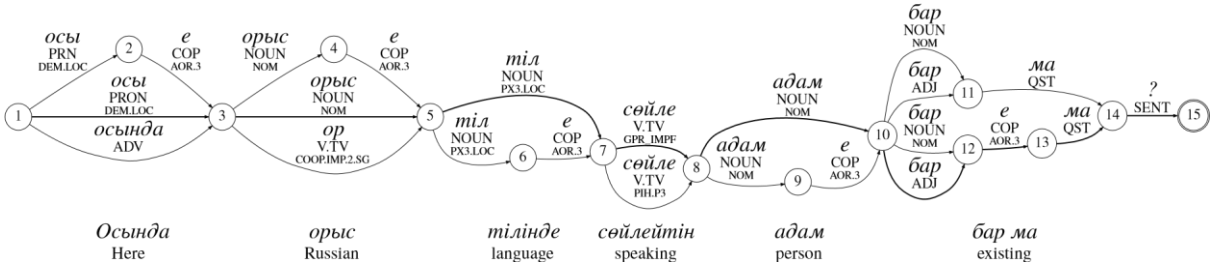


Figure 1: An example of ambiguous tokenisation for the sentence *Осында орыс тілінде сөйлейтін адамбарма?* "Is there a person here who speaks Russian?" in Kazakh. The tokenisation path expressed in the treebank is in bold.

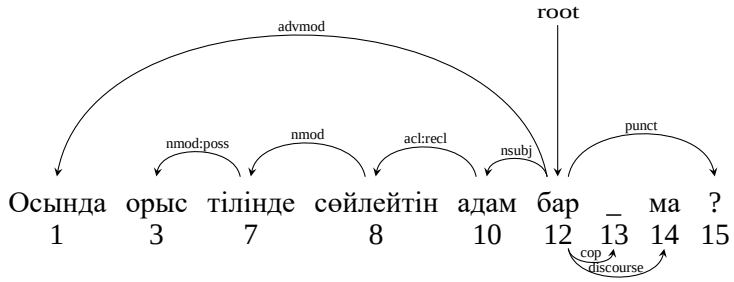


Figure 2: The dependency tree for the sentence given in Figure 1.

resenting the ambiguous output of the morphological analysers. The morphological analysers also perform tokenisation on the basis of a left-to-right longest match algorithm described in Garrido-Alenda et al. (2002). The mapping between space-separated 'surface forms' and syntactic tokens is non-trivial. In some cases a single surface token is equivalent to a single syntactic token (as in *сөйлейтін* 'speaking' in Figure 1), in other cases, multiple surface tokens may result in multiple syntactic tokens (as in *бар ма* 'is existing?' in Figure 1). There are a number of factors involved in determining if multiple surface tokens should be treated as a single token, including: does the token undergo any morphophonological processes? (e.g. the question suffix *ма* may also appear as *ме*, *ба* or *бе* depending on the ending of the previous token), and does an extra syntactic token (e.g. the zero-copula in the third person aorist) need to be introduced?

5.2 Morphological disambiguation

Having performed the baseline experiments, we have set out to develop the combined syntactic and morphological parser. We took our syntactic parser as a base and added the capability to do morphological disambiguation. We treat morphological disambiguation as a classification task, similar to determining the best next transition in the dependency parser. In this case, the items to classify also are configurations, and the label assigned to each is a concatenation of the part-of-speech and the morphological tags of the first word in the buffer. In reality, the entity classified is the first word in the buffer, but because we use the features from other parts of the configuration, technically, we classify configurations. We have chosen to perform disambiguation of the first word in the buffer. On one hand, it is best to disambiguate as late as possible, so that the syntactic parser can benefit from additional information for as long as

Language	Pipeline	Combined	Oracle
Kazakh	61.4	63.2	61.8
Tuvan	50.4	58.4	51.0
Crimean Tatar	64.0	62.4	73.8

Table 3: Results

possible. On the other hand, all transitions in the syntactic parser assume that the tokens are already disambiguated, and that the words that may participate in transitions are the first word in the buffer and the first word on the stack. Therefore, disambiguation happens just before the word can potentially participate in any transitions, but not earlier. We check if the buffer front needs disambiguation before predicting every following transition. We also accommodate for ambiguous tokenisation when the analyser may need to split a single 'surface form' into several structural tokens, which later form the dependency relations (for example, tokens *бар ма* in Figure 1). In this case, these tokens are *unwrapped*, and both the buffer and the underlying sentence shift to make place for the extra tokens.

The features we use for morphological disambiguation are the same as for dependency parsing, with a minor change. Because we only disambiguate the token when it reaches the buffer front, the features concerning other items in the buffer were modified to work with ambiguous tokens in the following way:

form: returns the form of the first analysis, or the unifying surface form for several syntactic tokens, if there are multiple;

part-of-speech: returns the ambiguity class of the token, i.e. a concatenation of all distinct part-of-speech tags seen in the analyses for this token. For the token *орыс* this would be noun|verb.

morphological features: returns nothing if the token is ambiguous.

The dependency parser drives the process: it moves the state from one configuration to another by determining the next transition, until the buffer is empty. The classifier for morphological disambiguation works as a supplementary tool at each step, selecting the best analysis for the buffer front as described in the section above.

6 Cross-lingual parsing

It was necessary to make a number of small changes to the annotation scheme for the Crimean Tatar and Tuvan treebanks as the annotation conventions for a number of phenomena are not yet completely standardised. The universal dependency relation `iobj` 'second obligatory argument' (typically indirect object) is called `arg` in the Tuvan data, and `nmod:rcp` in the Crimean Tatar data. These were standardised to `iobj`. The Crimean data used `acl:relcl` for relative clauses, where the Tuvan and Kazakh data used `acl`. We standardised on `acl`. Finally, both Crimean Tatar and Tuvan distinguished clausal subjects,

`csubj` from nominal subjects `nsubj`, a distinction which is currently not made in the Kazakh treebank, so we collapsed both of these labels into a single `subj` label.

There were also a number of idiosyncracies left in place, for example `nsubj:caus` for causative subjects if verbs in Crimean Tatar. There were no examples of this phenomenon in the Tuvan or Kazakh data.

7 Discussion

7.1 Error analysis

We analysed the errors made by the dependency parser and the morphological disambiguation component. This section reports the common error patterns we discovered. We have observed, although to a lesser extent than in the pipeline models, the error accumulation effect: if the morphological classifier selected an incorrect analysis for a given token, it will very likely enter an incorrect dependency relationship, which will at least partly affect the parse tree. The errors at this point may be incorrect tokenisation (cases when one surface form is analysed as several lemmas), incorrect part-of-speech label, or incorrect morphological analysis. Predictably, the parser also makes errors of its own, assigning incorrect head and/or dependency labels when the morphology has been determined correctly. We should note that the mistakes are not languagespecific, and repeat across different corpora — which is not very surprising, provided we used the same model to parse them. The first, and perhaps the most expected category of errors deals with part of speech ambiguity. Words like *bu* / *бо* ‘this’ and o ‘that’ can be classified as determiners, demonstrative pronouns, or personal pronouns (*o* as ‘that’ vs ‘he’); *bir* ‘one’ can be a numeral or an indefinite determiner, the distinctions not always correctly made by our model. It also tends to select the substantive interpretation over an attribute adjective, an error which has surfaced in the Tuvan and Kazakh corpora. Some of part of speech errors are common; others are made due to lack of training data. For example, the Kazakh word *көп* ‘many’ was misclassified once as an adjective as opposed to a determiner, and once correctly classified as an adverb, but it never occurred in the training corpus. In cases when part of speech has been determined correctly, the morphological information may have not been. The most common source of such errors is the distinction between verb forms. There are cases when passive transitive verbs have been classified as intransitive, and when the tag for a participle form has been assigned instead of (the correct) tag for verbal adverb form. These particular distinctions, however, have been up to debate in the annotation guidelines of the corpus, and can also depend on the interpretation. Other morphological errors are more straightforward and reveal that the parser may be rather ignorant about the surrounding words. For example, the verb *бастады* ‘started’ was classified twice as plural rather than singular, even though it had a correctly determined singular subject. In general, having more context may improve parsing accuracy, although at a cost of considering more possibilities at each step. A common error that speaks in favour of this is finding multiple subjects in rather simple sentences — a pattern that is infrequent in training data, and that could have been better learned. Consider a the Crimean Tatar sentence in Figure 3 in, where three words – my brother, every, and day – have been tagged as subjects.

We suspect that in such cases the parser (especially having made an incorrect part of speech decision) assigns the relation that is most likely given a local context. It does not

consider the likelihood of a 6-word sentence having 3 subjects, knowledge of which would significantly improve its performance.

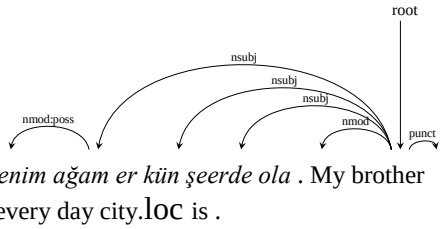


Figure 3: Multiple subjects in a simple sentence

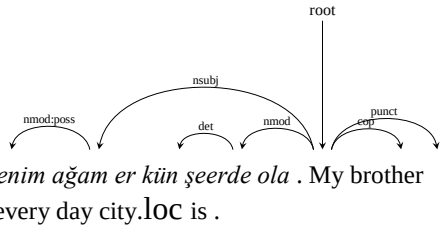


Figure 4: Treebank parse for Figure 3

Finally, a small fraction of errors come from rare categories, which have not been encountered often enough during training. For example, the only instance of a 'vocative' relation in the Kazakh testing corpus has not been correctly determined, but it only occurred twice in the training corpus. Another example is the dependency label 'parataxis', which signifies a relation between the main verb and the clause after a colon or a semicolon. This relation has no overt signs of coordination or subordination, and is therefore difficult to learn, especially given the 10 instances (0.3%) in the training corpus.

7.2 Future work

There are a number of avenues for future work. One aspect we would like to improve concerns the output of the morphological analyser. At the moment the lattice we give to the parser is unweighted, that is all of the analyses are considered equally probable. However, this is unlikely to be the case. A noun reading for the word *орыс* 'Russian' is far more likely than the cooperative imperative of the verb *ор* 'reap with me!'. It is possible to apply weights to a finite-state transducer either using corpora, or linguistic knowledge, and this is something we would like to incorporate into the model. On a similar vein, we would like to experiment with adding rulebased constraints. Given a small or non-existent treebank (in the case of cross-lingual parsing), is it possible to write simple rule-based constraints which could be incorporated into the model ? These constraints could be of the type "A clause may have at most one subject", "The copula verb cannot be the root of a sentence" or "A personal pronoun in nominative is never a nominal modifier". In considering the model, we would like to implement a real joint model, where we have a single classifier which predicts both the best transition and morphological disambiguation in a single step. Looking at adding word-embeddings (Mikolov et al., 2013), which can be calculated from inexpensive monolingual corpora would also be an interesting avenue for future work. Finally, we would like to apply this work to other Turkic languages, and possibly use the parser to bootstrap treebanks for other Kypchak languages.

8 Concluding remarks

This work has been concerned with cross-lingual dependency parsing enhanced by morphological disambiguation. We have developed a combined syntactic and morphological parser, which is transition-based and operates with two independent classifiers. We have shown that it is possible to use the classifiers trained on Kazakh data to parse corpora in Crimean Tatar and Tuvan. After adding morphological disambiguation to the dependency parsing process, we have improved the parsing quality for Kazakh and Tuvan over the baseline scores. All of the

code and data used in the experiments can be found on GitHub.[8]

Acknowledgements

Removed for review

References

Anderson, G. and Harrison, K. D. (1999). Tyvan. Lincom Europa.

Assylbekov, Z., Washintgon, J. N., Tyers, F. M., Nurkas, A., Sundetova, A., Karibayeva, A., Abduali,B.,andAmirova,D.(2016).Afree/open-sourcehybridmorphologicaldisambiguation tool for Kazakh". 1st International Workshop on Turkic Computational Linguistics. In: *Proceedings of the 1st International Workshop on Turkic Computational Linguistics*.

Bohnet, B. and Nivre, J. (2012). A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing. In: *Proceedings of EMNLP*.

Bohnet, B., Nivre, J., Boguslavsky, I. M., Farkas, R., Ginter, F., and Hajič, J. (2013). Joint morphological and syntactic analysis for richly inflected languages. In: *Transactions of the Association for Computational Linguistics* 1, pp. 429–440.

Çetinoğlu, Ö. and Kuhn, J. (2013). Towards joint morphological analysis and dependency parsing of Turkish. In: *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*, pp. 23–32.

Cohen, S. B. and Smith, N. A. (2007). Joint morphological and syntactic disambiguation. In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 208–217.

Garrido-Alenda, A., Forcada, M. L., and Carrasco, R. C. (2002). Incremental construction and maintenance of morphological analysers based on augmented letter transducers. In: *Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation*, pp. 53–62.

Jiang, W., Huang, L., Liu, Q., and Lü, Y. (2008). A cascaded linear model for joint Chinese word segmentation and part-of-speech tagging. In: *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*.

Kavitskaya, D. (2010). Crimean Tatar. Lincom Europa.

Kruengkrai, C., Uchimoto, K., Kazama, J., Wang, Y., Torisawa, K., and Isahara, H. (2009). An error-driven word-character hybrid model for joint Chinese word segmentation and POS tagging. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Vol. 1, pp. 513–521.

Kübler, S., MacDonald, R., and Nivre, J. (2009). Dependency Parsing. Morgan & Claypool.

Lefever, E. and Hoste, V. (2010). Semeval-2010 task 3: Cross-lingual word sense disambiguation. In: *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 15– 20.

[8] `http://github.com/Sereni/joint-parser`

УДК 81 `32

NAMED ENTITY RECOGNITION FOR KAZAKH USING CONDITIONAL RANDOM FIELDS

Tolegen Gulmira, Kazakhstan, 100000, c. Astana, *National Laboratory Astana* , gtolegen13@fudan.edu.cn
Toleu Alymzhan, Kazakhstan, 100000, c. Astana, *National Laboratory Astana* , atoleu13@fudan.edu.cn
Xiaoqing Zheng, China, 200433, c. Shanghai, *Fudan University*, zhengxq@fudan.edu.cn

We addressed the Named Entity Recognition (NER) problem for the Kazakh language by using conditional random fields. Kazakh is a typical agglutinative language in which thousands of words could be generated by adding prefixes and suffixes to the same root, which arises a serious data sparsity problem for many NLP tasks. To reduce the data sparsity problem, a necessary preprocessing step is to split the words into their roots and morphemes by morphological analysis. In this study, we designed a CRF-based NER system for Kazakh, which leveraged the features derived from the results of a new-developed morphological analyzer, and found that the performance can be boosted by introducing such derived features. Moreover, we assembled a NER corpus which was manually annotated with location, organization and person names.

Keywords: Kazakh language, agglutinative language, named entity, NER, CRF

ИЗВЛЕЧЕНИЕ ИМЕНОВАННЫХ СУЩНОСТЕЙ ИЗ ТЕКСТА НА КАЗАХСКОМ ЯЗЫКЕ С ИСПОЛЬЗОВАНИЕМ УСЛОВНЫХ СЛУЧАЙНЫХ ПОЛЕЙ

Толеген Гулмира, Казахстан , 100000, г. Астана, *Национальная Лаборатория* , gtolegen13@fudan.edu.cn
Толеу Алымжан , Казахстан , 100000, г. Астана, *Национальная Лаборатория* , atoleu13@fudan.edu.cn
Xiaoqing Zheng , КНР , 200433, г. Шанхай, *Фудан Университет* , zhengxq@fudan.edu.cn

Мы обратились к проблеме извлечения именованных сущностей (Named Entity Recognition, NER) для казахского языка, используя условные случайные поля (Conditional Random Fields, CRF). Казахский – типичный агглютинативный язык, в котором тысячи слов могли бы быть получены путем

добавления префиксов и суффиксов к тому же корню, что создает серьезную проблему разреженности данных для решения многих задач NLP. Чтобы уменьшить проблему разреженности данных, необходим шаг предварительной обработки для разделения слов в их корни и морфемы путем морфологического анализа. В данном исследовании мы разработали CRF-систему NER для казахского языка, которая использовала возможности, полученные из результатов новоразвитого морфологического анализатора, и обнаружили, что производительность может быть повышена путем внедрения таких производных функций. Кроме того, мы собрали NER корпус, который был вручную аннотированный с названиями мест, организаций и лиц.

Ключевые слова: Казахский язык, агглютинативный язык, именованные сущности, NER, CRF

1 Introduction

Named Entity Recognition is an important task in many Natural Language Processing (NLP) applications nowadays. State-of-the-art NER systems have been produced for several languages, but despite these recent improvements, developing an NER system for Kazakh is still a challenging task due to the structure of the language. Kazakh is the official state language of Kazakhstan. It is an agglutinative and highly inflected language. To illustrate this, consider the following English phrase "*people who live in the city*" which can be translated into Kazakh with only one word "*қаладағылардың*" which can decompose into the root and additional suffixes: "*қала+да+ғы+лар+дың*".

This productive nature of Kazakh results in a single root which may produce hundreds or thousands of word forms, thus causing the data sparsity problem. In order to prevent this behavior in our NER system, we build a Finite State Transducer (FST)-based morphological analyzer, which is able to decompose complex words into their constituent root and morphemes. We also developed a perceptron algorithm based morphological disambiguation system.

In this paper, a CRF-based system used to extract named entities from Kazakh text is presented. In order to obtain accurate generalization, we used several syntactic and semantic features of the text, including: more fine-grained morphological features and word type information. The morphological features are important for Kazakh; these features are extracted by morphological analyzer, and effectively used in this study to increase the recognition performance. We evaluated our models on Kazakh NER corpus (KNC) that we have annotated with location, person and organization names.

The rest of the paper is organized as follows: Section 2 gives brief information about FST-based morphological analyzer and describes features used in this work. Section 3 gives detailed information about the corpus of named entity and reports

the results of experiments. Section 4 reviews the existing work. Section 5 concludes with possible future work.

2 The CRF-based NER system

The CRF-based state-of-art NER systems for agglutinative languages often use a large feature set to improve system performance (Yeniterzi, 2011; Şeker and Eryiğit, 2012). As described before, Kazakh words may contain different morphemes to determine their meaning, and these morphemes can be viewed as a set of morphological features. In order to extract such features and prevent data sparseness problems, developing a morphological analyzer and disambiguator will be a crucial step for Kazakh NER.

2.1 Morphological analysis and disambiguation

Morphological analysis is a problem of breaking word such as "*бөлмесін*" (*don't let him split sth.*) into the stem and morphemes, *бөл* (*split*) and *-ме - сін*. These features specify the additional information about the stem. There are mainly two approaches, FST-based and data-driven approaches. Finite state techniques (Oflazer and Güzey, 1994) are the most suitable technique to represent the morphosyntactic rule of agglutinative languages.

In order to develop a FST-based morphological analyzer, we need three components: 1) a lexicon of roots annotated with some information (part of speech etc.), to determine which morphological rules apply to them, 2) a morphotactic component that describes the word formation by specifying the ordering of morphemes, and 3) Morphophonemics rules description.

For the lexicon list, we collected a new lexicon of 151,463 roots. To compile this lexicon and to ensure the correct spelling of the words, we manually checked and annotated with POS tags. We have considered almost all inflectional suffixes of nouns, verbs, adjectives, pronouns, adverbs, e.g. plural, case, possession, predication, tense, mood etc. We edited 3 ,039 word-formation rules and developed an FST-based morphological analyzer for Kazakh. The morphological analyzer achieved 95% coverage on news corpus (800K words), and 91% on Kazakh 's Wikipedia (10M words).

In order to select the most probable analysis of the words, depending on the context, we developed a perceptron algorithms (Collins, 2002; Sak et al., 2008) based morphological disambiguator and manually disambiguated a corpus (15K words) for model training, and the best model achieved 90.54 % accuracy on the test data. At the next stage, we use the information coming from the raw data and the disambiguated morphological analysis to extract named entity features.

2.2 Feature categories

We generalize named entities (NEs) by using a set of features that are capable of describing various properties of the text. In this study, these features can be grouped into two categories: morphological and word type information. In morphological features, since the part-of-speech of NEs will always be a noun, instead of using morphological features as sequence morpheme, we extracted more useful morpheme tags that attached to the noun from inflectional and derivational morphemes, and used theses as atomic units, and then each feature can be combined with other features effectively.

Morphological features:

- **Root Feature (RF)**: the root of word as a feature. Although the lexicon of morphological analyzer may not contain all the proper nouns and cannot analyze some proper nouns, but the root of the surrounding words of the proper nouns has an influence on the entity recognition.
- **Part of speech (POS)**: the Part-of-speech tag of the root as a feature.
- **Inflectional suffixes (NC, PL, PS, and PR)**: these four features that are extracted from all inflectional suffixes, the feature *NC* includes: Nominative, Genitive, Locative, Accusative, Dative, Ablative, and Instrumental suffixes; PL - plural; *PS* - possessive; *PR* - predication.
- **Derivational suffixes (SP, SN)** : these two features extracted from derivational suffixes.
- **Proper Noun (NP)**: this feature is stated as "1" when the selected morphological analysis includes tag "np".
- **Name suffixes (NS)**: some suffixes most used in the Kazakh surname formations such as (+ *ев, +ов, +ин, +ева, +ова, +ина, +ұлы, + қызы*).

Word type features:

- **Latin words (LW)**: Latin spelling words.
- **Acronym (AC)**: this should help to identify organizations and persons abbreviated as acronyms.
- **Case Feature (CF)**: the information about lowercase and uppercase letters used in the current word.
- **Start of the Sentence (SS)**: this feature indicates whether or not the current token represents the beginning of a sentence.

In this work, we provided atomic features within a window of {-4, 4}, and the feature template are designed by using wrapper methods (Kohavi, 1997). NER is typically treated as a sequence labeling task. We utilized CRF (Lafferty et al., 2001), which provides advantages over other statistical approaches such as the Hidden Markov Model and enables the use of any number of features.

2.3 The CRF model

Let $X = \{x_1, x_2, \ldots, x_n\}$ be the sequence of words in sentence, and $Y = \{y_1, y_2, \ldots, y_n\}$ the hidden labels. A linear chain Conditional Random Field defines a conditional probability

$$P(Y \mid X) = \frac{1}{Z} exp\left(\sum_{i}^{N}\sum_{j}^{K} \lambda_j f_j(y_{i-1}, y_i, X, i)\right) \quad (1)$$

where λ_j is the weight for feature f_j, and must be learned, the scalar Z is the normalization factor. For NER task, each y_i is named entity label and f_j is feature function which produces a real binary value. Training involves finding the λ_j and maximize the conditional log-likelihood of the training data $\mathcal{D}$

$$\sum_{(X,Y) \in \mathcal{D}} \log p(Y|X, \lambda) \quad (2)$$

In this work, we used CRF++ [9], which is an open source CRF sequence labeling toolkit.

3 Experiments

We have conducted several sets of experiments to explore the effectiveness of features on NER task for Kazakh. We used the CRF and examined the cumulative contribution of the features. For evaluation, we use the Kazakh NER corpus (Section 3.1). This corpus is divided into training (80%), development (10%) and test (10%) set. The development set was used for choosing hyper- parameters and model selection. We adopted IOB tagging scheme (Tjong Kim Sang, 2002) for all experiments. For evaluation, we used the *conlleval*[10] evaluation script to report the F1-score, precision and recall values.

3.1 Corpus Construction

A major obstacle to Kazakh NER is the scarcity of publicly available annotated corpora. We created a corpus by manually annotating the text of 2500 articles from general news media[11]. These articles were randomly selected from all articles. Only body text was extracted from the chosen articles for inclusion in the corpus. The annotations have been executed manually by native speakers of

[9] CRF++: Yet Another CRF toolkit.
[10] www.cnts.ua.ac.be/conll2000/chunking/conlleval.txt
[11] Available at: http://www.inform.kz

Kazakh. In order to assist the annotation process and to improve the correctness of the annotation, we have developed a web-based annotation tool. We followed the MUC-7 NE task definition (Chinchor, 1998) as a guideline for annotations.

The current corpus was annotated with person, location and organization names and the annotations were double-checked. We have measured the inter- annotator agreement (IAA) that was calculated using Fleiss' kappa. We randomly sampled 500 articles for this purpose. Each article was annotated by two annotators. The IAA is 0.93 without following guidelines or discussing difficult cases with each other. After discussions, the IAA achieved over 0.98.

Web text often contains errors. In this corpus, we did not correct grammatical and typographical errors, as we wished that the corpus remains as similar as possible to the source text. The final corpus was created by correcting annotation mistakes. This corpus consists of 18,054 sentences and 270,306 words with 4 ,292 person names, 7,391 location names, and 2,560 organization names.

3.2 Experimental Result

In CRF model, we adopted wrapper methods (Kohavi, 1997) to feature selection and tuned the feature template on the development set [12]. After find the best feature template, we grouped these features combinations into different classes, and each class is related to each feature category as described in Section 2.2.

In order to explore the contribution of each feature, we first evaluated the baseline (Word) performance of a CRF model, in which the tokens are only used in their surface form, and then added each feature combination to it. The results are evaluated with respect to the CoNLL metric and shown in Table 1. A plus (+) sign before the feature name indicates that these feature combinations[13] are added on top of the rest with suitable feature templates.

The system achieved a 69.91 % F1 using only word surface form. The F1 is improved by +4.57% when using the root feature *(+RF)* with Word. This result indicates that the root of the surrounding words of the proper nouns have an enormously positive impact on system performance. The F1 is improved by+6.34% when using *+POS* and improved by +3.51% when using morphological features such as *+NC, +PL, +PS, +PR*, which are extracted from all inflectional suffixes. As we can see, the inflectional suffix features not bring a significant improvement for organization names due to the complex of structure organization names, such as long NEs, mixed with other language and abbreviation etc. Then the feature *(+PS)* may hurt system performance. The morphological features such as *+SP, +SN* are extracted from derivational suffixes, these features also improves (+0.54%) system performance. These results show that using

[12] These experiments are not added to the paper due to the space constraints.

[13] These feature combinations are selected by wrapper methods and related to each feature category.

morphological features can significantly improves the NER from Kazakh texts due to the agglutinative nature of the language. This also indicates there is still room for improvement using more fine grained usage of these inflectional and derivational suffixes. Using name suffixes *(+NS)*, the person F1 improved by +2.94%. As shown, the feature *(+LW)* may hurt system performance (-0.31%). The F1 improved by +0.18 % using *+AC*. Since only the proper nouns and the initial words of the sentences start with a capital letter, the case features *(+CF)* improved the system significantly for all labels.

	Development set				**Test set**			
Feature	**LOC**	**ORG**	**PER**	**Overall**	**LOC**	**ORG**	**PER**	**Overall**
Word	77.86	61.26	65.19	71.73	77.56	66.67	57.45	69.91
+RF	85.26	66.34	69.21	77.95	82.97	69.52	61.19	74.48
+POS	88.03	70.26	74.64	**81.57**	87.99	76.79	69.65	**80.82**
+NC	88.19	73.10	78.79	83.16	88.58	79.21	74.36	82.81
+PL	88.47	73.02	79.95	83.58	89.55	79.65	76.92	**84.10**
+PS	88.93	74.77	79.51	**83.99**	89.75	77.66	76.62	**83.76**
+PR	89.01	73.42	80.32	84.03	89.94	77.75	78.21	84.33
+SP	89.79	74.89	80.64	84.79	90.47	78.71	77.48	84.57
+SN	89.81	74.67	81.06	84.87	90.84	77.59	78.52	84.87
+NP	90.77	74.27	83.03	85.93	90.53	77.49	80.66	85.38
+NS	90.65	74.94	**84.11**	86.26	90.85	77.42	**83.60**	86.37
+LW	90.83	74.89	84.55	86.46	90.35	77.16	83.60	86.06
+AC	90.70	75.16	84.86	86.51	90.64	77.49	83.46	86.24
+CF	93.15	78.49	91.89	**90.46**	91.45	82.38	89.83	**89.43**
+SS	93.15	78.59	91.91	**90.47**	**91.71**	**83.40**	**90.06**	**89.81**

Table 1 F1-score of the CRF when each feature is added cumulatively.

4 Related Work

There are large number of studies have been performed on NER for many languages. Here we review the literatures most relevant to this work.

For Turkish, (Tatar and Cicekli, 2011) proposed an automatic rule learning method for Turkish and achieved an averaged F1-score of 91.08% on the data-set, the experiment result show that morphological features can significantly improve the NER performance. (Yeniterzi, 2011) obtained an F1-score performance of 88.94% by using CRF and exploiting the effect of morphology used inflectional features as tokens. In the same direction (Şeker and Eryi ğit, 2012) proposed a successful CRF model for Turkish NER and extracted Proper Noun and Noun Case two features from all inflection suffixes, they report the result (89.55% in CoNLL metric) without using gazetteers on general news text. Few papers have been published in relation to Kazakh NER and this is one of the first systems to perform NER for Kazakh.

5 Conclusion

In this study, we have developed a CRF-based NER system for Kazakh language. Through a set of extensive experiments, the features of the CRF model have been carefully optimized for Kazakh NER. The experimental result shows that the features derived from the results of morphological analysis significantly improve the system's performance (from 69.91% to 89.81% in F1) by alleviating the data sparsity problem brought by the properties of agglutinative languages. Moreover, we created and manually annotated a Kazakh NER corpus (KNC). In the future, we plan to use deep learning approaches for Kazakh NER task.

Acknowledgments

The authors would like to thank Akmaral T., Olzhas S., Nazigul M., Galymzhan T., Duman S. for their help with data annotation. The authors would also like to thank Kakesh Kaiyrzhan for providing Kazakh language resources.

References

1. Erik F. Tjong Kim Sang. 2002. Introduction to the CoNLL-2002 Shared Task: Language-Independent Named Entity Recognition. In Proceedings of CoNLL-2002, pages 155-158.
2. John D. Lafferty, Andrew McCallum, and Fernando C.N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, pages 282-289.
3. J.L. Fleiss et al. 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin, pages 378-382.
4. Kessikbayeva Gulshat and Cicekli Ilyas. 2014. Rule Based Morphological Analyzer of Kazakh Language. In Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM, pages 46-54.
5. Michael Collins. 2002. Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms. In Proceedings of EMNLP 2002.
6. N. Chinchor. 1998. MUC-7 named entity task definition, version 3.5 In Proceedings of the Seventh Message Understanding Conference.
7. Oflazer K., Güzey C. 1994. Spelling correction in agglutinative languages. In Proceedings of the Fourth Conference on Applied Natural Language Processing, ANLP '94, pages 194-195.
8. Rabiner Lawremce R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. In Proceedings of the IEEE, 1989, pages 257-286.

9. Ron Kohavi and George H. John. 1997. Wrappers for Feature Subset Selection. Artificial Intelligence Archive, 97:273C324, 1997.
10. Seker Gö han Akın and Eryiğit Gülsen. Initial Explorations on using CRFs for Turkish Named Entity Recognition. In: Proceedings of COLING. 2012 , pp. 2459–2474.
11. Serhan Tatar and Ilyas Cicekli. Automatic rule learning exploiting morphological features for named entity recognition in Turkish. In: Proceedings of the 28th International Symposium on Computer and Information Sciences. 2013 , pp. 137 –151.
12. Sak, H., Güngör, T., Saraçlar, M. A stochastic finite-state morphological parser for Turkish. In: Proceedings of the ACL-IJCNLP. Conference Short Papers, 2009, pp. 273–276.
13. Tür, G., Hakkani- Tu¨r, D., and Oflazer, K. 2003. A statistical information extraction system for turkish. Natural Language Engineering, 9:181C210.
14. Yeniterzi Reyyan. 2011. Exploiting Morphology in Turkish Named Entity Recognition System. In Proceedings of the ACL 2011 Student Session, pages 105–110.

УДК 519.642.7:681.5.015.24

СИНГУЛЯРНЫЕ ВОЗМУЩЕНИЯ В ЛИНЕЙНОЙ ЗАДАЧЕ УПРАВЛЕНИЯ С МИНИМАЛЬНОЙ ЭНЕРГИЕЙ

Иманалиев З.К., *к.ф.-м.н., проф. каф. «Прикладная математика», КГТУ*
Баракова Ж.Т., *к.т.н., доц., заф. каф. «Информационные системы и технологии в телекоммуникациях», КГТУ, Janna05_05@mail.ru*
Кадыров Ч.А., *к.т.н, доц., декан фак. «Энергетический», КГТУ*

В статье рассмотрен способ решения задачи оптимального управления разнотемповыми системами при минимуме функционала типа нормы, который оценивает энергии управляющего воздействия. Рассмотренный способ основан на совместного использования идеи проблемы моментов и метода разделения движений. Исходная задача разделяется на две задачи оптимального управления, которые имеют меньшие размерности. Такое разделение переменных сокращает объём вычислительных процедур при решении конкретных практических задач.

Ключевые слова: сингулярные возмущения, разнотемповые системы, функционал, малый параметр, проблемы моментов, разделение движений, медленные и быстрые движения, управление, аналитическая форма.

SINGULAR PERTURBATIONS IN LINEAR PROBLEM OF CONTROL WITH MINIMUM ENERGY

Imanaliev Z.K., Barakova Z.T., Kadyrov Ch.A.

In the article describes the method of solving the problem of optimal control multirate systems with a minimum of functional type of rules which assesses the energy control action. The considered method is based on the idea of sharing the moment problem and the method of separation of motions. The original problem is divided into two optimal control problems which have smaller dimensions. This separation of variables reduces the amount of computational procedures for solving specific practical problems.

Keywords: singular perturbations, multirate systems, functional, small parameter, the moment problem, separation movements, slow and fast motions, control, analytical form.

ВВЕДЕНИЕ. Решение задач управления движениями сложных систем и оптимизации встречает существенные аналитические и вычислительные трудности, которые связаны с необходимостью решения краевых задач для многомерных систем дифференциальных уравнений (иногда с недостаточными краевыми условиями), а также с практической реализацией «точных» законов управления в виде алгоритмов при помощи вычислительных средств в реальном масштабе времени. Поэтому оказывается актуальной разработка эффективных приближенных аналитических и полуаналитических методов, допускающих существенное упрощение решения задачи.

Разработка таких методов как правило, основывается на сочетании методов теории возмущений с методами оптимального управления (принцип максимума, динамическое программирование и т.д.).

В данной работе предложен новый способ решения задачи оптимального управления разнотемповыми системами при минимуме функционала типа нормы, который оценивает энергии управляющего воздействия. Предложенный способ основан на совместного использования идеи проблемы моментов и метода разделения движений.

Рассмотрим задачу оптимального управления, включающую условие минимума функционала

$$J(u) = \int_{t_0}^{t_1} u(t)u(t)dt \qquad (1)$$

на траекториях системы $\dot{y} = A(t)y + B(t)u,$ (2)

$$y(t_0) = y^0,\ y(t_1) = y^1 \qquad (3)$$

$$\dot{x} = A_1(t)x + A_2(t)z + B_1(t)u$$

$$\varepsilon\dot{z} = A_3(t)x + A_4(t)z + B_2(t)u,$$

где $y = (x, z)$, $A(t) = \begin{pmatrix} A_1(t) & A_2(t) \\ A_3(t) & A_4(t) \end{pmatrix}$, $B(t) = \begin{pmatrix} B_1(t) \\ B_2(t) \end{pmatrix}$,

ε- малый положительный параметр, $x \in R^n$, $z \in R^m$, $u \in R^r$.

Считается, что система (2) вполне управляема. Функционал (1) можно рассматривать как квадрат нормы функции u(t) в пространстве $L_2[t_0, t_1]$. Так как норма $\|u\|_{L^2[t^0,t^1]}$ в $L_2[t_0, t_1]$.

достигает своего минимума со своим квадратом, то нам нужно выбрать среди допустимых решений задачи об управлении (задача о переводе системы (2) из заданной начальной точки (t_0, y^0) в конечную точку (t_1, y^1) такое решение, которое имеет минимальную норму в $L_2[t_0, t_1]$. Такое решение (с минимальной нормой) существует, если компоненты импульсной переходной вектор функции линейно независимы [1]. Это условие выполняется, если система вполне управляема [1], а по предположению система (2) вполне управляема.

Как показано в [6], что систему (2) можно заменить эквивалентной системой, у которой разделены медленные и быстрые составляющие вектора состояния:

$$\dot{\tilde{x}} = \tilde{A}_1(t)\tilde{x} + \tilde{B}_1(t)u, \qquad (4)$$

$$\varepsilon\dot{\tilde{z}} = \tilde{A}_4(t)\tilde{z} + \tilde{B}_2(t)u,$$

где $\tilde{x} = x - Nz$, $\tilde{z} = z - Hx$,

$\tilde{A}_1 = A_1 - A_2H$, $\tilde{A}_4 = A_4 - HA_2$, $\tilde{B}_1 = B_1 - NB_2$, $\tilde{B}_2 = B_2 - HB_1$, матрицы $H = H(t, \mu)$, $N = N(T, \mu)$ определяется из уравнения

$$\mu H' = \tilde{A}_1 H - HA_3 + A_4 H,$$
$$\mu N' = A_1 N - NA_2 + A_4. \quad (5)$$

Граничные условия системы (4) записываются в форме:

$\tilde{x}(t_0) = \tilde{x}_0, \tilde{z}(t_0) = \tilde{z}_0$,

$\tilde{x}(t_1) = \tilde{x}_1, \tilde{z}(t_1) = \tilde{z}_1$, где $z^i = z^i - H(t_i)x^i$, $\tilde{x}^i = x^i - N(t_i)z^i$, $i = 0;1$. (6) При замене исходную систему (2) эквивалентной системой (4) предполагается, что матрицы A_0, A_4 -устойчивы и при достаточно малых значений параметра μсобственные значения матрицы $\tilde{A}_1$ и $\tilde{A}_4$ были также отрицательными и близкими к собственным значениям матриц A_0, A_4.

Эти предположения выполняются, если выполняются условия теоремы, доказанной в [7]. А система (4) состоит из двух подсистем, которые связаны только по управлению. Теперь сформулируем задачу об управлении с минимальной нормой для системы (4) следующим образом: среди всех допустимых управлений требуется найти управление такое, которое доставляет минимум функционалу (1) при ограничениях (4)-(6). Назовем эту задачу возмущенной и обозначим ее символом P_μ.

Обращение в нуль параметра в системе (4) качественно изменяет структуры системы и её порядок. При $\mu = 0$ из (4) получаем

$$\dot{x} = A_0(t)x + B_0(t)u, \quad x(t_0) = x^0, \quad x(t_1) = x^1$$
$$(7)\ 0 = A_4^{-1}(t)A_3(t)x(t) - A_4^{-1}(t)B_2(t)u.$$

Систему (7) называют порождающей системой [5]. Задачу (1), (7) будем называть невозмущенной (предельной) по отношению к задаче P_μ обозначим её через P_0.

Второе уравнение системы (7) не является дифференциальным и заданные граничные условия для функции z будут лишними. Поведение системы (2) или (4) в окрестности граничных точек существенно отличается от поведения порождающей системы (7).

В этом случае интерес представляет следующий вопрос: будет ли предел решения задачи P_μ является решением задачи управления меньшей размерности, которую мы называли предельной или невозмущенной задачей P_0?

Следует заметить, что система (4) образуется через матрицы H, N,которые получаются из уравнения (5) в виде степенных рядов относительно малого параметра μ. Эти ряды сходятся равномерно. Назовем систему асимптотической (с точностью $O(\mu^{k+1})$ по отношению к системе (4), если её коэффициенты образованы через матричных полиномов

$H_k(t,\varepsilon), N_k(t,\varepsilon)$ по ε степени k. Такая система аппроксимирует систему (2) с точностью порядка малости $O(\varepsilon^{k+1})$.

Рассмотрим систему

$$\dot{x} = A_0(t)x + B_0(t)u,$$

$$\varepsilon\dot{\tilde{z}} = A_4(t_0)\tilde{z} + B_2(t_0)u. \qquad (8)$$

Эта система аппроксимирует систему (2) с точностью порядка ε, т.е. она является асимптотической с точностью $O(\varepsilon\varepsilon)$.

Система (8) получается из (4) при следующих приближениях:

$H(t,\varepsilon) = H_0(t) + \varepsilon A_4^{-1}(t)A_3(t), N(t,\varepsilon) = N_0(t) + \varepsilon A_2(t)A_4^{-1}(t),$

$\tilde{A}_1(t,\varepsilon) = A_0(t) + A_1(t) + A_2(t)A_4(t)A_3(t), \quad \tilde{B}_1(t,\varepsilon) = B_0(t) + B_1(t) + A_2(t)A_4(t)B_2(t), \tilde{A}_4(\varepsilon t) = A_4(t_*), \quad \tilde{B}_2(t_*) = B_2(t), \quad \tilde{z} = z + A_4^{-1}(t)A_3(t)x,$

t_*

$\overline{t} = t^*$ - растянутое время , $0 \le t \le t_1 \le \varepsilon^{-1}$, $t^* \in [t_0, t_1]$.

Заметим, что системы (7) и (8) отличаются только вторыми уравнениями. Поэтому здесь решение задачи P_ε построим для системы (8). Поправка к первому приближению не представляет трудности, т.е. все изложенные процедуры для системы (8) аналогично повторяется для высших приближений. Число слагаемых, образующие матричных полиномов H_k, N_k на алгоритм решения задачи P_ε ссущественных изменений не оказывают. Следует отметить, что быстрая подсистема рассматривается на большом промежутке времени, поэтому коэффициенты этой подсистемы считаются медленно меняющимися функциями [5].

Под решением возмущенной задачи P_ε следует понимать оптимальные траектории $x^0(t,\varepsilon), \tilde{z}(t,\varepsilon)$, управление $u^0(t,\varepsilon)$ и минимальное значение J_ε^0.

Подобные задачи рассмотрены в работах [2,3]. Здесь в отличие от указанных работ для исследуемых систем заданы краевые условия. Иначе говоря, эти системы называются системы с закрепленными конечными состояниями. При решении задачи P_ε и P_0 будет использован метод моментов.

РЕШЕНИЕ ЗАДАЧИ P_ε. Решения уравнения из (8) можно представить в виде:

$$x_0(t) = \Phi(t,t_0)x^0 + \int_{t_0}^{t}\Phi(t,s)B_0(s)u(s)ds, \qquad (9)$$

$$\sim z_0(t,\varepsilon) = e^{A_4(t_*)/\varepsilon}\sim z_0 + \frac{1}{\varepsilon}\int_{t_*}^{t} e^{A_4(t^*)/\varepsilon}\,ds \int_{t_0}^{t} B_2(t)u(s)\,ds, \qquad (10)$$

где $\Phi(t,t_0)$ - переходная матрица медленной подсистемы (8).

Решим задачу P_0. В силу соотношения (9) ограничение для медленной подсистемы $x(t_1) = x^1$ приводит к тому, что искомое управление $u = u^*(t)$ должно удовлетворять условию

$$\int_{t_0}^{t_1} \Phi(t_1,t)B_0(t)u(t)\,dt = a_1, \qquad (11)$$

где a_1-вектор, определяемый формулой $a_1 = x^1 - \Phi(t_1,t_0)x^0$. Управление $u = u^*(t)$, удовлетворяющее моментное соотношение (11) и доставляющее минимум функционалу (1) определяется формулой [1]:

$$u(t) = B_0(t)\Phi(t_1,t)W^{-1}(t_1,t_0)\left[x^1 - \Phi(t_1,t_0)x^0\right], \qquad (12)$$

где $W(t_1,t_0) = \int_{t_0}^{t_1} \Phi(t_1,t)B_0(t)B_0(t)\Phi(t_1,t)\,dt.$

Тогда оптимальные траектории $x^*(t)$, $\sim z^*(t)$ порождающей системы (7) соответствующие управлению $u^*(t)$ записываются в виде:

$$x^*(t) = \Phi(t,t_0)x^0 + \int_{t_0}^{t} \Phi(t,s)B_0(s)u^*(s)\,ds, \qquad (13)$$

$$\sim z^* = A_4^{-1}(t)B_2(t)u^*(t), \qquad (14)$$

где $\sim z^* = \sim z^*(t) = A_4^{-1}(t)A_3(t)x^*(t), u^*(t)$ определяется формулой (12). При $t = t_0, t = t_1$ из (14) получим

$$\sim z^*(t_0) = A_4^{-1}(t_0)B_2(t_0)u^*(t_0), \qquad \sim z^*(t_1) = A_4^{-1}(t_1)B_2(t_1)u^*(t_1). \qquad (15)$$

Как было отмечено выше, что мы должны иметь такое решение задачи P_ε которое при $\varepsilon \to 0$ переходит в решение задачи P_0. В силу соотношений (10), (14) и (15) разность векторов $\sim z(t,\varepsilon) - z^*(t)$ определяет следующую формулу:

$$\frac{t-s}{\varepsilon}$$

$$\tilde z(t,\varepsilon) = e^{A_4 (t - t_0)/\varepsilon}\left(\tilde z^0 + A_4^{-1}(t_0)B_2(t_0)u^*(t_0)\right) + \ldots B_2(t)u(s)\,ds, \qquad (16)$$

$$A_4^{-1}(t)B_2(t)u^*(t) + \frac{1}{\varepsilon}\int_{t_0} e \ldots$$

Для задачи P_ε определим управление следующим образом:

$$u(t) = \begin{cases} u^*(t), & t_0 \le t \le t_1, \\ u^*(t) + V\left(\frac{t_1 - t}{\varepsilon}\right), & 0 \le \frac{t_1 - t}{\varepsilon} \le \frac{t_1 - t_0}{\varepsilon}, \end{cases} \qquad (17)$$

$\tau = \frac{t_1 - t}{\varepsilon}$ - пограничная функция, которая имеет экспоненциальный характер убывания.

где $V(\tau)$

Управление $u^0(t) = u^*(t)$, имеющее минимальную норму и переводящее медленную подсистему из начального состояния $\overline{x(t_0) = x^0}$ в конечное состояние $x(t_1) = x^1$ нам уже известно. Теперь остается построить $V\left(\frac{t_1 - t}{\varepsilon}\right)$. При $t = t_1$ с учетом (17) из (1), (16) получим:

$$a_2 = \int_0^{\infty} e^{A_4(t_1)\tau} B_2(t_1) V(\tau)\, d\tau, \qquad (18)$$

$$\int_0^{\infty} V^2(\tau)\, d\tau \to \min \qquad (19)$$

где

$$a_2 = z_1 e^{A_4(t_0)(t_1)} \tilde{z}_0 + A_4(t_0)\int_{t_0}^{t_1} u^*(t_0)\,dt + A_4(t_1)\int_{t_1}^{t_1} B_2(t_1)\,u^*(t_1)\,dt, \quad t_1 = t_1 - t_0. \tag{20}$$

Решение задачи (18), (19) согласно проблемы моментов [2] записывается в виде:

$$V(t) = B_2^T(t_1)\,e^{A_4^T(t_*)}\,W^{-1}a_1, \tag{21}$$

где $W_* = \int_0^{t_*} e^{A_4(t_*)}B_2(t_1)B_2^T(t_1)e^{A_4^T(t_*)}\,dt_*$, $t_1 = t$.

Управление $u_0(t) = V(t)$ переводит быструю подсистему из начального состояния $\tilde{z}(t_0) = \tilde{z}^0$ в конечное состояние $\tilde{z}(t_1) = \tilde{z}^1$, имеют минимальную норму и при $t_1 - t_0 \to 0$ стремится к нулю. С учетом (21), из (16) будем иметь:

$$\tilde{z}_0(t,t_0) = e^{A_4(t_0)(t_1)}\tilde{z}_0 + A_4(t_0)\int_{t_0}^{t_1} B_2(t_0)\,u^*(t_0)\,dt + A_4(t)\int_{t_1}^{t} B_2(t)\,u^*(t)\,dt \tag{22}$$

$$W(t_1,t_0) = e^{A_4(t_1)}W_*^{-1}\tilde{z}_1 - e^{A_4(t_0)(t_1)}\tilde{z}_0 + A_4(t_0)\int_{t_0}^{t_1} B_2(t_0)\,u^*(t_0)\,dt + A_4(t_1)\int_{t_1}^{t} B_2(t_1)\,u^*(t_1)\,dt,$$

$$W(t_1,t_0) = \int_{t_0}^{t_1} e^{A_4(t_*)}B(t)B^T(t)e^{A_4^T(t_*)}\,dt,$$

где t_0

$t = t^1$, $t_1 = t_1 - t_0$, $t_0 = t_0 - t_1$, $t_1 = t_1 - t_0$.

$t - t$

ТЕОРЕМА. Если $A_4(t)$- устойчивая матрица, то при управлении $u(t) = V^T(t)W^{-1}$

вектор переменных состояния быстрой подсистемы $\tilde{z}_0(t,t_0)$ дается формулой

$$\tilde{z}_0(t,t_0) = e^{A_4(t_*)(t_0)}\tilde{z}_0 + A_4(t_0)\int_{t_0}^{t} B_2(t_0)u^*(t_0)\,dt + A_4(t)\int_{t}^{t} B_2(t)u^*(t)\,dt$$

$$- W(t)e^{A_4(t_*)(t_0)}W_*e^{A_4^T(t_1)(t_0)}\int_{t_0}^{t_1} e^{A_4(t_*)(t_0)}W_*^{-1}a_2. \tag{23}$$

ДОКАЗАТЕЛЬСТВО. Матрица $W(\tau,\tau_0)$ удовлетворяет дифференциальному уравнению

$$\frac{dW}{d\tau} = A_4(t_1)W - WA_4(t_1) - B_2(t_1)B_2'(t_1) \quad (24)$$

$$W(\tau,\tau_0) = 0$$

По условию теоремы $A_4(t)$-устойчива, то несобственный интеграл

$$W_* = \int_0^\infty e^{A_4(t_*)}B_2(t_1)B_2'(t_1)e^{A_4'(t_*)}dt,$$ сходится и является единственным решением

матричного алгебраического уравнения Ляпунова [7]

$$A_4(t_1)W - WA_4(t_1) - B_2(t_1)B_2'(t_1) = 0. \quad (25)$$

Введем в (24) замену переменной $V = W - W_*$. Тогда с учетом (25) из (24) получим

$$\frac{dV}{d\tau} = A_4(t_1)V - VA_4(t_1), \quad V(\tau_0,\tau_0) = -W_* \quad (26)$$

Решение (26) можно записать в виде:

$$V(\tau,\tau_0) = e^{A_4(t_*)(\tau-\tau_0)}W_*e^{A_4'(t_*)(\tau-\tau_0)} - \int_{\tau_0}^{\tau} e^{A_4(t_*)(\tau-t)}W_*e^{A_4'(t_*)(\tau-t)}dt. \quad (27)$$

Отсюда будем иметь:

$$W(\tau,\tau_0) = W_* - e^{A_4(t_*)(\tau-\tau_0)}W_*e^{A_4'(t_*)(\tau-\tau_0)} \quad (28)$$

Подставляю (28) в (22) получим (23), ч.т.д.

Как мы отметили выше, что управление $V(t) = -B_2'(t)W_*^{-1}z(t)$ переводит быструю подсистему из начального состояния $z(t_0) = z_0$ в конечное состояние $z(t_1) = z_1$ и соответствующая оптимальная траектория $z_0(t,\tau)$ при $\tau\to 0$ стремится к решению порождающей системы

(7), причем в ней содержатся левые и правые пограничные функции, которые имеют существенные значения в окрестности граничных точек и при удалении от них быстро убывают. Таким образом, мы полностью определили первое приближение решения задачи P_ε.

При определении высших приближений решение задачи $P_{\square}$, выше изложенные процедуры аналогично повторяются. Как отмечены выше, что число слагаемых, образующих матричных полиномов $H_k\ \square t,\square\square$, $N_k\ \square t,\square\square$ на алгоритм решения задачи существенных изменений не оказывают. Члены ряда $H\square t,\square\square$, $N\square t,\square\square$ определяются из две независимых рекуррентных процедур, которые образованы согласно уравнения (5) [8].

ЗАКЛЮЧЕНИЕ. В заключении отметим следующее:

Равномерное нулевое приближение к оптимальному управлению $u^0\ \square t,\square\square$ может быть получено в результате декомпозиции исходной задачи на две задачи оптимального управления меньшей размерности, что позволяет сократить объём вычислительных процедур при решении конкретных практических задач;

Для каждой задачи оптимального управления меньшей размерности эффективно применялся метод моментов, что дало возможность определить оптимальное управление в замкнутой аналитической форме, при этом одновременно решается и краевая задача. Кроме того, предложенный приближенный способ позволяет изучить магистральные свойства оптимальных траекторий процесса;

Список литературы

1. Красовский Н.Н. Теория управления движениям - М.: Наука, 1968,47 с.
2. Иманалиев З. К. О приближенном решении сингулярно возмущенной задачи оптимального управления. //Иссл. По интегро - дифф. уравнениям. - Бишкек: Илим, 1997.ВЫП.26.-С.152- 155.
3. Глизер В.Я., Дмитриев М.Г. Асимптотика решения одной сингулярно возмущенной задачи Коши, возникающей в теории оптимального управления.//Дифф. урав. - 1978. -Т.14, №4. - С.601 -612.
4. Васильева А.Б., Бутузов В.Ф. Асимптотические разложения решений сингулярно - возмущенных уравнений. - М.: Наука, 1973, 272с.
5. Иманалиев З.К., Аширбаев Б.Ы. О переходных матрицах медленных и быстрых подсистем с малым параметром. // Материалы межд. науч. конф. посв. 70 -летию акад. М.И. Иманалиева, «Асимптотические топологические и компьютерные методы в математике», Бишкек, 2001.-С.235-239.
6. Иманалиев З.К., Пахыров З.П., Аширбаев Б.Ы., Баракова Ж.Т. Разделение быстрых и медленных движений системы управления с малыми параметрами. // Материалы международной научной конференции: «Современные технологии и управления качество в образовании, науке, производстве. Опыты, адаптации и внедрения», Ч.1., Бишкек: 2001, С.244-250.
7. Стрыгин В.В., Соболев В.А. Разделение движений методом интегральных многообразий - М.: Наука, 1988, 256с.

УДК 621.326.77

ЭНЕРГОСБЕРЕГАЮЩИЕ ЛАМПЫ: ПЛЮСЫ И МИНУСЫ

Ташматов Абдыманап Суранбаевич, *к.т.н., доцент, КГТУ им. И.Раззакова,* ***Нарынбаев Алишер Фархатович,*** *студент, КГТУ им. И.Раззакова, Бишкек, Кыргызская Республика. 720044 г. Бишкек, пр. Мира 66, e-mail:aebrat@mail.ru*

Цель статьи – сравнение эксплуатационных, физических и экономических параметров энергосберегающих ламп и традиционных ламп накаливания. Рассмотрены практические примеры применения энергосберегающих ламп, их принцип работы и структурные особенности. В статье подробно рассматриваются преимущества и недостатки энергосберегающих ламп в эквивалентном сравнении с традиционными лампами накаливания, проиллюстрированные на практических и наглядных расчётах с целью выявления экономической выгоды с позиции задач энергосбережения.

Ключевые слова: энергосбережение, энергосберегающие лампы, лампы накаливания, освещение, экономия, электроэнергия, электрическая мощность, расходы, световой поток.

ENERGY SAVING LAMPS: ADVANTAGES AND DISADVANTAGES

Tashmamatov Abdymanap Suranbaevich, *Ph,D, Associate Professor, Kyrgyzstan, 720044, c. Bishkek, KSTU named after I.Razzakov,*
Narynbaev Alisher Farhatovich, *student, KSTU named after I.Razzakov, Kyrgyzstan, Bishkek, Miraave. 66, e-mail: aebrat@mail.ru*

The purpose of this article – comparison of operating, physical and economic parameters of energy-saving lamps and traditional incandescent bulbs. Here considered the practical examples of using energy-saving lamps in equivalent comparison with traditional incandescent lighting, illustrated in practical calculations to identify the economic benefits from the perspective of energy saving tasks.

Keywords: energy-saving, energy-saving lamps, incandescent lightings, lighting, economy, electric energy, electric power, costs, light stream.

Как известно, тьма – это отсутствие света. Человек 80% информации о мире воспринимает глазами. Так как естественного освещения со временем стало недостаточно, человечество изобрело осветительные приборы. Сначала это были примитивнейшие конструкции, но со временем они стали совершеннее во всех сферах своих возможностей. Сегодня, когда увеличивается количество используемых нами бытовых приборов, затраты на электроэнергию вырастают в разы.В каждой семье есть холодильник, телевизор, стиральная машина. Все чаще в наших квартирах «прописываются» компьютеры, посудомоечные машины, кухонные комбайны, электрочайники и другие приборы. Изрядное количество электроэнергии расходуется на освещение. Электроэнергия поступает в наши дома с электростанций различного типа и для ее производства сжигаются уголь, нефть, газ.

Экономное использование электроэнергии позволит сократить объемы использования этих энергетических ресурсов, а значит снизить выбросы вредных веществ в атмосферу, сохранить чистоту водоемов. Тем самым каждый из нас может внести свой посильный вклад в общее дело сохранения природы. Проблема энергосбережения даёт о себе знать и многие страны мира уже обратили своё внимание на складывающуюся ситуацию вокруг. На Кубе из-за проблем с энергетическими ресурсами власти приняли решение повсеместно использовать энергосберегающие люминесцентные лампы. 27 государств-членов ЕС уже сделали первый шаг на пути к энергосберегающим технологиям: здесь запрещен оборот 100ваттных ламп накаливания. Список стран, начавших переход к энергосберегающим технологиям, готова пополнить Россия. Не исключено, что и в нашей стране скоро

задумаются об этом шаге, и тогда о лампе накаливания, которая сейчас есть практически у каждого, мы будем говорить как о части истории.

В городе Бишкек, в декабре 2008 года в целях экономии потребления электроэнергии и соблюдения установленного лимита, мэрия города издало постановление по переходу на использование энергосберегающих ламп накаливания всех муниципальных организаций и предприятий города Бишкек к определенному сроку. Частично это постановление было исполнено - часть бюджетных зданий действительно перешла на энергосберегающие лампы, но преткновением стало то, что вопросы утилизации большого количества ртутносодержащих ламп на законодательном уровне не решаются. В нашей стране вопросы запрета на продажу ламп накаливания стоят под большим вопросом ввиду наличия отечественного производителя ламп накаливания - ОАО «Майлуу-суйский электроламповый завод». В исследовании, проведённом по заказу Министерства экономического развития и торговли КР при поддержке USAID было обозначено, что переход на энергосберегающие лампы в Кыргызстане мог бы обеспечить ежегодную экономию электроэнергии в республике до 1 миллиарда кВт. часов. Согласно проведённому исследованию, более 95 процентов источников света, используемых в Кыргызстане в бытовых целях, являются лампами накаливания.

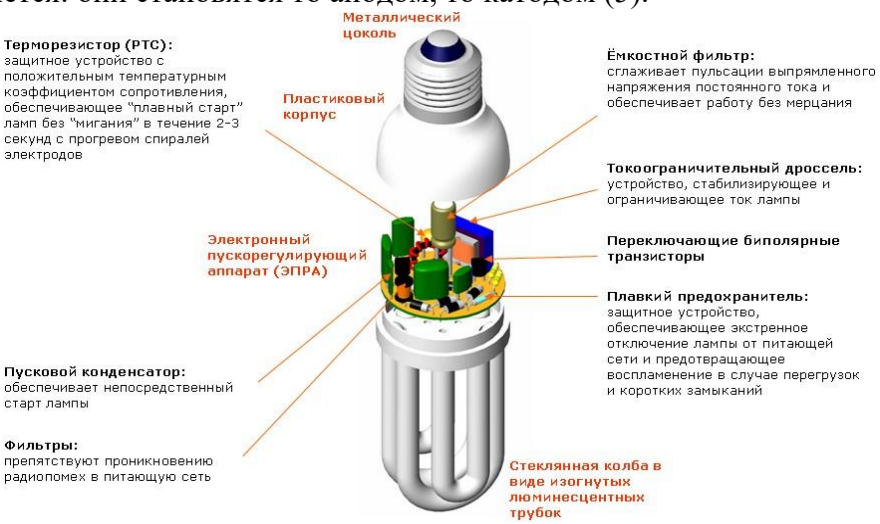
Так как же устроена энергосберегающая лампа, и почему её использование так благотворно сказывается на экономической составляющей потребления электроэнергии?

Итак, энергосберегающая лампа состоит из:

- стеклянной колбы(трубки), заполненной газом(чаще всего ртутью). На внутреннюю часть стеклянной колбы нанесено специальное вещество - люминофор - пускового устройства(стартера)

Пусковое устройство находится под пластиковым корпусом.

Принцип работы энергосберегающих ламп: Стеклянная трубка имеет на концах два электрода, которые нагреваются до 900-1000 градусов и испускают множество электронов, которые сталкиваются с атомами аргона и ртути. Возникающееультрафиолетовое излучение, попадая внутреннюю поверхность, покрытую люминофором, преобразуется в видимый свет. К электродам, разумеется, подводится переменное напряжение, поэтому их функция постоянно меняется: они становятся то анодом, то катодом (5).



Основные преимущества энергосберегающих ламп:

1. Высокая световая отдача, превышающая тот же показатель ламп накаливания в несколько раз.

2. Значительный срок службы. От 6 до 15 тысяч часов, то есть от 8 месяцев до около 2 лет непрерывного свечения. Эта цифра превышает срок службы обычных ламп накаливания приблизительно в 20 раз.
3. Энергосберегающая лампа мощностью в 11 Ватт даёт такой свет, как лампа накаливания с мощностью в 60 Ватт. Очевидная экономия – 49 Ватт
4. Незначительное тепловыделение, которое позволяет использовать компактные люминесцентные лампы большой мощности в хрупких светильниках и люстрах.
5. Свет распределяется равномернее, чем у ламп накаливания. Это объясняется тем, что в лампе накаливания свет идет только от вольфрамовой спирали, а

   энергосберегающая лампа светится по всей своей площади
6. КПД энергосберегающих ламп более 45% Недостатки:

1. Высокую цену по сравнению с традиционными лампами накаливания.
2. Неработоспособность в низком диапазоне температур (менее -15 ... -20 ºС)
3. Срок службы энергосберегающих ламп ощутимо зависит от режима эксплуатации, в частности, они не любят частого включения и выключения.
4. Фаза разогрева у них длится от 5 сек. до 5 минут, то есть, им понадобится некоторое время, чтобы развить свою номинальную яркость и "заявленную" цветовую температуру
5. Наличие ртути и фосфора, которые, хоть и в очень малых количествах, присутствуют внутри энергосберегающих ламп.
6. При снижении напряжения в сети более чем на 10%, довольно часто, энергосберегающие лампы просто не зажигаются.

Ключевым пунктом является то, что КПД энергосберегающих ламп превышает 45%. Этот пункт и является главным преимуществом перед традиционными лампами накаливания, коэффициент полезного действия которых в среднем составляет ничтожные 8% (4).

В качестве расчёта экономии, возьмём для примера однокомнатную квартиру с 5 лампами накаливания по 100 ватт. В среднем, лампы работают непрерывно по 7 часов в день. Тариф на электроэнергию за 1 квт-час составляет 0, 70 сом. Проведя несложные расчёты, рассчитаем расходы на лампу мощностью 100 Вт в течение месяца:

0,1 квт * 210 часов * 0, 70 сом = 14, 7 сом

Так как ламп у нас 5, месячный расход = 73, 5 сом

Рассчитаем годовой расход: 0,1 квт * 2500 часов * 0, 70 сом * 5 = 875 сом

Заменив лампы накаливания на эквивалентные по мощности энергосберегающие лампы(100 Вт эквивалентно 20 Вт) получим:

0,02 квт * 210 часов * 0, 70 сом = 2, 94 сом

Месячный расход на 5 энергосберегающих ламп = 14, 7 сом

Годовой расход: 0,02 квт * 2500 часов * 0, 70 сом * 5 = 175 сом

Таким образом, получается, что энергосберегающая лампа, несмотря на высокую стоимость, экономичнее почти в 3 раза (!), чем дешевая лампа накаливания.

Даже если считать, что лампа накаливания не выходит из строя целый год, на одной лампе мы сэкономим 140сомов, соответственно на 5 лампах экономия составит700сомов в год. Эти цифры могут показаться незначительными, вдобавок к тому, что стоимость энергосберегающих ламп превышает стоимость ламп накаливания. Однако, расходы на энергосберегающую лампу окупаются долгим сроком службы (3-4 года), чем не может похвастаться лампа накаливания. В будущем тарифы на стоимость электроэнергии неизбежно будут повышаться, и тогда выбора уже не останется – мир перейдёт на светодиодное освещение и полностью откажется от лампочек Ильича. Некоторые люди и вовсе считают, что возникновение парникового эффекта связано с работой ламп накаливания, которые и «нагрели» нашу атмосферу. Рассмотрев все «за и «против» мы можем сделать свой выбор в пользу энергосберегающих ламп. К счастью, у нас он ещё есть, хотя многие страны в мире такого выбора уже лишены.

История электрической лампочки началась в 1872 г. в Санкт-Петербурге. Именно тогда профессор физики Василий Владимирович Петров пропустил электрический ток по двум стержням из древесного угля. Между ними дугой перекинулось пламя. Обнаружились неизвестные ранее свойства электричества — возможность давать людям яркий свет и тепло. Основной вклад в создание электрической лампочки внесли трое людей, по иронии судьбы родившихся в один и тот же 1847 год. Это были русские инженеры Павел Николаевич Яблочков, Александр Николаевич Лодыгин и американец Томас Эдисон. Лодыгин начал опыты с электрической дугой, но очень быстро от них отказался, так как увидел, что раскаленные концы угольных стержней светят ярче, чем сама дуга. Угольный стерженек помещался между двумя медными держателями в стеклянный шар, по нему пропускался электрический ток. Уголь давал свет довольно яркий, хотя и желтоватый. Угольный стержень выдерживал примерно полчаса. Академия наук присвоила Лодыгину Ломоносовскую премию за то, что его изобретение приводит к «полезным, важным и новым практическим применениям». В это же время собственную конструкцию лампы разрабатывал Яблочков. Она представляла собой два угольных стержня, между которыми вспыхивала электрическая дуга. По мере сгорания стержней их сближал механический регулятор. Ток давала гальваническая батарея. Его занимала одна проблема: как построить лампу, не нуждающуюся в регуляторе. Решение оказалось простым: вместо того, чтобы располагать стержни один против другого, их надо было поставить параллельно, разделив прослойкой тугоплавкого вещества, не проводящего электрический ток. Для прослойки между электродами Яблочков выбрал каолин — белую глину, из которой делают фарфор.

Изобретение имело огромный успех. Магазины, театры, улицы Парижа были освещены «свечами Яблочкова» (3). В Лондоне ими осветили набережную Темзы и корабельные доки. Яблочков стал одним из самых популярных в Париже людей. Газеты называли его изобретение «русским светом». Говоря о вкладе Эдисона в развитие электрической лампочки, следует отметить, что перед созданием своей лампочки в его руках побывала лампочка Лодыгина. Эдисон начал работу над лампой с угольной нитью накаливания, помещенной в стеклянный шар, из которого выкачан воздух. Он нашел способ выкачивать воздух из баллона лучше, чем это удавалось другим изобретателям. Однако главное было найти материал для угольной нити, который бы обеспечил долгий срок службы. Для этого Эдисон перепробовал около шести тысяч растений из разных стран мира. В конце концов он остановился на одном из видов бамбука. Несмотря на все это, Томас Эдисон получил патент не на изобретение лампочки, а лишь на усовершенствование, поскольку приоритет оставался за Лодыгиным.Он и нашел самый подходящий материал для нити, использующийся до сих пор — вольфрам.

В XX веке у лампочек накаливания появился конкурент — газосветные лампы, или лампы дневного света. Они наполнены газом и дают свет, не нагреваясь. Сначала появились цветные газосветные лампы. В стеклянную трубку с обоих концов вплавлялись металлические пластины — электроды, к которым подводился ток. Трубка наполнялась газом или парами металла. Под воздействием тока газ начинал светиться.

В 1901 году инженер-изобретатель из США Питер Купер Хьюитт создал первую люминесцентную лампу. Однако, особой популярности лампа не получила, так как излучаемый ею свет был голубовато-зеленый, неприятный для глаза человека.

В 1926 году группа изобретателей под руководством Эдмунда Гермера создала лампу с нанесенным флуоресцирующим покрытием. Эта лампа уже излучала белый свет и не создавала никаких неудобств человеку.

В 1939 году на нью-йоркской выставке впервые была представлена лампа U-образной формы.

В 1976 году была разработана лампа спиралевидной формы, но из-за своей дороговизны, в серийное производство она не была запущена.

И наконец, в 1995 году китайские производители запустили в массовое производство энергосберегающие лампы.

Выводы:

1. Эффективное использование энергии — ключ к успешному решению экологической проблемы - сохранение невозобновляемых источников энергии.
2. Люминесцентные лампы являются одним из наиболее экономичных источников света. Отношение светового потока к истребляемой электроэнергии в 10 раз лучше, чем у ламп накаливания.
3. Срок службы энергосберегающей лампы колеблется от 6000 до 12000 часов (как правило, длительность срока службы указывается производителем на упаковке товара) и превышает срок использования лампы накаливания в 6 - 15 раз.
4. Энергосберегающая лампа, несмотря на высокую стоимость, экономичнее в 3 раза, чем дешевая лампа накаливания.
5. Серьезный недостаток энергосберегающих ламп – это использование ртути в их производстве. Ртуть – токсичное вещество, поэтому содержащие ее приборы требуют специальной утилизации.

Список литературы

1. Трофимова Т.И., Курс физики Учебное пособие для ВУЗов. Москва. Высшая школа, 1990 г., 478 с.

2. Википедия Свободная энциклопедия Компактная люминесце нтная

ла мпаhttp://ru.wikipedia.org/wiki

3. Малинин Г. Изобретатель "русского света". – Саратов: Приволжское кн.изд-во, 1999г.

4. Данилов Н.И., Тимофеева Ю.Н., Щелоков Я.М. «Энергосбережение для начинающих» Екатеринбург,2004г.

5. Энергосберегающие лампы http://www.goodpc.narod.ru/

ПРИКЛАДНАЯ МАТЕМАТИКА И МЕХАНИКА

УДК 517

ЕДИНСТВЕННОСТЬ РЕШЕНИЯ СИСТЕМЫ НЕЛИНЕЙНОГО ИНТЕГРАЛЬНОГО УРАВНЕНИЯ ФРЕДГОЛЬМА ПЕРВОГО РОДА

Сапарова Гульмира Баатыровна, *Ошский Технологический Университет, кафедра «Прикладная математика», г. Ош, E-mail:gulya141005@mail.ru*

Доказано существование и единственность решения системы нелинейного интегрального уравнения Фредгольма первого рода с разрывным ядром в пространстве C [*a*, *b*].

Ключевые слова: *интегральные уравнения, первого рода, единственность, резольвента, разрывное ядро, решение, существование, условие Липшица, регуляризация.*

SINGLE SOLUTION SYSTEM OF NON – LINED INTEGRAL EQUATION OF FREDGOLM OF FIRST TYPE.

G. Saparova, *Osh Technological University, chair "Applied Mathematics", Osh city*

The existence and uniqueness of solutions system of non – lined integral equation of Fredgolm of first type with discontinuous kernel in sphere C[a,b] is provid

Key words: integral equation of the first kind, discontinuous kernel, solution, existence, uniqueness, resolvent, Lipschitz condition, regularization.

Рассмотрим систему

$$\int_a^t H(s)u(s)ds + \int_t^b N(s)u(s)ds + \int_a^t K(t,s,u(s))ds = g(t) \quad t\in[a,b], \qquad (1)$$

где *H(s),N(s)* – n x n – мерные известные матричные функции, K(t,s,u(s)) – известная непрерывная n – мерная вектор-функция, g(t) и u(t) – известная и искомая функции.

Наряду с системой (1) будем рассматривать систему

$$\varepsilon v(t,\varepsilon) + \int_a^t H(s)v(s,\varepsilon)ds + \int_t^b N(s)v(s,\varepsilon)ds + \int_a^t K(t,s,v(s,\varepsilon))ds = g(t) + \varepsilon u(a), \qquad (2)$$

где u(t) – решение системы (1), $0 < \varepsilon$ - малый параметр.

Введем обозначения:

1) $\langle\cdot,\cdot\rangle$ - скалярные произведения в R^n, ||A||, ||u|| - нормы соответственно n x n – мерной матрицы A и n – мерного вектора u, то есть для любых

$$u = (u1,u2,.....,un), v = (v1,v2,.....,vn) \in R^n, \langle u,v\rangle = u1v1 + u2v2 + ...+ unv_n,$$

$\|u\| = \sqrt{u,u}$;

2) $C_n[a,b]$- пространство n – мерных векторов с элементами из C[*a*,*b*], $\|\cdot\|_c$ -норма в $Cn[a,b]$, то есть для любого $u(t) \in Cn[a,b]$

$$\| u(t) \|_c = \max_{t \in [a,b]} \| u(t) \|$$

Предполагаем выполнение следующих условий:

а) $\det[H(t) + N(t)] \neq 0$ при всех t∈[*a*,*b*]. Не ограничиваясь общности будем считать $(H(t)+N(t)) = I_n$ при всех t∈[a,b], где I_n –*nxn*- мерная единичная матрица. Так как в случае $\det[H(t) + N(t)] \neq 0$, t∈[*a*,*b*] в системе (1) можем сделать замену

$$u(t) = [H(t)+N(t)]^{-1} u1(t);$$

б) при t ≠ τ для любых $(t,s,u1),(\tau,s,u1),(t,s,u2),(\tau,s,u2) \in G \subset R$
справедлива оценка:

$$|K(t,s,u1)-K(\tau,s,u1)-K(t,s,u2)+K(\tau,s,u2)| \le M1(t-\tau)(u1-u2)$$

где $G = \{(t,s): a \le s \le t \le b\}$, 0 < M_1 – известная постоянная.

В дальнейшем используется следующая теорема.

ТЕОРЕМА 1. Если выполняются условия а) и б). Тогда система (1) имеет единственное решение u(t) ∈C_n[*a*,*b*], тогда и только тогда, когда

$$g(0) = \int_a^b N(s)u(s)ds \qquad \| g(t)\| \in C^1[a,b], a$$

где u(t) – решение системы t

$$u(t) = a + \int_a^K t(t,s,u(s))ds + g(t) \qquad t \in [a,b], \tag{3}$$

ДОКАЗАТЕЛЬСТВО: Из системы (1) имеем

$$\int_a^t [H(s) + N(s)]u(s)ds - \int_a^b N(s)u(s)ds - \int_a^t K(t,s,u(s))ds = g(t),$$

$$\int_a^t u(s)ds + \int_a^t K(t,s,u(s))ds = g(t) + \tau, \tag{4}$$

где

$$b$$

$$\square\square\ \square N(s)u(s)ds.\ a$$

Берем производную по t с обеих сторон системы (4), получаем

t

$$u(t)\ \square a\square^{K}{}_{t}\square(t,s,u(s))ds\ \square\ g\square(t) \qquad \text{при } t\square[a,b].$$

Отсюда имеем систему (3).

Системы (3) и (1) эквивалентны тогда и только тогда, когда $g(0)\square\square\square 0.$

Теорема 1 доказана.

Предположим выполнения следующих условий:

в) $\det(^{I}{}_{n}\ \square M(\square))\ \square\ 0$ при $\square > 0$, где

$$M(\square)\ \square\ \square\square^{\frac{1}{}}ba\square N(s)X(s,a,\square)ds, X(t,s,\square)\ \square\ I_{n}e^{\square\ \square}{}_{\underline{t}}\ \square\underline{s}\ ;$$

г) $\square i(t), i\ \square 1,2,.....,n\square$ собственные значения матрицы $\frac{1}{2}[H(t)\square\ N(t)\square\ H\square(t)\square\ N\square(t)]$ и

$\min\square i(t)\ \square\square(t)\ \square\square\square\ 0$, при $t\square[a,b]$;

д) K(t,t,u)=0, для любых (t,u)$\square$[*a,b*]xR;

е) K(t,s,0)=0, при (t,s) $\square$G={(t,s): $a \le s \le t \le b$} и

$$\| X\ (t,s,\square)\ \|\square\ e\square\underline{t}\square\square\underline{s}$$

$$N_{0}\square \sup_{t\square[a,b]}\|N(s)\|$$

Подставляем в систему (2) формулу

$$v(t,\square)\square u(t\ )\square\square(t,\square) \qquad (5)$$

где u(t) – решение системы (1).

$$\square\square(t,\square)\ \square H(s)\square(s,\square)ds\ \square\ \square\ N(s)\square(s,\square)ds\ \square\ \square\square[K(t,s,u(s)\square\square(s,\square))\ K(t,s,u(s))]ds$$

$$\square\square[u(t)\square u(a)].$$

Вводим обозначения

$$g(t,\square)\ \square\ \square\square[K(t,s,u(s)\square\square(s,\square))\square K(t,s,u(s))]ds\square\square[u(t)\square u(a)] \qquad (6)$$

Отсюда в силу следствия теоремы и применяя формулу Дирихле, имеем

$$\square(t,\square)\ \square\ \frac{1}{2}\ X(t,a,\square)\square(In\ \square\ M(\square))^{\square 1}\square b_{\square}b_{\square}\ N(s)[K(s,\square,u(\square)\square\square(\square,\square))\ K(s,\square,u(\square))]d\square ds\ \square$$

$$\square \qquad\qquad a\square$$

— □ 1 *X*(*t*,*a*,□)□(*In* □1 *bN*(*s*)[*u*(*s*)□*u*(*a*)]*ds* □ □

M(□)) □

□ *a*

□ 1/3 *X*(*t*,*a*,□)□(*In* □ *M*(□))□1 *b*□*b*□ *s*□ *N*(*s*)*X*(*s*,□,□)□[*K*(□,□,*u*(□)□□(□,□))□*K*(□,□,*u*(□))]*d*□

□ *a*□□

dsd□□ 1/2 *X*(*t*,*a*,□)□(*In* □ *M*(□))□1 *ba*□*as*□ *N*(*s*)*X*(*s*,□,□)[*u*(□)□*u*(*a*)]*d*□*ds* □

□

□ 1 *t*□ *K*(*t*,*s*,*u*(*s*)□□(*s*,□))□ *K*(*t*,*s*,*u*(*s*))*ds* □[*u*(*t*)□*u*(*a*)]□

□*a*

□ 1/2 *at*□□*t*□ *X*(*t*,*s*,□)□[*K*(*s*,□,*u*(□)□□(□,□))□ *K*(*s*,□,*u*(□))]*d*□*ds* □□1 *at*□
X(*t*,*s*,□)□[*u*(*s*)□*u*(*a*)]*ds*

□

Введем, новые обозначения $H1(t,□,□(□,□),□), H2(t,□,□(□,□),□), □(t,□)$, и получаем новое уравнение

$$□(t,□) □ \int_a^b H_1(t,□,□(□,□),□)d□□\int_a^t H_2(t,□,□(□,□),□)d□□□(t,□)$$

(7) где

$$H_1(t,□,□(□,□),□) □ \frac{1}{2} X(t,a,□)(I_n - M(□))^{-1} \int bN(s)[K(s,□,u(□)□□(□,□))-K(s,□,u(□))]ds-$$

□ □

$$-\frac{1}{3} X(t,a,□)(In - M(□))^{-1} \int □b□s□ N(s)X(s,□,□)[K(□,□,u(□)□□(□,□))□K(□,□,u(□))]d□ds,$$
(8)

□ □□

$$H2(t,\square,\square(\square,\square),\square) \square \square\frac{1}{\square}[K(t,s,u(s)\square\square(s,\square))\square K(t,s,u(s))]\square$$

$$\square \frac{1}{2}\square X(t,\square,\square)[\square K(\square,\square,u(\square)\square\square(\square,\square))\square K(\square,\square,u(\square))]d\square\,, \qquad (9)$$

$$\square(t,\square)\square\square\frac{1}{\square}X(t,a,\square)(In \square M(\square))^{\square 1}\int_a^b \square N(s)[u(s)\square u(a)]ds \square$$

$$\square \frac{1}{2}\square X(t,a,\square)(In \square M(\square))^{\square 1}\int_a^b \square\int_a^s \square N(s)X(s,\square,\square)[u(\square)\square u(a)]d\square ds \square$$

$$\square[u(t)\square u(a)]\square \frac{1}{\square}\int_a^t \square X(t,s,\square)[u(s)\square u(a)]ds, \qquad (10)$$

ЛЕММА 1. Пусть выполняются условия а),б),в), и г) $\|(In \square M(\square))\square 1\| \square M0$, где известная постоянная $0 \square M_0$ не зависит от $\square \square 0$,

$H1(t,\square,\square(\square,\square),\square), H2(t,\square,\square(\square,\square),\square),\square(t,\square)$ определены по формулам (8),(9),(10), тогда справедливы следующие оценки:

$$\|H1(t,\square,\square(\square,\square),\square)\| \square \frac{1}{\square}\|e^{\frac{1}{\square}(t\square a)M_0N_0M_1(b\square a)}\| \text{при всех t,}\square\square\text{[a,b];} \qquad (11)$$

$$\|H2(t,\square,\square(\square,\square),\square)\| \square M_1 \quad \text{при всех} \quad (t,\square)\square G\ ; \qquad (12)$$

$$\square(t,\square) \square (M_0N_0L(b\square a)e^{\frac{1}{\square}(t\square a)} \square L\square), \text{при всех } t \square[a,b] \qquad (13)$$

где

$$N0 \square \sup_{t\square[a,b]} |N(s)|,$$

а функция u(t) удовлетворяет условию Липшица с коэффициентом L, то есть для любых t,s □ [a,b], при t > s справедлива оценка

$|u(t)□u(s)|□ L(t □s)$

ДОКАЗАТЕЛЬСТВО. Пусть $a□t □b,$ тогда, учитывая формулу Дирихле преобразуем функцию H_1 (t,□, □(□,□) ,□),

H1(t,□,□(□,□),□) □ □ 1/2 X(t,a,□)(In □M(□))□1 □N(s)□□b□ {[K(s,□,u(□)□□(□,□))□
_ K(s,□,u(□))]□e □□ 1(s□□) □ 1 s□ X(s,□,□)[K(s,□,u(□)□□(□,□))□ K(s,□,u(□))□
K(□,□,u(□)□□(□,□))□ K(□,□,u(□))]d□ □
□□
}ds;

|| H_1(t,□,□(□,□),□) b1/2 ||□||X (t,a,□)M0N(s)□{[K(s,□,u(□) □□(□,□))□
K(s,□,u(□))]□
□ □
1
—
□□(s □t) — □ 1 s□ X (s,□,□)[K(s,□,u(□) □□(□,□))□ K(s,□,u(□))□ K(□,□,u(□)
□□(□,□))□
□e
□□

□ K(□,□,u(□))]d□}ds ||□□ 12— e□□(t □a)M0N0□b□{M1(s □□)e□□(s □ □) □□ —1 □s□e□ 1 (s □ □) □

□(s□□)d□}|□(□,□) | ds ||□ 1 e□□ 1(t □ a)M0 □N0M1(b□a) (14)
□

Из формулы (9) имеем

1 1 t

*H*2(*t*,□,□(□,□),□) □ □□ _[*K*(*t*,*s*,*u*(*s*) □□(*s*,□)) □ *K*(*t*,*s*,*u*(*s*))]□ □ ___2 □□X(t,□,□)[K(□,□,*u*(□) □□(*s*,□))-

-K(□,□,*u*(□))]d□□

1

□ _(*t* □□)

□□ _1[*K*(*t*,*s*,*u*(*s*)□□(*s*,□))□*K*(*t*,*s*,*u*(*s*))]*e* □ □

□

1 t

□ ___2 □X(t,□,□)[*K*(*t*,*s*,*u*(*s*)□□(*s*,□))□*K*(*t*,*s*,*u*(*s*))□*K*(□,□,*u*(□)□□(□,□))□ *K*(□,□,*u*(□))]*d*□

□ *s*

Отсюда

1

□ _(*t* □□)

|| *H*2 (*t*,□,□(□,□),□) ||□|| □ _1 [*K*(*t*,*s*,*u*(*s*)□□(*s*,□))□*K*(*t*,*s*,*u*(*s*))]*e* □ □

□

1 t

□ 2 □ X(t,□,□)[*K*(*t*,*s*,*u*(*s*) □□(*s*,□)) □ *K*(*t*,*s*,*u*(*s*)) □ *K*(□,□,*u*(□) □□(□,□)) □ *K*(□,□,*u*(□))]*d*□||□

□ *s*

1 1 1

– 1 □ □ _ (*t* □*s*) □ 1 t□*e*□ (*t* □ □)M (t - – 1 □ _ (*t* □*s*)

(*t* □*s*)*e*

□ □*M*1 □2 *s* 1 □)d□□ □*M*1(*t* □*s*)*e* □ □

1

–

□ *M*1 [t□*e*□ (*t* □ *s*)(t -s)ds □*M*1(1□*e*□ 1(*t*□*s*)) □ *M*1 , (t,s)□G (15)

□

□

Из формулы (10) имеем

□(*t*,□) □□1(*t*,□)□□2(*t*,□) ,

где

$$\Phi_1(t,\lambda) = \frac{1}{\lambda} X(t,a,\lambda)(In - M(\lambda))^{-1} \int_a^b N(s)[u(s) - u(a)]ds +$$

$$+ \frac{1}{2} X(t,a,\lambda)(In - M(\lambda))^{-1} \int_a^b N(s)X(s,\lambda,\lambda)[u(\lambda) - u(a)]d\lambda ds,$$

$$\Phi_2(t,\lambda) = [u(t) - u(a)] - \frac{1}{\lambda} \int_a^t X(t,s,\lambda)[u(s) - u(a)]ds.$$

Преобразуем вектор – функцию $\Phi_1(t,\lambda)$:

$$\Phi_1(t,\lambda) =$$

$$= X(t,a,\lambda)(I_n - M(\lambda))^{-1} \int_a^b N(s)\{[u(s) - u(a)] - \frac{1}{\lambda} \int_a^s X(s,\lambda,\lambda)[u(\lambda) - u(s)]d\lambda - [u(s) - u(a)] -$$

$$- \frac{1}{\lambda} \int_a^s X(s,\lambda,\lambda)[u(\lambda) - u(s)]d\lambda - [u(s) - u(a)]e^{\frac{1}{\lambda}(s-a)}\}ds = X(t,a,\lambda)(In - M(\lambda))^{-1} -$$

$$- \int_a^b N(s)\{\frac{1}{\lambda} \int_a^s X(s,\lambda,\lambda)[u(\lambda) - u(s)]d\lambda - [u(s) - u(a)]e^{\frac{1}{\lambda}(s-a)}\}ds. \quad a < a$$

$$\|\Phi_1(t,\lambda)\| \le \frac{1}{|\lambda|} e^{\alpha(t-a)} M 0 b$$

$$N 0 L \{$$

$$a$$

$$u - s \le, du - \le d$$

$$\frac{1}{\lambda}$$

$$\left\| \int_a^t \left\{ \int_a^s e^{\alpha(s-\tau)}(\alpha-\beta)d\tau - (s-a)e^{\alpha(s-a)} \right\} ds \right\| \le \int_a^t \left| e^{\alpha(s-\tau)}d\tau - (s-a) \right| v \le e^{\alpha(s-a)}$$

$$\le \frac{1}{\alpha}(s-a) \quad v \le e^{\alpha(s-a)} \quad \frac{1}{\alpha}(t-a) \quad b \le$$

$$\le e^{\alpha(t-a)} \quad M0 \le N0L\left(\frac{1}{\alpha} - e^{\alpha(t-a)}\right)ds \le e^{\alpha(t-a)} \quad M0N0L(b-a) \qquad (16)$$

Преобразуем вектор – функцию $\Phi 2(t,\alpha)$

$$\Phi 2(t,\alpha) = \alpha[u(t) - u(a)] + \int_a^t X(t,s,\alpha)[u(s) - u(a)]ds = \alpha[u(t) - u(a)]$$

$$= \int_a^t X(t,s,\alpha)[u(s) - u(t)]ds + \int_a^t X(t,s,\alpha)[u(t) - u(a)]ds = \alpha[u(t) - u(a)]$$

$$= \int_a^t X(t,s,\alpha)[u(s) - u(t)]ds + \frac{1}{\alpha}\left(e^{\alpha(t-a)} - 1\right)[u(t) - u(a)] - \int_a^t X(t,s,\alpha)[u(s) - u(t)]ds$$

$$\Phi \frac{1}{\alpha}(t-a)$$

$$\le e^{\alpha(t-a)} - 1 \le [u(t) - u(a)].$$

Имеем оценку

$$\Phi \frac{1}{\alpha}(t-a) \le$$

$$\le \Phi 2(t,\alpha) \le \alpha \int_a^t X(t,s,\alpha)[u(s) - u(t)]ds + [u(t) - u(s)]e$$

$$\Phi \frac{1}{\alpha}(t$$

$$\le L\{(t - s)e^{\alpha(t-s)} - s) \mid ss - ta - at - e^{\alpha(t-s)}ds - (t - a)e^{\alpha(t-a)}\}$$

$$\square\underline{1}(t \qquad \underline{1}$$

$$\square| L\{\square(t\ \square a)e\ \square\ \square a)\ \square\ t\square e\square\square(t\ \square\ s)ds\ \square(t\ \square a)e\square\square\underline{1}(t\square a)\}\square$$

$$L\square(1\square e\square\square\underline{1}(t\square a))\ \square\ a$$

$$\square\ L\square; \qquad (17)$$

Так как $|\square(t,\square)|\square|\square 1(t,\square)|\ \square\ |\square 2(t,\square)|$, то

$$|\square(t,\square)|\square\ (\overset{M}{0}\ \overset{N}{0}L(b\square a)e^{\square\ \square\ L\square}), \text{ где } t\ \square[a,b] \qquad (18)$$

Лемма 1 доказана.

Список литературы

1. Лаврентьев М.М. Об интегральных уравнениях первого рода. Докл.АН СССР. – 1959 -Т.127, № 1-с.31-33 2. Иманалиев М.И. Методы решения нелинейных обратных задач и их приложения. – Фрунзе: Илим, 1977 – 348 с.

3. Иманалиев М.И., Асанов А.А. Регуляризация, единственность и существование решения для интегральных уравнений Вольтерра первого рода. //Исследования по интегро – дифференц. уравнениям. – Фрунзе: Илим, 1988 –Вып. 21.-с.3-38

4. Саадабаев А.С. Оценка точности приближенного решения интегрального уравнения первого рода в равномерной метрике.// там же - с. 77-83.

УДК 517

РЕГУЛЯРИЗАЦИЯ РЕШЕНИЯ СИСТЕМЫ НЕЛИНЕЙНОГО ИНТЕГРАЛЬНОГО УРАВНЕНИЯ ФРЕДГОЛЬМА ПЕРВОГО РОДА

Сапарова Гульмира Баатыровна *Ошский Технологический Университет, кафедра «Прикладная математика», г. Ош E-mail:gulya141005@mail.ru*

Построена регуляризация решения системы нелинейного интегрального уравнения Фредгольма первого рода в пространстве С[*a,b*].

Ключевые слова: интегральные уравнения, первого рода, единственность, резольвента, разрывное ядро, решение, существование, условие Липшица, регуляризация.

REGULARIZATION SOLUTION SYSTEM OF NON – LINED INTEGRAL EQUATION OF FREDGOLM OF FIRST TYPE.

Gulmira Saparova *Osh Technological University, chair "Applied Mathematics", Osh city*
E-mail:gulya141005@mail.ru

Partial regularization of solutions system of non – lined integral equation of Fredgolm of first type with discontinuous kernel in sphere C[a,b] is provid

Key words: *integral equation of the first kind, discontinuous kernel, solution, existence, uniqueness, resolvent, Lipschitz condition, regularization.*

Рассмотрим систему

$$\int_a^t H(s)u(s)ds + \int_t^b N(s)u(s)ds + \int_a^t K(t,s,u(s))ds = g(t) \quad t\in[a,b], \qquad (1)$$

где $H(s),N(s)$ – n x n – мерные известные матричные функции, K(t,s,u(s)) – известная непрерывная n – мерная вектор-функция, g(t) и u(t) – известная и искомая функции.

Наряду с системой (1) будем рассматривать систему

$$v(t,\varepsilon) + \int_a^t H(s)v(s,\varepsilon)ds + \int_t^b N(s)v(s,\varepsilon)ds + \int_a^t K(t,s,v(s,\varepsilon))ds = g(t) + \varepsilon u(a), \qquad (2)$$

где u(t) – решение системы (1), $0 < \varepsilon$ - малый параметр.

Введем обозначения:

3) $\langle\cdot,\cdot\rangle$ - скалярные произведения в R^n, ||A||, ||u|| - нормы соответственно n x n – мерной матрицы A и n – мерного вектора u, то есть для любых

$$u=(u1,u2,.....,un), v=(v1,v2,.....,vn)\in R^n, \langle u,v\rangle = u1v1+u2v2+.....+unv_n,$$

$$\|u\| = \sqrt{\langle u,u\rangle};$$

4) $C_n[a,b]$- пространство n – мерных векторов с элементами из C[a,b], $\|\cdot\|_c$ -норма в $Cn[a,b]$, то есть для любого $u(t)\in Cn[a,b]$

$$\| u(t) \|_c = \max_{t\in[a,b]} \| u(t) \|$$

Предполагаем выполнение следующих условий:

а) $\det[H(t) + N(t)] \neq 0$ при всех t$\in$[a,b]. Не ограничиваясь общности будем считать $(H(t)+N(t)) = I_n$ при всех t$\in$[a,b], где I_n $-nxn$- мерная единичная матрица. Так как в случае $\det[H(t) + N(t)] \neq 0$, t$\in$[a,b] в системе (1) можем сделать замену

$$u(t) = [H(t)+N(t)]^{-1}u1(t);$$

б) при $t \neq \tau$ для любых $(t,s,u1),(\tau,s,u1),(t,s,u2),(\tau,s,u2)\in G\subset R$

справедлива оценка:

$$|K(t,s,u1)-K(\tau,s,u1)-K(t,s,u2)+K(\tau,s,u2)|\le M1(t-\tau)(u1-u2)$$

где $G=\{(t,s): a\le s\le t\le b\}$, $0 < M_1$ – известная постоянная.

Предположим выполнения следующих условий:

в) $\det(I_n - M(\lambda)) \neq 0$ при $\lambda > 0$, где

$$M(\lambda) = \int_a^b N(s)X(s,a,\lambda)ds, X(t,s,\lambda) = I_n e^{\lambda(t-s)};$$

г) $\lambda_i(t), i=1,2,.....,n$ собственные значения матрицы $\frac{1}{2}[H(t)+N(t)+H^*(t)+N^*(t)]$ и $\min\lambda_i(t)+\lambda_j(t)\neq 0$, при $t\in[a,b]$;

д) K(t,t,u)=0, для любых (t,u)$\in$[a,b]xR;

е) K(t,s,0)=0, при (t,s) $\in$G={(t,s): $a \le s \le t \le b$} и

$$\|X(t,s,\lambda)\| \le e^{N_0(t-s)}$$

$$N_0 = \sup_{t\in[a,b]}\|N(s)\|$$

Подставляем в систему (2) формулу

$$v(t,\lambda)+u(t)=\varphi(t,\lambda) \qquad (3)$$

где u(t) – решение системы (1).

$$\varphi(t,\lambda)+\int_a^t H(s)\varphi(s,\lambda)ds+\int_a^t N(s)\varphi(s,\lambda)ds+\int_a^t [K(t,s,u(s)+\varphi(s,\lambda))-K(t,s,u(s))]ds$$

$$+\int_a^t [u(t)-u(a)].$$

Вводим обозначения

$$g(t,\lambda)=\int_a^t [K(t,s,u(s)+\varphi(s,\lambda))-K(t,s,u(s))]ds+\int_a^t [u(t)-u(a)] \qquad (4)$$

Отсюда применяя формулу Дирихле, имеем

$$\varphi(t,\lambda)=\frac{1}{2}X(t,a,\lambda)(In-M(\lambda))^{-1}\int_b\int_b N(s)[K(s,\tau,u(\tau)+\varphi(\tau,\lambda))-$$

$$K(s,\tau,u(\tau))]d\tau ds+$$

□ *a*□

□ _ 1 *X*(*t*,*a*,□)□(*In* □*M*(□))□1*b*□ *N*(*s*)[*u*(*s*)□*u*(*a*)]*ds* □

□ *a*

□ 1 3 *X*(*t*,*a*,□)□(*In* □*M*(□))□1*b*□*b*□ *s*□

N(*s*)*X*(*s*,□,□)□[*K*(□,□,*u*(□)□□(□,□))□*K*(□,□,*u*(□))]*d*□

□ *a*□□

dsd□□□ 1 2 *X*(*t*,*a*,□)□(*In* □*M*(□))□1*ba*□*as*□*N*(*s*)*X*(*s*,□,□)[*u*(□)□*u*(*a*)]*d*□*ds* □

□ 1 *t*□*K*(*t*,*s*,*u*(*s*)□□(*s*,□))□ *K*(*t*,*s*,*u*(*s*))*ds* □[*u*(*t*)□*u*(*a*)]□

□*a*

□ 12 *at*□□*t*□ *X*(*t*,*s*,□)□[*K*(*s*,□,*u*(□)□□(□,□))□ *K*(*s*,□,*u*(□))]*d*□*ds* □□ 1 *at*□

X(*t*,*s*,□)□[*u*(*s*)□*u*(*a*)]*ds*

□

(5)

Введем, новые обозначения $H_1(t,□,□(□,□),□)$, $H_2(t,□,□(□,□),□),□(t,□)$, и получаем новое уравнение

$$□(t,□) \quad □ \quad a□H_1(t,□,□(□,□),□)d□□_a□H_2(t,□,□(□,□),□)d□□□(t,□)$$

(6) где

$H_1(t,□,□(□,□),□)$ □ 1 2 X(t,a,□)(I_n -M(□))-1b□ N(s)[K(s,□,*u*(□) □□(□,□))-

K(s,□,u(□))]ds-

□ □

$$-\frac{1}{3}X(t,a,\lambda)(In-M(\lambda))^{-1}\int_b\int_s N(s)X(s,\lambda,\lambda)[K(\lambda,\lambda,u(\lambda)\lambda\lambda(\lambda,\lambda))\lambda K(\lambda,\lambda,u(\lambda))]d\lambda ds, \quad (9)$$

$$H2(t,\lambda,\lambda(\lambda,\lambda),\lambda) \lambda \lambda\frac{1}{}[K(t,s,u(s)\lambda\lambda(s,\lambda))\lambda K(t,s,u(s))]\lambda$$

$$\lambda \frac{1}{2}\int_1^t X(t,\lambda,\lambda)[\lambda K(\lambda,\lambda,u(\lambda)\lambda\lambda(\lambda,\lambda))\lambda K(\lambda,\lambda,u(\lambda))]d\lambda , \quad (7)$$

$$\lambda(t,\lambda)\lambda\lambda\frac{1}{}X(t,a,\lambda)(In\lambda M(\lambda))^{\lambda 1}\int_b a\lambda N(s)[u(s)\lambda u(a)]ds\lambda$$

$$\lambda \frac{1}{2}X(t,a,\lambda)(In\lambda M(\lambda))^{\lambda 1}\int_b\int_s N(s)X(s,\lambda,\lambda)[u(\lambda)\lambda u(a)]d\lambda ds\lambda$$

$$\lambda[u(t)\lambda u(a)]\lambda\frac{1}{}\int_a^t X(t,s,\lambda)[u(s)\lambda u(a)]ds, \quad (8)$$

ЛЕММА 1. Пусть выполняются условия а), б), в), и г) $\|(In\lambda M(\lambda))\lambda 1\| \lambda M0$, где известная постоянная $0\lambda M_0$ не зависит от $\lambda\lambda 0$, $H1(t,\lambda,\lambda(\lambda,\lambda),\lambda), H2(t,\lambda,\lambda(\lambda,\lambda),\lambda),\lambda(t,\lambda)$ определены по формулам (6),(7),(8), тогда справедливы следующие оценки:

$$\|H1(t,\lambda,\lambda(\lambda,\lambda),\lambda)\|\lambda\lambda 1^{1}\lambda^{\lambda}(t\lambda a)M_0N_0M_1(b\lambda a)e \quad \text{при всех } t,\lambda\lambda[a,b]; \quad (9)$$

$$\|H2(t,\lambda,\lambda(\lambda,\lambda),\lambda)\|\lambda M_1 \quad \text{при всех } (t,\lambda)\lambda G ; \quad (10)$$

$$\|\lambda(t,\lambda)\|\lambda(M_0N_0L(b\lambda a)e^{\lambda\lambda\frac{1}{}(t}a)\lambda L\lambda), \text{при всех } t\lambda[a,b] \quad (11)$$

где

$$N0 = \sup_{t \in [a,b]} |N(s)|,$$

а функция u(t) удовлетворяет условию Липшица с коэффициентом L, то есть для любых t,s ∈ [a,b], при t > s справедлива оценка

$$|u(t) - u(s)| \le L(t - s)$$

ТЕОРЕМА 1. Пусть выполняются условия а), б), в) и система (1) имеет решение $u(t) \in Cn[a,b]$ и

$$N1 = M0M1N0(b-a)e^{M1(b-a)} < 1,$$

$$M0 = (In - M(\lambda))^{-1} \ne 0$$

$$N0 = \sup_{t \in [a,b]} |N(t)|$$

Тогда, $v(t,\lambda)$ - решение системы (2) при $\lambda \to 0$ сходится по норме L[a,b] к u(t). При этом справедлива оценка

$$\int_a^b |v(t,\lambda) - u(t)|dt \le \frac{1}{(1 - N1)}(M0N0L\lambda(b-a) + L\lambda)eM^1(b - a) \qquad (12)$$

ДОКАЗАТЕЛЬСТВО: Получив оценки функций $H_1(t,\lambda,\lambda(\lambda,\lambda),\lambda), H_2(t,\lambda,\lambda(\lambda,\lambda),\lambda),\lambda(t,\lambda)$, имеем

$$\int_a^t |\lambda(t,s)|ds \le M0M1N0(b-a)\int_a^b |\lambda(\lambda,\lambda)|d\lambda \le M1\int_a^b (a|\lambda(\lambda,\lambda)|d\lambda)dt \le$$

(13)

$$\le (\mathrm{M}0N0L\lambda(b-a) + L\lambda) \qquad t \in [a,b]$$

где

$$\sup_{t\in[a,b]} |N(t)| \le N0$$

$$N3 = M0M1N0(b-a)a\int|\Gamma(\tau,\tau)|\,d\tau$$

$$N4 = M0N0L(b-a) + L\Gamma$$

Тогда, имеем

$$a\int|\Gamma(t,s)|\,ds \le M1a\left(a\int|\Gamma(\tau,\tau)|\,d\tau\right)dt + N3 + N4, \quad t\in[a,b]$$

После применения неравенства Гронуолла-Беллмана, имеем

$$a\int|\Gamma(t,s)|ds \le \{M0M1N0(b-a)a\int|\Gamma(\tau,\tau)|d\tau + (M0N0L(b-a) + L\Gamma)\}eM^{1(b-a)} \qquad (14)$$

Накладываем условия

$$M1(b-a) < 1, \quad N1 = M0M1N0(b-a)e$$

где $M0 = (In - M(\Gamma))^{-1} > 0$

Тогда из (13), имеем

$$(1- N1)a\int|\Gamma(\tau,\tau)|\,d\tau \le (M0N0L(b-a) + L\Gamma)e^{M(b-a)}$$

Отсюда получаем оценку (12).
Теорема 1 доказана.

Список литературы

5. Лаврентьев М.М. Об интегральных уравнениях первого рода. Докл.АН СССР. – 1959 -Т.127, № 1-с.31-33

6. Иманалиев М.И. Методы решения нелинейных обратных задач и их приложения. – Фрунзе: Илим, 1977 – 348 с.

7. Иманалиев М.И., Асанов А.А. Регуляризация, единственность и существование решения для интегральных уравнений Вольтерра первого рода. //Исследования по интегро – дифференц. уравнениям. – Фрунзе: Илим, 1988 –Вып. 21.-с.3-38

8. Саадабаев А.С. Оценка точности приближенного решения интегрального уравнения первого рода в равномерной метрике.// там же - с. 77-83.

9. Сыдыков Т. Приближенное решение интегрального уравнения Фредгольма первого рода в пространстве С(0,1).// Всесоюз. конф. по некорректно поставленным задачам.- Фрунзе: Илим, 1979

10. Асанов А.А, Сыдыков Т., Сапарова Г. Регуляризация интегрального уравнения Фредгольма первого рода с разрывным ядром.//Сб. науч. статей. – Бишкек:ИГЗ КГПУ им. И.Арабаева, 2002. – с. 225-231.

ТРАНСПОРТ И МАШИНОСТРОЕНИЕ

УДК 625.712.1(575.2)

КЫРГЫЗ РЕСПУБЛИКАСЫНДАГЫ ШААР ЧЕТИНДЕГИ КАЛКТУУ АЙМАКТАРЫНДАГЫ ЖОЛ КЫЙМЫЛЫН ИЗИЛДӨӨ

Кадыров Эрмек Тургамбаевич, *И.Раззаков атындагы Кыргыз мамлекеттик техникалык университетинин "Ташууларды уюштуруу жана кыймыл коопсуздугу" кафедрасынын окутуучусу. Бишкек ш. kadet-dosoi@mail.ru*

Макалада Кыргыз Республикасындагы шаар четиндеги калктуу аймактарындагы жол кыймылынын изилдөө боюнча суроолор каралган.

Түйүндүү сөздөр: жол кыймылын изилдөө, жол кыймылынын изилдөөсүнүн классификациясы, кыймыл коопсуздугу, шаар четиндеги калктуу аймактардагы жол кыймылынын көйгөйлөрү.

ИССЛЕДОВАНИЕ ДОРОЖНОГО ДВИЖЕНИЯ В ПРИГОРОДНЫХ НАСЕЛЕННЫХ ПУНКТАХ КЫРГЫЗСКОЙ РЕСПУБЛИКИ

Кадыров Эрмек Тургамбаевич, *Преподаватель кафедры «Организации перевозок и безопасность движения» Кыргызский государственный технический университет им. И. Раззакова, г. Бишкек. kadet-dosoi@mail.ru*

В статье рассмотрены вопросы по проведенному исследованию дорожного движения в пригородных населенных пунктах Кыргызской Республики.

Ключевые слова: исследование дорожного движения, классификация исследований дорожного движения, безопасность движения, проблемы безопасности движения в пригородных населенных пунктах.

A STUDY OF THE ROAD IN A SUBURBAN TOWNS OF THE KYRGYZ REPUBLIC

Kadyrov Ermek Turganbaevich, *Lecturer of the Department "Organization of transportation and traffic safety"Kyrgyz State Technical University. I. Razzakova, Bishkek. kadet-dosoi@mail.ru*

In the article the questions in the survey of the road in a suburban towns of the Kyrgyz Republic.

Keywords: traffic study, classification of traffic studies, traffic safety, the problems of traffic safety in suburban settlements.

Азыркы учурда Кыргыз Республикасынын жол тармагы унаа агымынын интенсивдүүлүгүнө туруштук бере албай келет, ошондой эле айдоочулардын жол кыймылынын эрежесине кайдыгерлиги, совет доорундагы иштелип чыккан эффективдүү эмес жол кыймылынын уюштуруу схемалары жана автожолдорунун элементтери заманбап унаалардын кыймыл шартына ылайыкталбаганы өндүү көйгөлөрү бар.

Изилдөө – жол кыймылы үчүн зарыл базис, изилдөөсүз жол кыймылынын иштешин жана өнүгүшүн элестетүү кыйын. Туура чечим чыгарууга изилдөөдөн гана алынган толук жана так маалымат болушу керек

1-сүрөттө керектүү маалымат алуу ыкмасына негизделген кеңири колодонулган жол кыймылынын мүнөздөрүн жана шарттарын изилдөө ыкмалары колдонулган[1, 78-79 б].

1-Сүрөт. Жол кыймылын изилдөө ыкмаларынын классификациясы

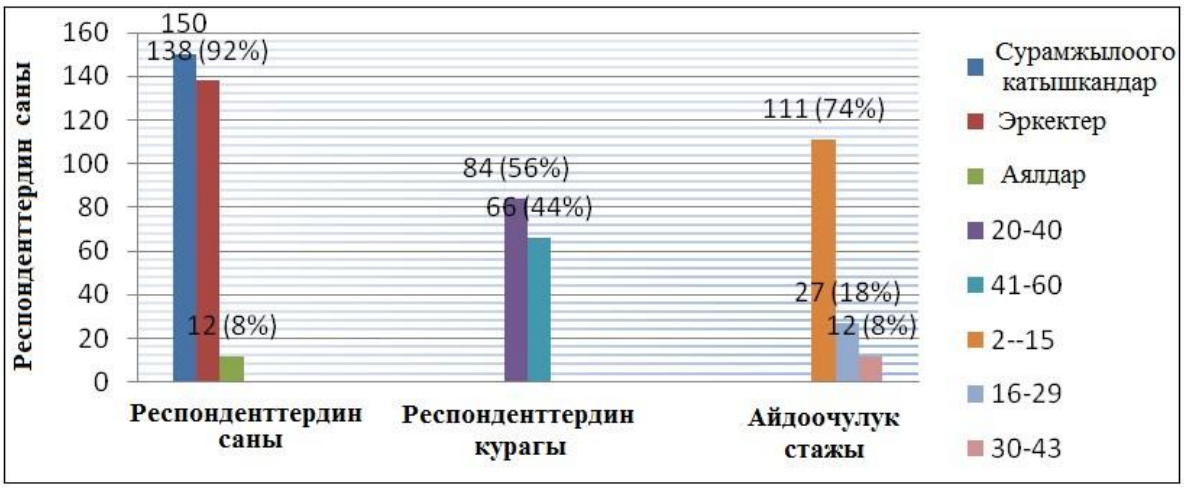
Документалдык изилдөө. Материалды изилдөө обьектине чыкпастан текшерүү (камералдык шарттарда).

Натурдук изилдөө. Аныкталган убакытта болуп жаткан жол кыймылынын айкын көрсөткүчтөрүн жана шарттарын аныктоо. Кеңири таралган, ар түрдүүлүгү менен айырмаланат, жол кыймылынын абалы тууралуу анык маалыматты алуучу жалгыз ыкма болуп саналат

Жол кыймылынын процессинин моделдөө, унаа агымынын баяндоо, сүрөттөө үчүн математикалык ыкмаларды колдонууда негизделет.

Жол кыймылынын заманбап абалын түздөн-түз жол кыймылынын катышуучулары – унаа каражаттарынын айдоочулары сезишет. Ушул айдоочулардын арасында сурамжылооанкеталык изилдөө жүргүзүлгөн. Бишкек шаарына күнүнө экиден кем эмес жолу каттаган айдоочулар тандап алынып, сурамжылоого катышышты. Анкетада, айдоочуга, унаа каражатына тиешелүү жана кыймылдын каттамын мүнөздөөчү суроолор камтылган.

Жүз элүү айдоочу сурамжылоого катышышты, басымдуу бөлүгүн шаар четиндеги каттамдар боюнча жеке ташуулар менен иш алып барган айдоочулар түздү, алардын ичинде менеджерлер, мугалимдер, жеке ишкерлер ж.б. бар. Респонденттердин курагы эркектердики 20дан 60 жашка чейин жана аялдардыкы 27ден 43 жашка чейин куракты түздүү Айдоочулук стажы эркектердики 27ден 43 жылга чейин жана аялдардыкы 2ден 9 жылга чейин (2-сүрөт).



2-сүрөт. Сурамжылоонун натыйжалары

Ошондой эле унаа каражаттарынын түрү, чыгарылган жылы жана узата огуна жараша рулдун жайгашканы эске алынды. Автомобилдер үч шарттуу топко бөлүндү: 1. Жеңил автомобилдер

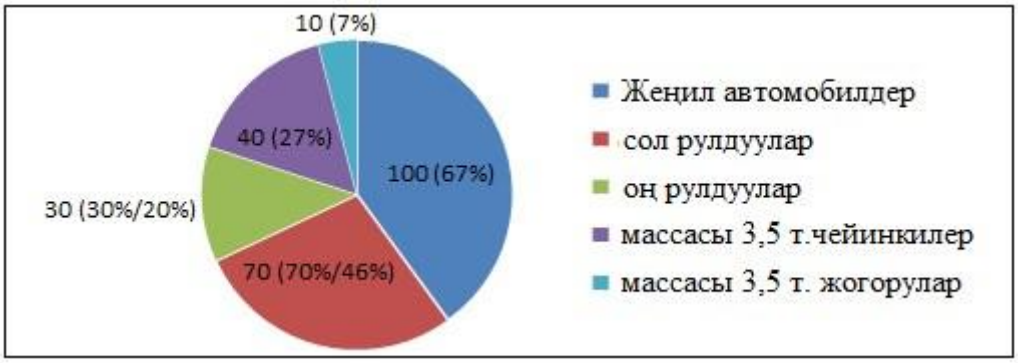
1.1. Автомобилдин сол тарабында жайгашкан рулдуулар (сол рулдуулар);

1.2. Оң рулдуулар.

2. Максималдуу массасы 3,5 тга чейинкилер, микроавтобустар кошулган;

3. Максималдуу массасы 3,5 тдан жогорулар

Жеңил автомобилдердин саны 100 даананы түздү, башкача айтканда изилдөөгө катышкан автомобилдердин жалпы санынын 67% анын ичинен 30у оң рулдуулар же болбосо изилдөөгө катышкан автомобилдердин жалпы санынын 20% же жеңил автомобилдердин 30% Максималдуу массасы 3,5 тга чейинкилер – экинчи топтун автомобилдери – 40 даана (27%) жана максималдуу массасы 3,5 тдан жогорулар10 даана (7%) (3-сүрөт).

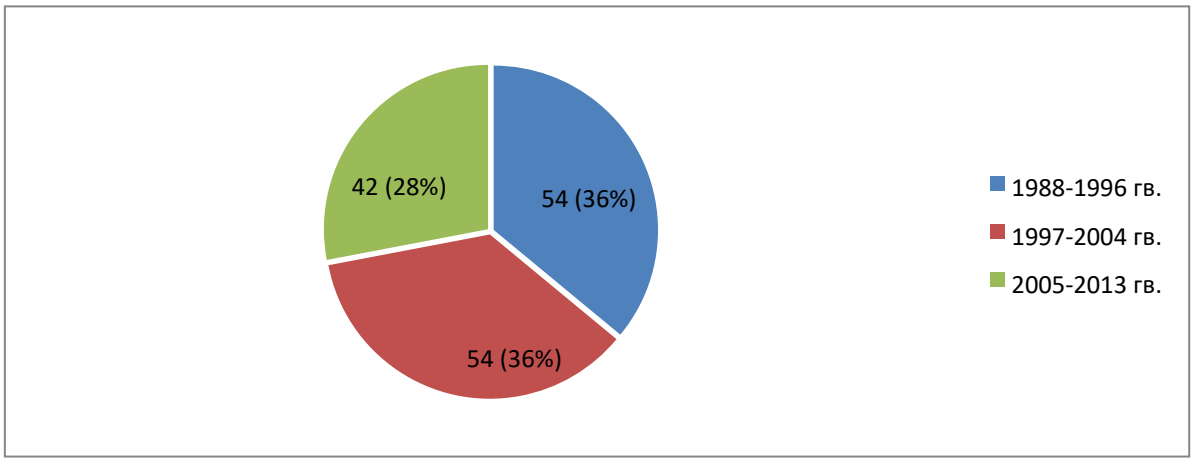


70 (70%/46%) – сол рулдуу автомобилдердин саны (% жеңил автомобилдердин санынан / % автомобилдердин жалпы санынан);

30 (30%/20%) – оң рулдуу автомобилдердин саны (%жеңил автомобилдердин санынан / % автомобилдердин жалпы санынан).

3-сүрөт. Изилдөөгө катышкан автомобилдердин саны

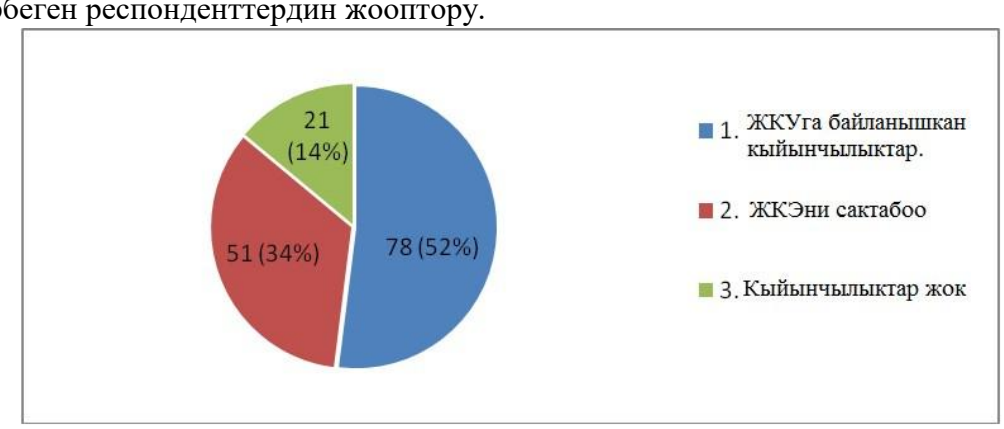
Автомобилдердин чыгарылган жылдары 1988-жылдан 2013-жылга чейинки аралыкты түздү. 1988-1996 жана 1997-2004 жылдары чыккан автомобилдердин саны ар бир аралыкта 54төн – 36%, ал эми 2005-2013 жылкы аралыктагы унаалардын саны 42 даананы автомобиддердин жалпы санынан 28% түздү(4-сүрөт).



4-сүрөт. Автомобилдердин чыгарылган жылы боюнча саны.

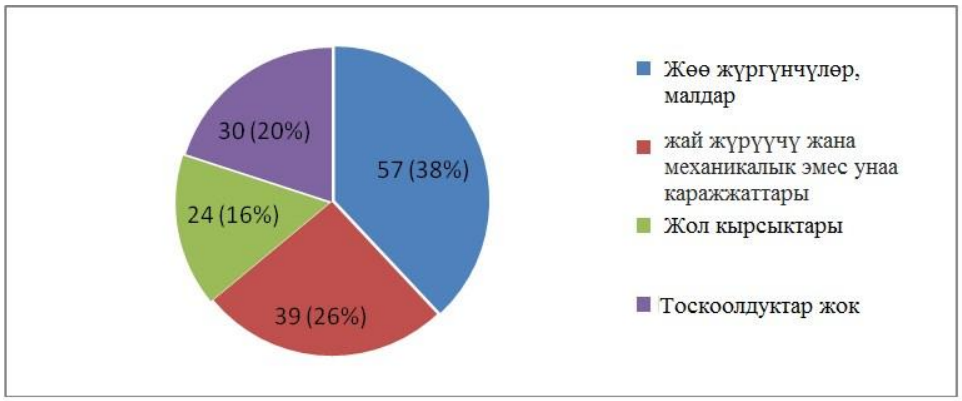
Каттамда (жолдо) кандай кыйынчылыктар пайда болот? Атту суроого респондеттердин көпчүлүгү жолдун начар абалын жана жол кыймылынын катышуучуларынын ылдамдык режимин жана жол кыймылынын эрежесинин жөнөкөй талаптарын сактабоосо, ошондой эле жолдун туурасынын жетишсиздигинен келип чыккан унаа тыгындарын белгилеп кетишти. Жөө жүргүнчүлөрдүн тоскол болуусу, начар көрүнүмдүүлүк жана аба ырайына байланышкан тоскоолдуктар деген жооптор жок эмес. Берилген суроого жоопторду иштеп чыгуу максатында, респонденттердин суроосун үч шартту топко бөлдүк (5-сүрөт):

1. Жол кыймылын уюштуруудагы (ЖКУ) кыйынчылыктар. Бул топко жолдун туурасынын кууштугу, унаа тыгындары, жол шарттарынын ылайыкталбаганы, жол кыймылын уюштуруучу техникалык каражаттарга (ЖКУТК) (жол белгилери, жол чийиндери) байланышкан кыйынчылыктар ж.б. киргизилген.
2. Жол кыймылынын эрежесинин (ЖКЭ) сактабоосо. Бул топто ылдамдык режимин сактабоо, жол кыймылынын катышуучуларынын ЖКЭнин сактабоосу, жүргүнчүлөрдүн тоскол болуусу (унаа ичинде жана сыртында жүргүнчүлөрдүн тоскоол болуусун микроавтобустун айдоочулары белгилеп кетишти) ж.б.
3. Кыйынчылыктар жок. Кыйынчылыктар байкалбаганын айткан жана суроого жооп бербеген респонденттердин жооптору.



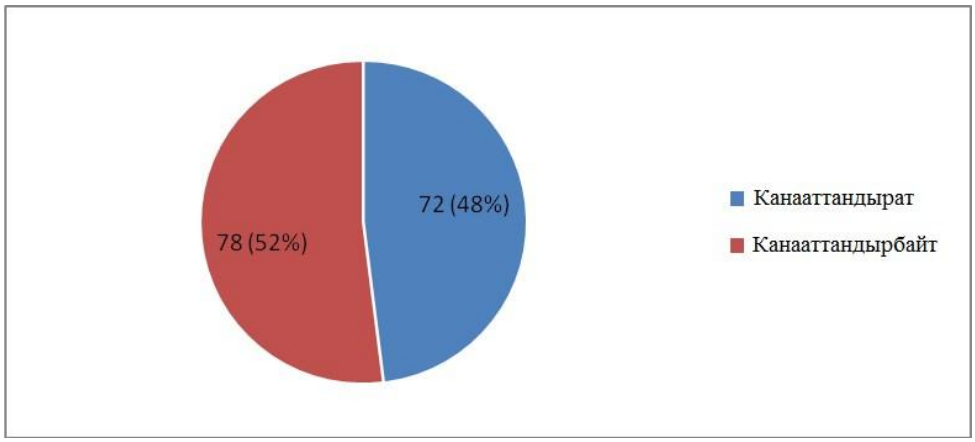
5-сүрөт. Каттамдагы кыйынчылыктарга байланышкан респондеттердин жооптору

Жолдогу тоскоолдуктары аныктөөгө көздөлгөн кийинки суроого жооптор төмөнкүдөй болду: жөө жүргүнчүлөр, малдар, айыл-чарба техникасы, жай жүрүүчү жана механикалык эмес унаа каражжаттары, жол кырсыктары, жол кайгуул кызматынын посттору ж.б.у.с. (6-сүрөт)



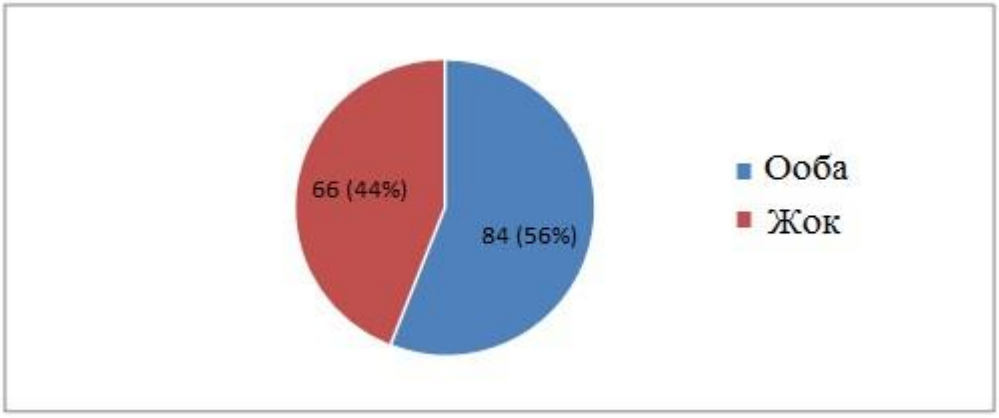
6-сүрөт. Каттамдагы тоскоолдуктарга байланышкан респондеттердин жооптору

Азыркы учурдагы жол шарттары катышуучулардын 48% - 72 адамды канааттандырат жана 52% - 78 адамды канааттандырбайт. (7-сүрөт).



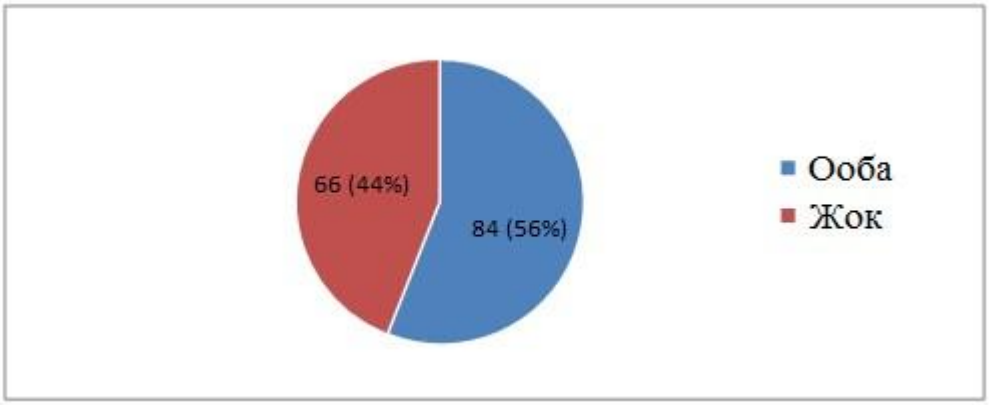
7-сүрөт. “Жол шарттары канааттандырабы?” суроосуна респондеттердин жообу

Каттамда коопту жол тилкелери (калктуу аймактардын, мектептердин жанында, белгилүү жерлер) көп кездешеби? Суроосуна айдоочулардын 56% көп кездешет деп жооп беришти. (8-сүрөт).



8-сүрөт. “Каттамда коопту жол тилклери көп кездешеби?” суроосуна жооптор

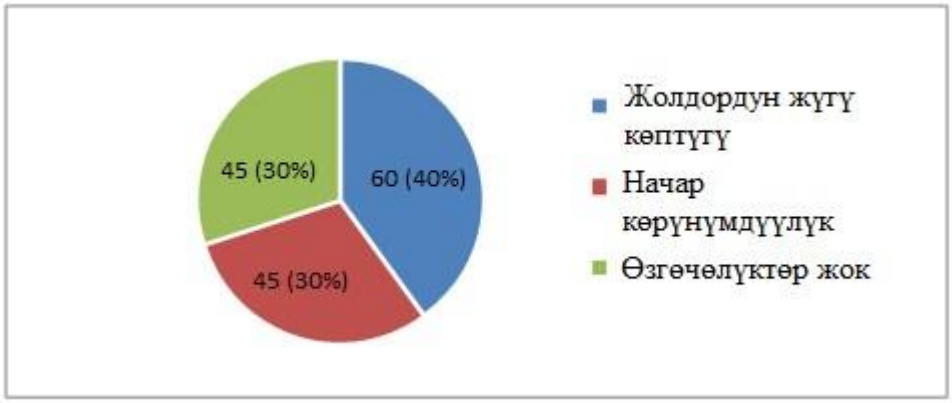
Заманбап автомобилдердин эксплуатациялык касиеттери Кыргыз Республикасынын автожолдорунун техникалык мүнөздөрүнө ылайык келбегени сыр эмес, ушуга айдоочулардын 56% макулдугун айтып өтүштү. (9-сүрөт)



9-сүрөт. “Сиздин автомобилдин эксплуатациялык касиеттерине автожолдордун техникалык мүнөздөрү ылайык келеби?” суроосуна респондеттердин жооптору

Кыймыл шарттары сутканын убактысы, метеорологиялык шарты боюнча кескин өзгөрөт. Жооптор кийинки критерийлер боюнча топтолду (10-сүрөт):

1. Жолдордун жүгү көптүгү. Транспорт тыгындары и т.п.
2. Көрүнүмдүүлүктүн начарлыгы. Карама каршы багыттагы автомобилдердин фарасынын жарыгы, күндүн жарыгы ж.б.
3. Өзгөчөлүктөр жок.



10-сүрөт. “Эртең мененки жана кечки убакыттагы каттамдагы жүрүү өзгөчөлүктөру?” суроосуна респондеттердин жообу

Сурамжылоого катышкан айдоочулардын көпчүлүгү, үчөөнөн башкасы, калктуу аймака жакындаганда ылдамдыкты төмөндөткөндүгүн жооп катары айтышты, себеби ЖКЭнин талаптарын (ылдамдык режимин) сактоосун белгилеп кетишти. Акыркы ушул көйгөйлөрдү чечүүчү кандай чечимдерди, жолдорду сунуштайсыз дегенде, айдоочулар бийликтин аракетсиздигинен баштап жолдорду заманбап талаптар боюнча кайра калыбына келтирүүсүнө чейин жоопторунда көрсөтүштү.

Өткөрүлгөн изилдөө Кыргыз Республикасындагы шаар четиндеги калктуу аймактарындагы жол кыймылынын абалын сүрөттөйт жана жол кыймылынын коопсуздугун жогорулатуу, өркүндөтүү боюнча ыкчам ж.б. чараларды көрүүгө мүмкүнчүлүк берет. Изилдөөнүн натыйжалары баалуу жана окуу, практикалык максатта колдонулушу толук мүмкүн.

Колдонулган адабияттардын тизмеси

1. Клинковштейн Г. И., Афанасьев М. Б. Организация дорожного движения: Учеб. для вузов.– 5-е изд., перераб. и доп. – М: Транспорт, 2001 – 247 с.

УДК 681.62

АКТУАЛЬНЫЕ ПЕРСПЕКТИВЫ ИССЛЕДОВАНИЯ ШУМА И ВИБРАЦИИ ПОСЛЕ ПЕЧАТНОГО ПОЛИГРАФИЧЕСКОГО ОБОРУДОВАНИЯ

Киянбекова Л.Р., *докторант PhD специальности «Машиностроение», КГТУ имени И.Раззакова*

Рассмотрены результаты работы по разработке технических стандартов для полиграфического оборудования. Показано, что стремительное изменения нормативной базы, регламентирующей технические аспекты обеспечения безопасности полиграфического оборудования, обуславливает необходимость совершенствования методики акустических испытаний при проведении лабораторных исследований в рамках декларирования соответствия для послепечатного полиграфического оборудования. Рассмотрено современное научное состояние проблемы виброакустической диагностики и прогнозирования технического состояния полиграфических машин в процессе их эксплуатации. Установлено, что современные исследователи, использующие информационно-компьютерные технологии (такие, например, как вейвлет-анализ) сосредоточили своё внимание на совершенствовании методов виброакустической диагностики и прогнозирования технического состояния диагностики печатных машин, не уделяя достаточного внимания послепечатному полиграфическому оборудованию. Обоснован вывод о перспективности и востребованности проведения исследований шума и вибрации послепечатного полиграфического оборудования в направлениях: а) совершенствование методики акустических испытаний при проведении лабораторных исследований в рамках декларирования соответствия; б) разработка методов виброакустической диагностики и прогнозирования технического состояния в процессе их эксплуатации с использованием современных информационно-компьютерных технологий

Ключевые слова: полиграфическое оборудование, послепечатное оборудование, межгосударственный стандарт, шум, вибрация, виброакустическая безопасность, акустические испытания, виброакустическая диагностика.

ACTUAL PROSPECTS OF RESEARCH OF NOISE AND VIBRATION OF THE POSTPRINTING EQUIPMENT

Kiyanbekova L.R., *PhD doctoral, specialty "Machine building"*

Results of development of technical standards for the printing equipment are considered. It is shown that prompt changes of the regulatory base regulating technical aspects of safety of the printing equipment are caused by need of improvement of a technique of acoustic tests when carrying out laboratory researches within declaring of compliance for the postprinting printing equipment. The current scientific state of a problem of vibroacoustic diagnostics and forecasting of technical condition of printing cars in the course of their operation is considered. It is established that the modern researchers using information and computer technologies (such, for example, as the veyvlet-analysis) have concentrated the attention on improvement of methods of vibroacoustic diagnostics and forecasting of technical condition of diagnostics of printing machines, without paying sufficient attention to the postprinting printing equipment. A valid conclusion about

prospects and a demand of carrying out researches of noise and vibration of the postprinting printing equipment in the directions: a) improvement of a technique of acoustic tests when carrying out laboratory researches within compliance declaring; b) development of methods of vibroacoustic diagnostics and forecasting of technical condition in the course of their operation with use of modern information and computer technologies

Keywords: printing equipment, postprinting equipment, interstate standard, noise, vibration, vibroacoustic safety, acoustic tests, vibroacoustic diagnostics.

Введение

Многообразие факторов, являющихся потенциальными источниками производственного травматизма обслуживающего персонала полиграфического оборудования, обуславливает необходимость разработки специальных требований по безопасности и охране труда для полиграфических предприятий. В РК такие правила были разработаны и приняты в 2005 году [19]; в дальнейшем республика перешла на межгосударственные стандарты, которые в последние годы постоянно совершенствуются. В 2012 году Межгосударственный совет по стандартизации, метрологии и сертификации принял ГОСТ 12.2.231-2012, регламентирующий требования безопасности и методы испытаний полиграфического оборудования, который был введен в действие в качестве национального стандарта РК с 1 июля 2014 г. [4]. Требования безопасности для конструирования и изготовления для полиграфических машины и оборудования. регламентируются ГОСТ EN 1010-1-2011 (общие требования) [3] и ГОСТ EN 1010-3-2011 (машины резальные) [6].

В настоящее время Межгосударственным техническим комитетом по стандартизации обсуждается проект нового варианта ГОСТ EN 1010-1, который планируют принять в 2016 году. Его текст фактически идентичен европейскому стандарту EN 1010-1:2004+A1:2010, регламентирующему требования безопасности для конструирования и изготовления печатных и бумагоперерабатывающих машин. После принятия этого стандарта, действующий межгосударственный стандарт ГОСТ EN 1010-1-2011 подлежит отмене [7].

Шум и вибрация являются одним из основных вредных факторов, которыми сопровождается работа большинства видов полиграфических машин. (Принципиальной разницы между шумом и вибрацией нет: в основе того и другого явления лежат колебательные процессы, но субъективно человек воспринимает шум слухом, а вибрацию осязанием).

В условиях столь стремительного изменения нормативной базы, регламентирующей технические аспекты обеспечения безопасности полиграфического оборудования, перед специалистами, изучающими шум и вибрацию полиграфического оборудования, приходится тщательно отбирать действительно актуальные и востребованные направления исследования. Цель данной работы – определить такие направления.

В соответствии с Техническим регламентом ТР ТС 010/2011 «О безопасности машин и оборудования», полиграфическое оборудование подлежит декларированию соответствия [13]. Процесс декларирования кардинально не отличается от сертификации, процедура декларирования включает в себя этапы подтверждения качества: проведение лабораторных испытаний, оформление протокола испытаний и декларации соответствия.

Акустические испытания являются наиболее сложным и трудоемким видом лабораторных испытаний полиграфического оборудования, существенно влияющим на трудоемкость и стоимость работ. В зависимости от выбора методики акустических испытаний и их условий оборудование может быть признано как соответствующим, так и не соответствующей требованиям безопасности [9]. Как отмечается в ГОСТ 27409-97, посвящённого нормированию шумовых характеристик стационарного оборудования,

предельные значения шумовых характеристик машин зависят от конкретных условий их эксплуатации. Наличия группы одновременно работающих машин, шум от которых оказывает совместное воздействие, величины шума, излучаемого каждой машиной, расположения машин относительно рабочего места и акустических характеристик помещения, в котором они эксплуатируются [5].

О сложности организации акустических испытаний свидетельствует, например, то, что в РФ к настоящему времени разработано 234 стандарта по шуму и вибрации [14]. Принят ГОСТ Р 53479-2009, регламентирующий методы определения шумовых характеристик для полиграфического оборудования, который является модифицированным по отношению к европейскому стандарту ЕН 13023:2003 «Методы измерения шума печатных, бумагоперерабатывающих, бумагоделательных машин и вспомогательного оборудования. Степень точности 2 и 3». Область применения ГОСТ Р 53479-2009 ограничена только печатными и бумагоперерабатывающими машинами [8] - на послепечатное полиграфическое оборудование они не распространяются.

Исходя из этого, мы считаем, что совершенствование методики акустических испытаний при проведении лабораторных исследований в рамках декларирования соответствия послепечатного полиграфического оборудования является одним из проблемных и перспективных направлений исследования шума и вибрации полиграфического оборудования.

Как показывает изучение статистики поломок полиграфического оборудования, 90% отказов возникает из-за скрытых внутренних дефектов (технологических погрешностей изготовления и сборки; дефектов, появляющихся в процессе эксплуатации в результате старения, износа, воздействия вибрации, температуры и т.д.) и только 10% отказов - вследствие неправильной эксплуатации. [12]

Как отмечает Г.Б. Куликов, в результате износа ухудшаются показатели работы элементов (деталей, сопряжений, узлов и агрегатов) машин: снижается точность приводки и качество печати в печатных машинах; нарушается точность фальцовки; ухудшается качество шитья в ниткошвейных автоматах; наблюдается снижение точности выполнения операций по обработке блоков и т.д. [10]

В своё время для решения проблемы надежности и устранения других недостатков системы технического обслуживания и ремонта по потребности была разработана система планово-предупредительных ремонтов (ППР). В ее основу были положены предупреждающие аварии машин плановые обслуживание и ремонт по назначенному заранее ресурсу, которые осуществлялись по мере их наработки.

Однако повышение сложности полиграфического оборудования и увеличение её номенклатуры привело к резкому увеличению расходов на техническое обслуживание и ремонт, выявил все недостатки системы ППР

Возрастающие потребности в повышении качества полиграфического оборудования и производимой на нем продукции привели к усложнению конструкций машин и увеличению скорости их работы. Это повлекло за собой столь значительное увеличение объемов ППР, что сделало эту систему экономически неэффективной. Куда более выгодной оказалась система мониторинга технического состояния и обслуживания полиграфических машин и оборудованияпо потребности. Ключевым элементов такой системы является виброакустическая диагностика, с помощью которой определяется потребность в ремонте.

Разработка методов виброакустической диагностики для отдельных видов полиграфического оборудования проводились в Москве с начала 70-х. Было установлено, что наиболее уязвимыми узлами являются опоры качения и кулачковые механизмы [16]. Исследования виброакустических параметров работы подшипников качения проводились во ВНИИполиграфмаше под руководством О.К. Постникова [15]; была разработана методика диагностики подшипников качения печатной пары малоформатной офсетной печатной

машины «Ромайор-314» [2]. Исследования по разработке методов технической диагностики кулачковых механизмов полиграфических машин проводились как под руководством О.К. Постникова [17;18], так и Г.Б. Куликова на кафедре брошюровочно-переплетных машин Московского государственного университета печати. [11].

Следует отметить, что до последнего времени исследователи, использующие современные информационно-компьютерные технологиями (такие, например, как вейвлетанализ) сосредоточили своё внимание на совершенствовании методов технической диагностики печатных машин [20;21]. До послепечатного полиграфического оборудования у исследователей пока «не доходят руки» - были разработаны только методы компьютерной диагностики для определения технического состояния привода качающегося стола ниткошвейного автомата [1; 12].

Однако номенклатура послепечатного полиграфического оборудования весьма широка. Она включает бумагорезальные машины, вырубные пресса, фальцевальные аппараты, листоподборочные установки, машины клеевого бесшвейного скрепления, скрепко - проволокошвейные машины, комбинированные фальцевально-швейные аппараты (буклетмейкеры), автономные нумераторы, пресса горячего тиснения, ламинаторы и т.п. На крупных полиграфических комбинатах установлены подборочно-брошюровальные линии или вкладочно-швейно-резальные автоматы, которые позволяют выполнять процессы изготовления книжных блоков с большой скоростью в автоматическом режиме. Все послепечатные операции и соответствующее оборудование современные специалисты рассматриваю как часть общего производственного процесса, от набора до переплета, с соответствующими информационными потоками данных.

Вторым, не менее востребованным направлением исследований, мы считаем разработку методов виброакустической диагностики и прогнозирования технического состояния в процессе их эксплуатации послепечатного полиграфического оборудования.

Вывод:

Таким образом, мы определили два наиболее перспективных и востребованных направления исследования шума и вибрации полиграфического оборудования:

-совершенствование методики акустических испытаний при проведении лабораторных исследований в рамках декларирования соответствия;

-совершенствование методов виброакустической диагностики и прогнозирования технического состояния в процессе их эксплуатации.

Оба эти направления объединяет общий объект исследования - послепечатное полиграфическое оборудование, которому, как мы показали, и исследователи методов виброакустической диагностики и прогнозирования технического состояния исследователи, и разработчики стандартов пока не уделяют достаточного внимания.

Список литературы

1. Абрамов В.В., Куликов Г.Б. Использование методов компьютерной диагностики для определения технического состояния привода качающегося стола ниткошвейного автомата БНШ-6 // Проблемы полиграфии и издательского дела. - 2007. № 4. - С. 12-23.

2. Быков А.В. Разработка методики диагностирования подшипников качения печатной пары: Дис. ... канд. техн. наук.- М., 2002. - 190 с.

3. ГОСТ EN 1010-1-2011. Оборудование полиграфическое. Требования безопасности для конструирования и изготовления. Часть 1. Общие требования. URL: http://www.internet-law.ru/gosts/gost/52858/ (дата обращения 15.03.2016)

4. ГОСТ 12.2.231-2012 Система стандартов безопасности труда. Оборудование полиграфическое. Требования безопасности и методы испытаний. // ИС Параграф URL: http://online.zakon.kz (дата обращения 15.03.2016)

5. ГОСТ 27409-97 Шум Нормирование шумовых характеристик стационарного оборудования Основные положения // ИС Параграф URL: http://online.zakon.kz (дата обращения 15.03.2016)

6. ГОСТ EN 1010-3-2011 «Оборудование полиграфическое. Требования безопасности для конструирования и изготовления. Часть 3. Машины резальные» // ИС Параграф URL: http://online.zakon.kz (дата обращения 15.03.2016) 7. ГОСТ EN 1010-1 Проект стандарта «Машины и оборудование полиграфические. Требования безопасности для конструирования и изготовления. Часть 1. Общие требования
(EN 1010-1:2004+A1:2010, IDT)» // Комитет технического регулирования и метрологии Министерства по инвестициям и развитию Республики Казахстан, 2016 URL: https://www.memst.kz/discussion/detail.php?ELEMENT_ID=10549&sphrase_id=19979(дата обращения 15.03.2016)

8. ГОСТ Р 53479-2009. Оборудование полиграфическое. Методы определения шумовых характеристик степени точности 2 и 3 // Открытая база ГОСТов URL: http://standartgost.ru/g/%D0%93%D0%9E%D0%A1%D0%A2_%D0%A0_53479-2009(дата обращения 15.03.2016)

9. Добромыслов Б.В. Определение шумовых характеристик полиграфических машин по измерениям в ближнем звуковом поле : Дисс. …канд. техн. наук Москва 2007. – 172 с.

10. Куликов Г. Б. Диагностика механических систем привода полиграфических машин с использованием искусственных нейронных сетей : дисс. ... д-ра. техн. наук.- Москва, 2008.- 293 с.

ГУМАНИТАРНЫЕ И СОЦИАЛЬНО - ЭКОНОМИЧЕСКИЕ НАУКИ

УДК 808.5:811.512.154-057.875

ОФИЦИАЛЬНО - ДЕЛОВАЯ РЕЧЬ В ПРАКТИЧЕСКОМ КУРСЕ КЫРГЫЗСКОГО ЯЗЫКА ДЛЯ РУССКОЯЗЫЧНЫХ СТУДЕНТОВ

Айтбаева Н.Б., *доцент, канд. пед. наук КГТУ им. И. Раззакова, (+996) 555-503-110, 720044, г. Бишкек, пр. Мира 66, e-mail: aitbaeva20@mail.ru*
Дуйшенкулова Д.Ш., *доцент кафедры кыргызского языка, КГТУ им. И. Раззакова, (+996) 550-07-77-54, 720044, г. Бишкек, пр. Мира 66, e-mail: duishenkulova.sh@mail.ru*

Рассматриваются вопросы обучения русскоязычных студентов составлению официально-деловых бумаг на государственном языке, обосновывается актуальность изучения этого вида письменной речи.

Определены основные коммуникативные навыки и умения, которыми должны овладеть студенты при обучении составлению деловых бумаг, поскольку в сумме они дают те необходимые компетенции, которые делают их востребованными на рынке труда.

Описаны основные функционально-стилистические, структурно-композиционные, морфологические и синтаксические особенности деловых бумаг.

Приведены примеры использования общепринятых речевых штампов, конструкции простых и сложных предложений, присущие данному стилю речи, которые представляют особую сложность для русскоязычных студентов.

Обосновывается использование метода перевода и комментария на этапе презентации грамматического и лексического материала на занятиях кыргызского языка.

Анализ деловых текстов дан на материале разрабатываемого авторами статьи учебного пособия «Расмий иш кагаздары», ориентированного на русскоязычную аудиторию.

Ключевые слова: официально-деловой стиль речи, коммуникация, умение и навыки, государственный язык, общепринятые речевые штампы, терминологическая лексика, стилистические и языковые особенности, родной язык, метод перевода

BUSINESS SPEECH IN A PRACTICAL COURSE OF KYRGYZ LANGUAGE FOR RUSSIAN SPEAKING STUDENTS

Aitbaeva N.B*., Associate Professor, PhD in Pedagogic sciences, KSTU named after I. Razzakov, (+996) 555-503-110, 720044, Bishkek, Mir ave. 66, e-mail: aitbaeva20@mail.ru*
Duishenkulova D. Sh*., Associate Professor at the Department of Kyrgyz language, KSTU named after I. Razzakov, (+996) 550-07-77-54, 720044, Bishkek, Mir ave. 66, e-mail: duyshenkulova.54@mail.ru*

The article considers the problem of practical teaching compilation of official papers in Kyrgyz language to Russian speaking students Information in the article grounds the actuality of learning this type of written speech.

The article identifies the key communicative stages and abilities that a student should master in the process of learning compilation of official papers and help students to become more employable, because together they provide the necessary competence.

The functional-stylistic, structural, morphological and syntactic features of official papers are described. Examples of using common speech patterns, the simple and complex syntax inherent in this style are shown.

Using the methods of translation and comment at the stage of presentation of grammatical and lexical material in Kyrgyz language is explained.

The analysis of business papers in this article is based on the textbook «Расмий иш кагаздары» developed by the authors of this article.

Keywords: official-business style, communication, communicative skills, official language, common speech patterns, terminological lexicon, structural and composite features of official style, native language, methods of translation

Введение. В связи с постепенным внедрением и поэтапным переводом делопроизводства на государственный язык, в рабочую программу практического курса кыргызского языка для русскоязычных студентов был включен модуль «Расмий иш кагаздары. Официально-деловые бумаги», цель которого – знакомство студентов с основными функционально-стилистическими особенностями речи официально-деловых бумаг и формирование навыков и умений составления деловых бумаг на государственном языке. Обучение этому типу коммуникации, имеющему ярко выраженные стилистические особенности, является важной задачей преподавания кыргызского языка как неродного.

Использование метода перевода и комментирование на этапе презентации грамматического и лексического материала методически оправдано тем, что перевод, особенно терминологической лексики, имеет важное значение для лучшей психологической адаптации студентов, позволяющей в определенной степени снять лингвистический барьер и тем самым положительно влиять на результат обучения.

Исследование. Следуя основным принципам Болонского процесса, в модульной рабочей программе по практическому курсу кыргызского языка для русскоязычных студентов был определен уровень формируемых навыков и умений в области письменной речи:

– умение заполнить анкету;

– умение составить заявление, расписку, доверенность, резюме; – умение написать автобиографию; – умение написать характеристику.

При кредитной технологии обучения крайне важно подчеркнуть, что будет знать и уметь студент по окончании изучения курса, поскольку в сумме они дают те необходимые компетенции, которые делают его востребованным на рынке труда. Так, по вышеуказанному модулю основными коммуникативными умениями, которыми должны овладеть студенты при составлении деловых бумаг являются:

– умение правильно располагать реквизиты служебного документа;

– умение использовать адекватный лексический материал, общепринятые речевые штампы;

– умение использовать соответствующие данному стилю речи конструкции простых и сложных предложений;

– умение употреблять по назначению формулы речевого этикета; – умение составить деловую бумагу на государственном языке.

Для автоматизации названных речевых умений студентам необходимо знать следующий лексико-грамматический материал:

– композиционную структуру деловых и официальных бумаг;

– присущие данному стилю речи лексические средства;

– соответствующие формулы речевого этикета;

– конструкции простых и сложных предложений, употребляющиеся при составлении деловых и официальных бумаг.

Приступая к работе над официально-деловым стилем, прежде всего надо иметь ввиду, что он неоднороден и может быть представлен в виде трех подстилей:

1. законодательный, включающий в свой состав нормативные акты высших органов государственной власти – законы, постановления, кодексы, конституции, уставы;

2. дипломатический, представленный такими жанрами, как международное

соглашение, договор, меморандум, конвенция, пакт, нота и другие;

3. административно-канцелярский подстиль – наиболее многожанровая разновидность официально-делового стиля. Сюда входят: приказ, распоряжение, предписание, инструкция, докладная записка, отчет, коммерческая корреспонденция, а также канцелярская документация – справка, заявление, доверенность, расписка, отчет и т.п. В своей практической деятельности студенты чаще всего встречаются с документами этого типа.

Постановление, международное соглашение, инструкция, заявление, доверенность – любой документ имеет юридическое значение, поэтому он составляется согласно правилам делового стиля, что требует особой тщательности и продуманности. Здесь важно не только выразить мысль, но и отобрать те языковые средства, которые необходимы в данной сфере речевого общения. Так, при обучении студентов этому стилю речи, используются такие задания:

Тапшырма.

“Морфологиялык жана синтаксистик өзгөчөлүктөгү расмий иш кагаздарынын тили” деген текстти дептериңерге жазгыла жана аны орус тилине которгула.

Законспектируйте текст “Морфологические и синтаксические особенности официально-деловой речи” и переведите его на русский язык.

Расмий иш кагаздарынын өзүнө мүнөздүү тилдик каражаттары бар.

Лексикалык каражаттар:

Расмий иш кагаздарында эмоцияны билдирүүчү жана экспрессивдик маанидеги сөздөр дээрлик колдонулбайт. Мындай өзгөчөлүк аларды көркөм публицистикалык стилден алыстатып, илимий стилге белгилүү даражада жакындаштырат.

Расмий иш кагаздарында адамдын ысмы, атасынын аты, теги (Асан Баатыр уулу, же Ысак Болот тегин), иштеген жери (Бишкек шаардык Мамкаттоосунун башчысы), дареги (Бишкек шаары “Улан кичирайону 7-үй, 42-батири, телефон № 435434), документе адамдардын иш абалын билдирүү арыз ээси, айыпкер, доокер, күбө, атуул же жаран, жашоочу, мекеменин же башкармалыктын аттары (Кыргыз Республикасынын Мамкаттоосу), башкармалыктын башчылары (Кыргыз Республикасынын Жогорку Кеңешинин спикери) жана башка аталыштар туура жана так жазылууга тийиш.

Морфологиялык каражаттар:

Көптөгөн аталыштар, терминдер сөз жасоочу мүчөлөрдүн жалганышы аркылуу пайда болгон: токтом – токто+ым; келишим – келиш+ым; мүнөздөмө - мүнөздө+мө, -ган атоочтугу+эмес, +ыл+ба+ган формалары алигиче маселе ордунан козголгон эмес. Ар бир башкармалык бул чечимдерди аткарбаган токтомдон: этиштин буйрук ыңгайларынын – дан+дыр, +ыл+сын формалары: Бул токтомдун аткарылышын көзөмөлдөө Бишкек башкы архитектурага (Нарбаев) жана Жерге жайгаштыруу жана кыймылсыз мүлккө укуктарды каттоо боюнча Бишкек шаардык башкармасына (Дюшембиев) жүктөлсүн. (токтомдон).

Синтаксистик каражаттар:

Расмий иш кагаздарында сүйлөмдөрдүн түзүлүшү жагынан өзгөчө болот. Мисалы, сүйлөмдүн баяндоочу ээден мурда келет: Берилди ушул справка Асылбековага ... (справка): Угулду: 1-маселе тууралуу башчынын каттоо боюнча орун басары Болот Берикбаев кыскача баяндайт. (токтомдон).

Сүйлөмдүн айкындооч мүчөлөрү ээ менен баяндоочтон мурун келет: 1993-жылдын 10-ноябрында Улуттук филармониянын чоң залында Кыргыз Республикасынын Эгемендүүлүгүнүн 14 жылдыгына арналган чоң программадагы концерти болот (Кулактандыруу).

Туруктуу эреже катары калыптанып калган тилдик штамптардын колдонулушу: отчеттун мезгил ичинде, каралуучу маселелер, иши жактырылсын, эске алынсын, чаралар көрүлсүн, жарыш сөзгө чыгып сүйлөштү, токтом кабыл алынды, алкыш (сөгүш) жарыялансын, ийгиликтер менен катар төмөндөгү кемчиликтер аныкталды.

Акыркы мезгилдерде канцелярдык иш кагаздарынын негизги белгилери: ой-пикир нечен жылдардан бери калыптанып калган туруктуу бир формада берилет; тилдик штамптар, даяр формулировкалар, иш кагаздарына тиешелүү сөздөр, сөз айкаштары, татаал синтаксисттик конструкциялар активдүү пайдаланылат.

Тапшырма.

Берилген сөздөрдүн кайсыл этиштер менен айкалышканын көрсөткүлө (алардын терминологиялык маанисинде).

Укажите, с какими глаголами обычно сочетаются данные слова (в их терминологическом значении).

Үлгү:

1) *акт – түзүү, берүү... акт– составить, предьявить...*
2) *апелляция – жазуу, берүү...*
   апелляция – написать, подать...
3) *алкыш – алуу, жарыялоо...*
   благодарность – получить, объявить...

Арыз – заявление - .....
Сөгүш – выговор - ....
Дело – дело -....
Келишим – договор - ....
Баяндама – доклад -....
Макала – статья -....
Мыйзам – закон -....
Мамиле – отношение - ....
Отчет – отчет - ....
Чакыруу кагазы – повестка - ....
Күн тартиби – повестка дня - ....
Токтом – постановление - ....
Сунуш – предложение - ....
Эскертүү – предупреждение - ....
Буйрук – приказ - ....
Программа – программа - ....
Протокол – протокол - ....

Тил кат – расписка - ....
Ишеним кат – доверенность - ....
Түшүнүк кат – объяснительная - ....
Мактоо мүнөздөмө-рекомендация - ...
Кабарландыруу – извещение - ....
Чечим – решение - ....
Талап – требование - ....
Справка – справка - ....
Өмүр баян – автобиография - ....
Мүнөздөмө - характеристика - ...
Тастыктама - удостоверение - …
Тескеме, буюрма - распоряжение - ….

Тапшырма.
Сөздөрдү катары менен уланткыла.
Продолжите ряд слов.
1. Официалдуу документтер: келишим, ...
2. Канцелярдык иш кагаздары: тил кат, ...

Сөздүк: буйрук, программа, справка, тил кат, указ, күбөлүк кат, устав, ишеним кат, конвенция, акт, арыз, билдирүү, токтом, меморандум, рапорт, отчет, нота, кулактандыруу, мүнөздөмө, сунуш, талап, эскертүү.

При обучении студентов составлению заявлений используются такие задания:
Тапшырма.

Кийинки зат атоочторду туура жөндөмөдө койгула. Жөндөмө аффикстерин кантип тандаганыңарды далилдегиле.

Следующие имена существительные поставьте в нужной форме. Аргументируйте свой выбор падежных аффиксов

Кимге? - ректор...., декан...., лаборант...., аспирант..., кафедра башчысы..., студенттик
Кому? профсоюздук уюмдун төрагасы..., көркөм ийримдин жетекчиси...., студенттик шаарчанын директору....., № 3 – жатаканын комендanты.

- Исхак Раззкаков атындагы Кыргыз мамлекеттик техникалык университет... ректор.... профессор М.Д.Джаманбаев....(профессор М.Б.Баткибекова....)
- Исхак Раззаков атындагы Кыргыз мамлекеттик техникалык университет... энергетика (технология, информациялык технологиялар) факультет... декан..., доцент (профессор) Т. Ш. Джунушалиева

- Исхак Раззаков атындагы Кыргыз мамлекеттик техникалык университет ....физика (химия, жогорку математика, колдонмо математика, дене тарбия,

	кыбачылык чийүү, механика жана мехатроника, жеңил өнөр жай буюмдарынын технологиясы, консервалоо жана тамак-аш азыктарынын технологиясы, кыргыз тили) кафедра ... башчы.... доцент Н.Б.Айтбаева....
Кимден? От кого?	- технологиялык (энергетикалык, информациялык технологиялар) факультет... студент....Иванов Сергей....Прыткова Света....

При анализе деловых текстов надо учитывать те или иные языковые особенности, которые могут варьироваться в большей или меньшей степени. Одной из важных особенностей стиля являются составляющие его лексические элементы. Это наглядно можно продемонстрировать на следующих заданиях.

Тапшырма.

Арыз жазууда колдонулуучу тилдик штамптарды көчүрүп жазгыла. Запишите речевые штампы, которые используются при составлении заявления.

1. Прошу предоставить – алып берүүңүздү суранам (өтүнөм)

Прошу зачислить – киргизүүңүздү суранам (өтүнөм)

Прошу перевести – которууңузду суранам (өтүнөм)

Прошу разрешить – улуксат берүүңүздү суранам (өтүнөм)

Прошу отозвать –мөөнөтүнөн мурда чакыртууңузду суранам

Прошу дать возможность – мүмкүнчүлүк берүүңүздү суранам (өтүнөм)

Прошу освободить от занятий – сабактан бошотууңузду суранам

Прошу оказать содействие – жардам көрсөтүшүңүздү өтүнөм

Прошу командировать – эмгектик өргүүгө жөнөтүшүңүздү суранам

Прошу выделить необходимые материальные средства, (путевку) – зарыл материалдык каражаттарды (жолдомо) бөлүштүрүп берүүңүз суранам.

Прошу оказать материльную помощь – материалдык жардам берүүңүздү суранам.

Прошу дать указание на... – көрсөтмө берүүңүздү суранам (өтүнөм)

Прошу дать разрешение на... – улуксат берүүңүздү суранам (өтүнөм)

Прошу допустить к участию в конкурсе на замещение вакантной должности... – бош орунду толуктоо сынагына катышууга уруксат берүүңүздү суранам (өтүнөм)

Прошу освободить от занимаемой должности – ээлеген орунумдан бошотууну суранам.

2. по семейным обстоятельствам – үй-бүлөлүк шартка байланыштуу // үй-бүлөлүк шартка жараша;
- в связи с болезнью – ден соолугума байланыштуу;
- в связи с непредвиденными обстоятельствами – күтүлбөгөн шартка байланыштуу;
- в связи с несчастным случаем – кырсыкка байланыштуу;
- ввиду актульности и неразработанности темы – теманын актуалдуулугуна жана иштетилбегендигине байланыштуу;
- по собственному желанию - өзүмдүн каалом менен.

Тапшырма.

Төмөндөгү көрсөтүлгөн конструкцияларды пайдаланып университеттин ректоруна, студенттик шаарчанын директоруна, факультеттин деканына, кафедранын башчысына арыз жазгыла.

Используя данные ниже конструкции, составьте заявления на имя ректора университета, директору студенческого городка, декану факультета, заведующему кафедрой:

- ооруп калгандыгыма байланыштуу, экинчи семестрден баштап, бир жылдык мөөнөткө академиялык эс алууга чыгууга уруксат берүүңүздү өтүнөм;
- спорттук мелдештерге катышып жаткандыгыма байланыштуу мени 2015-жылдын 20ноябрынан 30-ноябрына чейин сабактардан бошотуп коюуңузду өтүнөм;
- үй-бүлөлүк шартыма байланыштуу мени сырттан окуу бөлүмүнө которуп коюңузду суранам.

Стандартизация, трафаретность выражений уместна и оправдана практикой деловой коммуникации, так как она здесь облегчает процесс составления и восприятия делового документа, способствует унификации деловых бумаг.

При обучении русскоязычных студентов составлению деловых бумаг нельзя пренебрегать возможностью использовать опору учащегося на родной язык. Эта мысль не противоречит мнению А.А.Леонтьева о том, что “даже если при организации обучения мы не будем учитывать возможность (и необходимость) опоры на родной язык, или точнее на навыки речи на родном языке, учащийся все равно будет опираться на эти навыки без нашей помощи и участия”.

Выводы: Многочисленные исследования убедительно показали зависимость результатов обучения от степени использования в процессе обучения родного языка учащихся. Сошлемся хотя бы на мнение Л.В.Щербы, что правильное изучение иностранного языка “заставляет вдумываться в самое существо человеской мысли; при этом происходит “преодоление родного языка”, выход из его магического круга. Вот почему знакомство русскоязычных студентов с основными функционально-стилистическими особенностями речи официально-деловых бумаг и формирование навыков и умений составление деловых бумаг на государственном языке идет в тесном контакте с языком родного языка (русского) учащихся.

Составление деловых бумаг требует большого умения и языковой культуры не меньше, чем другие виды письменной речи. Поэтому развитие навыков стилистически дифференцированной речи и деловой речи в частности, представляет для студентов несомненный практический интерес.

Список литературы

1. Жапарова Б.Б., Жусупова Н.Л. Делопроизводство и деловая переписка на государственном и официальном языках. Бишкек 2009
2. Кыргыз Республикасынын Мамлекеттик тили боюнча расмий документтердин жыйнагы. -Б, 2014
3. Костомаров В.Т. Еще раз о понятии “родной язык” // Русский язык в СССР 1991 № 1.
4. Сироткина З.И. Учет родного языка на продвинутом этапе // Русский язык за рубежом 1987, № 5.
5. Шпанова Э.М. Лингвистические особенности официально-деловой речи // Русский язык за рубежом 1983

References

1. Japarova B.B., Jusupova N.L. Document management and business correspondence on official language. Bishkek, 2009

2. The collection of official documents on the state language of Kyrgyz Republic. Bishkek, 2014
3. Kostomarov V.G. Once again about the concept of "native language" // Russian in USSR 1991, № 1.
4. Sirotkina Z.I. Accounting the native language in the advanced stages // Russian language abroad 1987, № 5.
5. Shpanova E.M. Linguistic features officially-business speech // Russian language abroad 1983

УДК 371.334:811.512.154

ДИСТАНЦИОННЫЕ ТЕХНОЛОГИИ ОБУЧЕНИЯ В ПРАКТИЧЕСКОМ КУРСЕ КЫРГЫЗСКОГО ЯЗЫКА КАК НЕРОДНОГО

Айтбаева Н.Б., *доцент, канд. пед. Наук КГТУ им. И. Раззакова, (+996) 555-503-110, 720044, г. Бишкек, пр. Мира 66, e-mail: aitbaeva20@mail.ru*
Дуйшенкулова Д.Ш., *доцент кафедры кыргызского языка, КГТУ им. И. Раззакова, (+996) 550-07-77-54, 720044, г. Бишкек, пр. Мира 66, e-mail: duyshenkulova.54@mail.ru*

В статье рассматриваются вопросы использования дистанционных технологий обучения в практическом курсе кыргызского языка как неродного, виды информационнокоммуникационных технологий, используемые в сфере обучения языкам, лингводидактические возможности современной электронной среды.

Обосновывается актуальность использования дистанционных технологий обучения языкам, поскольку на их основе можно строить принципиально новые формы обучения, позволяющие сделать процесс овладения кыргызским языком более естественным.

Представлена презентация электронного учебного пособия «Стандартный лексический минимум в электронном практикуме по лексике кыргызского языка», разрабатываемый преподавателями кафедры кыргызского языка КГТУ им. И.Раззакова на примере тематического блока «Быт. Турмуш-тиричилик».

Описана методика работы с лексическим материалом от слова вне какого-либо контекста (значение, графический облик) к освоению его коммуникативных возможностей, его способностей участвовать в достижении реальных целей речевого общения.

Ключевые слова: дистанционное обучение, инновационные технологии, лексические единицы, электронные коммуникации, стандартный лексический минимум, тематические блоки, интернет-технологии, мотивационная основа

FORMING PROFESSIONAL COMMUNICATIVE COMPETENCE IN KYRGYZ LANGUAGE TEACHING FOR RUSSIAN SPEAKING STUDENTS

Aitbaeva N.B., *Associate Professor, PhD in Pedagogic sciences, KSTU named after I. Razzakov, (+996) 555-503-110, 720044, Bishkek, Mir ave. 66 e-mail: aitbaeva20@mail.ru*
Duishenkulova D. Sh*., Associate Professor at the Department of Kyrgyz language, KSTU named after I. Razzakov, (+996) 550-07-77-54, 720044, Bishkek, Mir ave. 66, e-mail: duyshenkulova.54@mail.ru*

The article considers the problem of using distance learning technology in practical course of Kyrgyz language as native, types of the information and communication technologies which can

be used in practice of language teaching; linguo-didactic peculiarities of the modern electronic environment.

The actuality of using the distance learning technology is explained, since it is the basis on which it will be able to build fundamentally new methods and forms of education for making the processes of learning Kyrgyz language more natural.

There is a presentation of the electronical textbook «Standard lexical minimum in the electronic practical work on the Kyrgyz language vocabulary» which is based on the thematic cluster «Быт. Турмуш-тиричилик» developed by the Teachers in the Department of Kyrgyz language, KSTU named after I. Razzakov. The method of working with lexical material from the word beyond any context (meaning, graphical form of word) to the harnessing its communicative potential, its ability to participate in the reaching the real goals of language communication.

Keywords: distance learning, innovative technologies, lexical units, electronic communication, basic word stock, thematic block, internet technologies, motivational basis

Введение. Переход на кредитную систему обучения потребовал кардинального пересмотра содержания и методов обучения языку и необходимости создания новых учебно-методических разработок, основной целью которых является формирование навыков практического владения языком в различных ситуациях повседневного речевого общения, введение использования компьютера в практике преподавания кыргызского языка как неродного.

Проблемы обучения лексике в рамках практического курса кыргызского языка как неродного до настоящего времени не получили всестороннего освещения в научных трудах, и лексический аспект по сравнению с другими в значительно меньшей степени обеспечен учебной литературой.

Практика показывает, что ставшее уже традиционным слияние аспектов «Лексика» и «Грамматика» в единый лексико-грамматический аспект приводит, как правило, к доминированию грамматики в процессе обучения, когда лексика отступает на второй план, начинает играть вспомогательную роль, становясь материалом для наполнения грамматических конструкций. В этом случае некоторым обязательным для изучения лексическим единицам не уделяется должного внимания или когда в учебных текстах появляется лексика, выходящая за пределы лексического минимума того или иного уровня и в последствии не находящая своего употребления в различных сферах и ситуациях общения.

В связи с этим возникает необходимость в разработке для русскоязычных учащихся таких средств обучения лексике, в которых был бы реализован принцип комплексного, последовательного и поэтапного освоения соответствующих определенному уровню владения кыргызским языком лексическим единиц.

Однако, представление лексики во всем объеме ее коммуникативных возможностей трудно реализовать только с помощью традиционных средств обучения. Полноценное использование в учебном процессе лексического минимума возможно только в электронной среде.

Исследование. Инновационные технологии в обучении языку связаны с использованием современных информационных технологий. Современные компьютерные средства дают возможность использовать учебные материалы нового поколения, интерактивные учебно-методические комплексы, включающие в себя всё необходимое для обучения, в частности кыргызскому языку, как в группах, так индивидуально и одновременно создающие благодаря Интернету интерактивные площадки дистанционной учебно-методической поддержки.

В сфере обучения языкам применяются следующие виды информационнокоммуникационных технологий (ИКТ): электронные учебники,

интерактивные обучающие пособия, справочно-информационные источники (онлайн-переводчики, словари), электронные библиотеки, электронные периодические издания и.т.д.

Использование дистанционных технологий обучения в учебном процессе оправдано рядом причин:

Во-первых, обеспечивается доступность обучения. Независимо от места нахождения любой желающий может получить те или иные образовательные услуги в индивидуальном режиме;

Во-вторых, применяются новые формы организации и представления информации: текст, графика, видео, анимация, огромный объем справочной, основной и сопроводительной информации;

В-третьих, вводятся новые формы сертификации знаний и умений путем использования тестов, рефератов, проектов и др.

Основные лингводидактические возможности современной электронной среды широко известны:

- мультимедийность, позволяющая использовать все форматы представления языковой, речевой и экстралингвистической информации, воздействуя на все каналы восприятия;
- гипертекстовая организация учебного материала, дающая возможность наглядно структурировать и дозировать больше информационные массивы, эффективно направляя самостоятельную деятельность субъекта;
- интерактивность, обеспечивающая оперативную и разнообразную реакцию дидактической интегрированной среды на действия пользователя.

Исследователи считают, что использование электронной коммуникации в качестве средства обучения помогает частично решить одну из основных задач обучения - создание естественной языковой среды, поскольку даёт дополнительные возможности общения на изучаемом языке.

Студенты могут пользоваться учебным сайтом, знакомиться с новым лексическим материалом, читать тексты, выполнять упражнения и отправлять на электронный адрес преподавателя.

Преподаватели кафедры кыргызского языка в настоящее время работают над созданием электронного учебника «Стандартный лексический минимум в электронном практикуме по лексике кыргызского языка», первая часть которого уже размещена в образовательном портале ИДО и ПК КГТУ им. И.Раззакова в разделе «Инновационные технологии обучения, применяемые в учебном процессе».

Выбор лексического минимума в качестве основы ресурса по лексике не случаен. Данный методический материал не только конкретизирует список лексических единиц, подлежащих обязательному усвоению на определенном уровне владения кыргызским языком, но и служит ориентиром для авторов учебников, учебных пособий и разработчиков текстовых материалов. Именно поэтому в последнее время как преподаватели, так и учащиеся все чаще стали использовать данный компонент тестирования в функции своеобразного учебного пособия, и с этой своей новой ролью (ролью обучающего) лексический минимум в бумажном варианте справляется с трудом. Для его полноценного использования в учебном процессе необходимо оснастить словник расширенным методическим аппаратом, что возможно только в электронной среде.

Цель создания электронного учебника - поддержка лексического аспекта практического курса кыргызского языка как неродного. Задачи обучения - помочь организовывать и направлять самостоятельную учебную речевую деятельность студентов,

способствовать индивидуализации процесса обучения, формировать устойчивую мотивационную основу для обращения к кыргызскому языку во внеаудиторное время, расширяя зону их контакта с изучаемым языком посредством расширения лексического запаса слов.

В структуре электронного учебника несколько тематических блоков:

- быт (квартира+предметы быта+мебель);
- питание (продукты + магазины +интернациональная кухня);
- время (часовое время +времена года +погода);
- здоровье (основные части тела +болезни + лечебные учреждения)

Каждый тематический блок открывается формулировкой коммуникативной задачи, которая направляет речевую деятельность студентов, а также определяет отбор конкретного лексического материала. Например, по тематическому блоку «Быт. Турмуш-тиричилик» студенты должны сформировать следующие речевые умения:

- давать описание мебели, обстановки жилища;
- давать краткое описание квартиры, задавать вопросы, связанные с жилищем;
- высказывать мнения по поводу жилья;
- научиться пользоваться лексикой, которая связана с вопросами жилья.

Четкая ориентация учащихся на решение конкретной коммуникативной задачи придает всей учебной работы в рамках тематического блока осмысленность, целенаправленность и системность.

В начале каждого тематического блока предлагаются тематические списки слов для первичного ознакомления, выбранные на основе их общеупотребительности в речи.

Знакомство со словом в электронном ресурсе происходит через совмещение его графического и визуального образов: при подведении курсора к изображению предмета на мониторе компьютера появляется изображение предмета и его написание на русском и кыргызском языках. Так, по тематическому блоку «Быт. Турмуш-тиричилик» предлагается следующий стандартный лексический минимум.

Дом - үй

1. план дома - үйдүн планы;
2. забор - кашаа, короо, дубал; сад огорожен забором - дубал менен тосулган бакча (короо)
3. почтовый ящик - почта ящиги;
4. проезд - баруу, жүрүү;
5. дверной звонок - эшиктин коңгуроосу;
6. фонарь - фонарь, чырак;
7. входная дверь - кире бирештеги эщик;
8. двор - короо;
9. веранда - веранда (үйгө кошулуп саланган, үстү жабылган же ачык, же терезеленген жай);
10. окно - терезе;
11. водосточная труба - суу аккан түтүк;
12. дымовая труба - мор;
13. крыша - крыша үйдүн төбөсү, үйдүн үстү, чатыр; жить под одной крышей с кем-либо - бирөө менен бир үйдө туруу

Кухня - ашкана

1. ручное полотенце - кол ааарчы (суулук, майлык);
2. сушилка - кургаткыч;

3. мусорный ящик - таштанды (шыпырынды) ящиги;
4. полка - текче;
5. холодильник - муздаткыч;
6. морозильник - тондургуч;
7. кофейница - кофейница;
8. заварной чайник - чай демдей турган кичинекей чайнек;
9. большой чайник - чоң чайнек;
10. духовка - духовка;
11. обеданный стол - тамак ичүүчү үстөл;
12. сковорода - көмөч казан (казан көмөч);
13. миксер - $_1$.аралаштыргыч $_2$.чалгыч;
14. доска - тактай;
15. стул - отургуч (орундук).

Методика работы с лексическим материалом предусматривает цепочку технологических шагов от внимания к слову вне какого-либо контекста (значение, графический облик) к освоению его коммуникативных возможностей, его способностей участвовать в достижении реальных целей речевого общения.

Все упражнения (и лексико-грамматические и ситуативные), построенные на основе тематической лексики, имеют одну практическую цель и направлены на расширение лексического запаса студентов.

Рассмотрим следующие типы упражнений:

Тапшырма.
Сүрөткө сөздөрдү тууралап койгула.
Соотнесите слова с рисунками.

1. Муздактыч
2. Сыналгы
3. Кир жуугуч машине
4. Музыкалык центромпьютер
5. Чаң соргуч
6. Үтүк
7. Лампа - ылампа, асмачырак - висячая лампа
8. Калькулятор
9. Батарейка
10. Фонарик, колчырак
11. Тостер
12. Газ плитасы
13. Өчургуч (свет)
14. Кофе кайнаткыч


TechHome.ru


CITIZEN
1234567890 12.

Тапшырма
Бөлмөдө турган эмеректерди атагыла.
Назовите мебель, стоящую в комнатах.

Образец/Үлгү.

Балдар бөлмөсү

Балдар бөлмөсү кичине, бирок жарык. Ал бөлмөдө бир керебет, китеп жана кийим салган шкаф турат. Бул бөлмөдө сабак окуган үстөл жана көп оюнчуктар бар.

Балдар бөлмөсү

Лексико-грамматические и ситуативные задания, построенные на сонове тематической лексики, имеют одну практическую цель и направлены на расширение лексического запаса учащихся.

Выводы: Таким образом, интернет-технологии обладают значительными образовательными возможностями, которые могут найти применение в преподавании кыргызского языка русскоязычной аудитории. Процесс в этой области деятельности не вызывает никакого сомнения. Наиболее перспективными, на наш взгляд, являются дистанционные технологии, поскольку на их основе можно строить принципиально новые формы обучения, а также программные оболочки, которые позволяют самим преподавателям создавать и размещать в интернете тренировочные и тестовые задания, и упражнения. Следует, конечно, понимать, что организация электронной (виртуальной) образовательной среды - сложный процесс, он требует совместных усилий коллективов методистов, программистов.

Отметим также, что материалы электронного ресурса востребованы и в связи с недостаточным количеством учебных часов по кыргызскому языку, большой наполняемостью учебных групп, сложностью демонстрации функционирования изучаемой лексики в аудитории. С их помощью можно прежде всего «нарастить» реальное количество учебных часов, активизируя и обеспечивая учебными материалами самостоятельную внеаудиторную деятельность студентов, индивидуализировать её, что весьма сложно сделать в аудитории в условиях занятия с большой группой.

Список литературы

1. Азимов Э.Г. Методика организации дистанционного обучения русскому языку как иностранному. М., 2006
2. Азимов Э.Г. Информационно-коммуникационные технологии в обучении русского языка как иностранного: состояние и перспективы //Русский язык за рубежом. 2011. № 6.
3. Бовтенко М.А. Компьютерная лингводидактика: Учебное пособие. М., 2005
4. Богомолов А.Н. Интернет-технологии в обучении русскому языку как иностранному // Вестник ЦМО МГУ. 2009. № 1.
5. Гарцев А.Д. Электронная лингводидактика в системе инновационного языкового образования. Автореф.дис..... д-ра пед.наук. М., 2009.

УДК 332.012.32:330.131.5

РАЗВИТИЕ МАЛОГО И СРЕДНЕГО БИЗНЕСА В РЕГИОНЕ И ЕГО ВЛИЯНИЕ НА СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ

Алибаева Дамира Какеновна, *соискатель, КГТУ им. И.Раззакова, Кыргызстан, 720044, г. Бишкек, пр. Мира 66,* e-mail: dk_alibaeva@mail.ru

Цель статьи - посвящена исследованию тенденций развития малого и среднего бизнеса Чуйской области и вклада сектора в результаты социально-экономических преобразований; приводятся результаты анализа удельных весов предприятий МСП области в основных экономических показателях ее развития.

Ключевые слова: малый и средний бизнес, развитие частного сектора, валовой региональный продукт, внешнеэкономическая деятельность, социально-экономические преобразования.

DEVELOPMENT OF SMALL AND MEDIUM-SIZED BUSINESSES IN THE REGION AND ITS IMPACT ON THE SOCIO-ECONOMIC

TRANSFORMATION

***Alibaeva Damira**, competitor , KSTU named after I.Razzakov , Kyrgyzstan , 720044, Bishkek c., prospect Mira 66,* e-mail:. dk_alibaeva@mail.ru

The purpose of the article is devoted to the study of development trends of small and medium-sized business sector and the Chui region 's contribution to the results of socio - - economic changes ; the results given to the analysis - weights of SME enterprises in the field of basic economic exponents - Lyakh its development.

Keywords: small and medium enterprises, private sector development, the gross regional product, foreign trade, social and economic transformation.

В большинстве стран мира сектор малого и среднего бизнеса (МСБ) очень важен, так как его развтие может гарантировать высокое экономическое развитие страны и международную конкурентоспособность и может содействовать правительству в решений важных экономических и социальных задач. Практика европейских стран с переходной экономикой показала, что динамичное развитие МСБ сделало возможным относительно быстрый переход от планово-директивной экономики к рыночной.

В Кыргызской Республике частный сектор составляет основу экономики, и важная роль при этом принадлежит малому и среднему бизнесу. Развитие частного сектора Кыргызской Республики невозможно рассматривать без учета ее общего экономического состояния. Удельный вес МСП в экономике республике велик в достаточной степени, чтобы влиять на тенденции развития экономики в целом [1]. С другой стороны, состояние всей экономики создает условия для развития частного сектора, и в частности сектора МСП.

Согласно статистическим данным в 2014г. на территории Кыргызской Республики действовало 13,5 тыс. предприятий МСП, из них 12,7 тыс. – малые и 0,8 тыс. – средние [2, стр. 38].

Количество занятых в МСП в общей численности занятых в экономике Кыргызстана в 2014 году составило 439 тыс. человек, или 19 %. Количество занятых на предприятиях сектора за пять лет возросло с 50 тыс. 200 человек до 52 тыс. Количественный рост субъектов этого сектора сопровождался ростом его доли в объеме валовой добавленной стоимости (рис. 1).

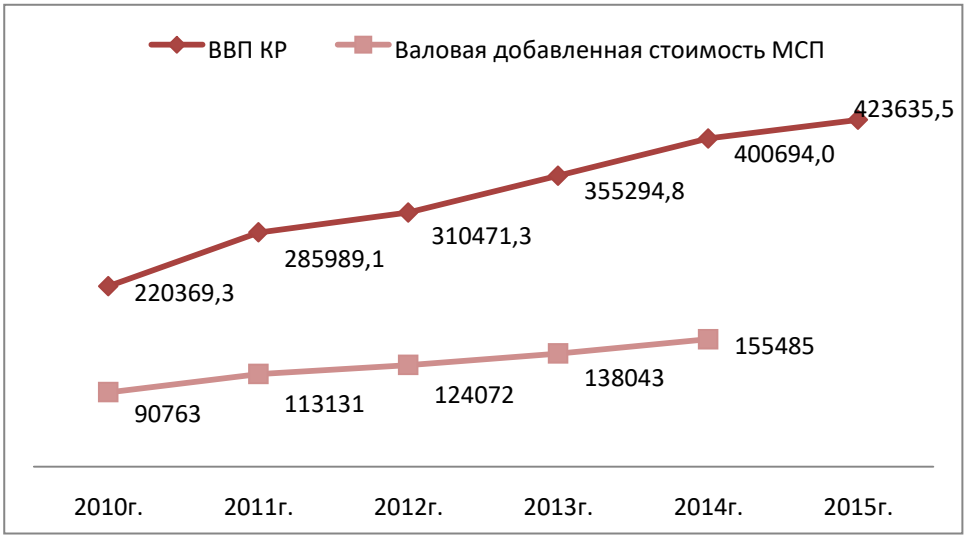


Рис. 1 – Рост объемов ВВП Кыргызской Республики и валовой добавленной стоимости МСП, млн. сом

Объем валовой добавленной стоимости, произведенной предприятиями МСП приведен в таблице 1 и отражает ее стабильный рост по всем категориям субъектов.

В результате сравнения темпов роста ВВП Кыргызской Республики, составивших 92,2% за период с 2010г., и объемов валовой добавленной стоимости (рост 71,3% за тот же период) выявлено, что темпы роста показателя МСП отстают от роста ВВП на 20,9 процентных пункта.

Это отражает негативный фактор в развитии сектора МСП, свидетельствующий о недостаточном использовании его потенциала.

Таблица 1 – Объем валовой добавленной стоимости, млн. сомов

Субъекты МСП	**2010г.**	**2011г.**	**2012г.**	**2013г.**	**2014г.**	**Январьмарт 2015г.**
Малые предприятия	16325,0	19580,6	23724,7	26519,8	30779,9	25909,0
Средние предприятия	11538,6	13860,6	13561,5	16074,6	17606,0	5171,9
Индивидуальные предприниматели	37420,9	48062,7	55367,3	62540,4	70515,9	3337,1
Крестьянские (фермерские) хозяйства	25478,7	31627,4	31418,6	32908,1	36583,6	3584,9

В среднем за 2010-2014гг. доля валовой добавленной стоимости, произведенной субъектами малого и среднего предпринимательства, составила около 40 процентов к ВВП (табл. 2). Однако выявлено, что по итогам 2014г. ее объем сложился в размере 155485,4 млн. сомов, или 39 процентов к ВВП.

Таблица 2 – Удельный вес объема валовой добавленной стоимости в ВВП, % [2, стр. 24]

Субъекты МСП	**2010г.**	**2011г.**	**2012г.**	**2013г.**	**2014г.**
Малые предприятия	7,4	6,8	7,6	7,5	7,7
Средние предприятия	5,2	4,8	4,4	4,5	4,4
Индивидуальные предприниматели	17,0	16,8	17,8	17,6	17,7
Крестьянские (фермерские) хозяйства	11,6	11,1	10,1	9,3	9,2

Данные, приведенные в таблице 2, отражают практическое отсутствие положительных изменений в размерах вклада МСП в результаты деятельности республики с 2010 года. Так, удельный вес объема валовой добавленной стоимости малых предприятий повысился лишь на 0,3 процентных пункта, индивидуальных предпринимателей – на 0,7 процентных пункта. При этом произошло снижение показателя средних предприятий на 0,6 процентных пункта, крестьянских (фермерских) хозяйств – на 1,4.

В Кыргызской Республике сектор МСП представлен неравномерно по регионам: в 2015г. 70% предприятий данного сектора зарегистрировано в г. Бишкек, в Чуйской области – 9%, в Ошской – 3%, в Джалал-Абадской – 4%, Иссык-Кульской – 3%, Нарынской – 2%, Таласской – 2% и в Баткенской области – 1%.

В общем количестве предприятий сектора МСП Кыргызской Республики предприятия Чуйской области имеют долю около 10% в разные годы исследуемого периода (табл. 3). Основным представителем МСП области является малый бизнес: его доля в общем показателе составляет 9,6% в 2010г. и 8,8% в 2014г. Средние предприятия имеют

значительно меньший вес: их количество сложилось на уровне 1% от общего количества предприятий МСП КР.

Таблица 3 – Доля предприятий МСП Чуйской области в общем количестве предприятий МСП КР, %

	2010г.	**2011г.**	**2012г.**	**2013г.**	**2014г.**
Всего	10,7	10,8	11	10,3	9,7
малые предприятия	9,6	9,8	9,9	9,4	8,8
средние предприятия	1,1	1	1	0,9	0,9

На 01.01.2016г. в Чуйской области зарегистрировано 11165 предприятий МСП, отсюда можно сделать вывод о наибольшей степени развитости области в сфере МСП после г.Бишкек.

Предприятия МСП Чуйской области функционируют в различных сферах деятельности. Так, 2304 предприятия из общего количества зарегистрированных (9808 субъектов) занимаются обслуживающей деятельностью, 2123 – оптовой и розничной торговлей. В обрабатывающих производствах заняты 1167 предприятий МСП, в сельском хозяйстве – 1138 предприятий [3, стр. 46]. Строительных частных предприятий МСП зарегистрировано 789 единиц, финансовым посредничеством занимается 504 предприятия.

Доля промышленных предприятий МСП в общем количестве предприятий области составляет почти 33% (рис. 2). Остальные предприятия МСП заняты в таких сферах, как оптовая и розничная торговля, строительство, транспорт, информация и связь, а также гостиницы и рестораны.

Рис. 2 – Доля предприятий МСП Чуйской области в общем числе предприятий области по видам экономической деятельности, %

Вклад МСП Чуйской области в результаты ее развития изучен через оценку доли сектора МСП в основных показателях деятельности Чуйской области.

Валовой региональный продукт Чуйской области [4] имеет выраженную положительную динамику: рост показателя с 2008 года составил 155%. По Нацстаткома, валовой региональный продукт Чуйской области в 2014 году составил 45 млн. 164 тыс. сомов; с 2013 года показатель снизился на 9,1%.

Объемы промышленной продукции МСП Чуйской области в период с 2010 года стабильно увеличивались (табл. 4). Наибольший вклад в данный показатель внесли

индивидуальные предприниматели; активно увеличивали объемы производства и средние предприятия. Малые предприятия в наименьшей степени заняты в промышленности.

Таблица 4 – Объем промышленной продукции, работ, услуг МСП Чуйской области, млн. сомов [2, стр. 134]

	2010г.	**2011г.**	**2012г.**	**2013г.**	**2014г.**
Сектор МСП	8898,5	9413,2	9826,8	11389,8	12112,6
малые предприятия	1430,2	1341,4	1685,8	2348,5	2 544,7
средние предприятия	2625,5	3203	3043,5	3468,8	4 424,8
индивидуальные предприниматели	4842,8	4868,8	5097,5	5572,5	5 143,1

Можно говорить об относительно низком уровне активности МСП в сфере промышленного производства Чуйской области.

Однако надо отметить, что к 2014 году именно малые предприятия наиболее активно увеличили показатель: его рост с 2010 года составил чуть менее 80%. Рост объемов производства промышленной продукции средними предприятиями с 2010 по 2014 г.г. составил 68,5%, индивидуальными предпринимателями – 6,2%.

Рост активности МСП присутствует и в сфере сельскохозяйственного производства: с 2010 года стабильно увеличиваются объемы производства сельхозпродукции (табл. 5). Субъектами малого и среднего предпринимательства, осуществляющими деятельность в сельском хозяйстве (включая крестьянские (фермерские) хозяйства), произведено продукции на 116360 млн. сомов. Традиционно, крестьянские хозяйства обеспечивают более половины всех объемов; их удельный вес колеблется от 54,2% в 2010г. до 57,3% в 2014г. Кроме того, они наиболее активно увеличивали данный показатель за период (рост на 69,8%). Рост объемов сельхозпродукции малыми предприятиями – 60,2%, средними – 16,8% в 2014г.

Динамика объемов оптовой и розничной торговли сектора МСП увеличивается в период с 2010 по 2014 г.г.; общий рост составил 94,5%. Увеличение объемов торговли малых предприятий – 32,2%, средних предприятий – 131%. И наиболее активно выросли объемы торговли индивидуальных предпринимателей: показатель увеличился на 163,5% (табл. 6):

Таблица 6 – Объем оптовой и розничной торговли МСП Чуйской области, млн. сомов

	2010г.	**2011г.**	**2012г.**	**2013г.**	**2014г.**
Сектор МСП	26 429,1	32 538,8	37 618,6	42 907,8	51 402,5
малые предприятия	13635,5	16734,8	15749	16782,9	18 042,7
средние предприятия	1077,5	1297	3690,4	4736,9	2 490,4
индивидуальные предприниматели	11716,1	14507	18179,2	21388	30 869,4

Анализ структуры объемов оптовой и розничной торговли Чуйской области показывает, что вся торговля сосредоточена в руках субъектов МСП. Наименьший удельный вес в показателе имеют средние предприятия, доля которых за период колеблется в диапазоне от 4% до 11%: максимальный вес сектора отмечен в 2012г. (9,8%) и в 2013г. (11%). Доля малых предприятий стабильно снижалась с 51,6% в 2010г. до 35,1% в 2014г., при этом вклад индивидуальных предпринимателей возрастал с 44,3% до 60,1% соответственно в 2010 и 2014г.г. Таким образом, отмечено изменение структуры оптовой и розничной торговли Чуйской области, которая постепенно все больше переходит в руки индивидуальных предпринимателей. [2, стр. 140]

В 2014г. внешнеторговый оборот субъектов малого и среднего предпринимательства Кыргызской Республики составил 3519,2 млн. долларов США (в текущих ценах) и уменьшился по сравнению с 2013г. на 4,8 процента, а по сравнению с 2010г. увеличился на 40,7 процента.

Экспортные поставки МСП Чуйской области за период с 2010 года снижаются: в 2014г. они сложились в размере 44066,2 тыс. долларов США, что на 4,3% меньше, чем в 2010г. и на 16 % больше, чем в 2013г. (табл. 7).

Тенденция снижения экспорта отмечена у малых предприятий на 34%, что ввиду определяющей их роли в общих объемах экспорта МСП (их доля составляет 60,1% МСП) отразилось и на общем снижении экспортных поставок МСП области. Средние предприятия повысили экспортную активность с 2010 года на 291%, индивидуальные предприниматели – на 18%, а крестьянские (фермерские) хозяйства – на 188%.

Таблица – Объем экспортных операций МСП Чуйской области, тыс. долларов США [2, стр. 158]

	2010г.	**2011г.**	**2012г.**	**2013г.**	**2014г.**
Сектор МСП	46263,4	38069,8	37660,9	37811,3	44066,2
малые предприятия	40391,3	22567,6	25905,6	22015,3	26 469,7
средние предприятия	3827,6	13278,1	9586	10491,7	14 967,7
индивидуальные предприниматели	1919,6	2036,5	1891	4969	2 268,9
крестьянские (фермерские) хозяйства	124,9	187,6	278,3	335,3	359,9

Доля МСП в общих объемах экспортных операций Чуйской области в 2010 году составляла 20%, увеличившись до 27,6% в 2014 году.

Ведущая роль в экспортных операциях МСП принадлежит малым предприятиям: их доля в общих объемах экспорта Чуйской области в 2010г. составляла 17,5%, увеличилась в 2011г. до 22,4% и уменьшилась в 2014г. до 16,6% (рис. 2.26). Средние предприятия активизировались в 2011 году, увеличив долю в экспорте области с 1,7% в 2010г. до 13,2%. На 2014 год их вклад составил 9,4%.

Импортные поступления от субъектов МСП Кыргызской Республики в 2014г. составили 3053,8 млн. долларов США и по сравнению с предыдущим годом уменьшились на 2,1%, а по сравнению с 2010г. увеличились на 49,1%.

Сектор МСП Чуйской области в 2014г. увеличил объемы импорта с 2010 года на 82,6%, а в сравнении с 2013 годом уменьшил показатель на 28,1%. Рост показателя с 2010 года обеспечили малые предприятия – на 32,8%, средние предприятия – на 253%, индивидуальные предприниматели – на 133,4%. И наибольший вклад в рост объемов импорта за пятилетний период внесли крестьянские хозяйства, рост объемов которых составил 618,8%. [2, стр. 162]

Средняя доля МСП в общих объемах импорта Чуйской области за период 20102014г.г. составила 56,2%.

Обобщив все данные приведенного выше анализа вклада сектора МСП Чуйской области в ее результаты, можно сказать, что этот вклад очень значителен. Такие сферы, как оптовая и розничная торговля, услуги гостиниц и ресторанов, а также подрядные работы, полностью обеспечиваются предприятиями МСП. Вклад сельскохозяйственных предприятий МСП находится на уровне порядка 60-ти процентов (включая крестьянские (фермерские) хозяйства и индивидуальных предпринимателей). Удельный вес внешнеторговых операций сектора колеблется по годам периода исследования: доля экспортных операций МСП не

достигает 40-ка процентов в общих объемах области. Импортные операции со стороны предприятий МСП вносят более весомый вклад и достигают почти 70-ти процентов в 2012 году, снижаясь до 44,5% в 2014 году.

Наименьший вклад в деятельность Чуйской области вносят предприятия сектора МСП, функционирующие в сфере производства промышленной продукции: его доля составляет от 15-ти до 20-ти процентов, различаясь по исследуемому периоду (рис. 3):

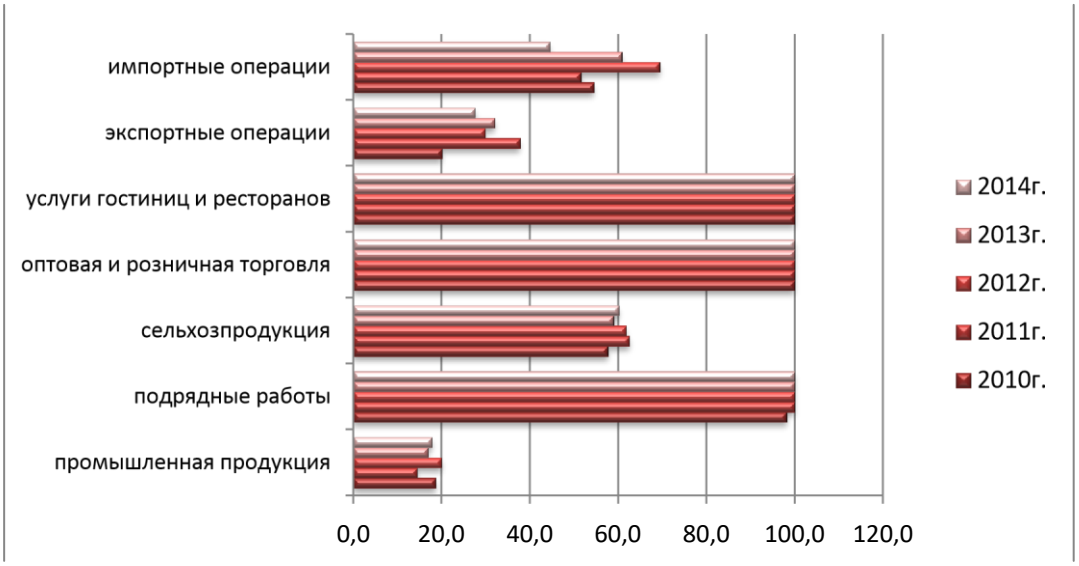


Рис. 3 – Удельный вес показателей в общем их объеме области, %

Предприятия МСП Кыргызстана в целом за период с 2010 по 2014г.г. имели скачкообразный неравномерный сальдированный финансовый результат. Однако основные финансовые показатели деятельности малых и средних предприятий Чуйской области отражают позитивную ситуацию. Так, в целом по Кыргызстану отмечена отрицательная динамика сальдированного финансового результата: снижение с 2010 по 2014 г.г. составило 240%. При общем снижении показателя сальдированного финансового результата предприятия МСП Чуйской области смогли повысить данный показатель за тот же период в 3,5 раза. По объемам выручки Чуйская область занимает вторую позицию после предприятий МСП города Бишкек [2, стр. 99]: 6842, 1 млн. сомов и 18792,8 млн. сомов.

Обобщая результаты оценки воздействия сектора МСП на развитие Чуйской области, можно сформулировать следующее.

В общем количестве предприятий сектора МСП Кыргызской Республики предприятия Чуйской области имеют долю около 10%, при этом основным представителем МСП области является малый бизнес. Основную долю в общем количестве предприятий МСП области занимают промышленные предприятия: их удельный вес в общем количестве предприятий области составляет почти 33%. По основным направлениям деятельности субъектов МСП в реальном секторе области происходили положительные изменения, объемы увеличивались за пятилетний период. Несмотря на значительное количество промышленных предприятий сектора МСП области, в результатах развития области они представлены недостаточно. Сектор МСП Чуйской области на сто процентов обеспечивает сферу торговли, услуг гостиниц и ресторанов, а также подрядных работ. Внешнеэкономическая активность предприятий МСП низка, уступая позиции крупным предприятиям и государственному сектору.

Список литературы

1. Сорокина В. Государственное регулирование малого бизнеса: опыт

Великобритании // Проблемы теории и практики управления – 2008. № 2. – С. 101-103.

2. Национальный Институт стратегических исследований Кыргызской Республики.

Электронный ресурс. Режим доступа – http://www.nisi.kg/ru-analytics-1319

3. Малое и среднее предпринимательство в Кыргызской Республике: 2010-2014 Нацстатком Кырг.Респ., 2015 – 220с.

4. Социально-экономическое положение *Чуйской области в январе - декабре* 2015. Источник: Национальный статистический комитет КР. С. 46. / Режим доступа: http://www.stat.kg/ru/statistika-chujskoj-oblasti/

5. Национальные счета: национальный статистический комитет. Режим доступа:

http://www.stat.kg/ru/statistics/nacionalnye-scheta/

УДК 342.553:330.131.5

РОЛЬ ГОСУДАРСТВА И МЕСТНОГО САМОУПРАВЛЕНИЯ В РЕГУЛИРОВАНИИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРЕОБРАЗОВАНИЙ

Алибаева Дамира Какеновна, *соискатель, КГТУ им. И.Раззакова, Кыргызстан, 720044, г. Бишкек, пр. Мира 66,* e-mail: dk_alibaeva@mail.ru

Цель статьи - посвящена обоснованию роли государства и местного самоуправления в регулировании процесса реформ в стране и поиску возможных путей и механизмов государственной поддержки развития социально-экономических преобразований на страновом и региональном уровне.

Ключевые слова: государственная поддержка, органы местного самоуправления, социально-экономические преобразования, стратегия экономического роста, социальная мобилизация.

THE ROLE OF THE STATE AND LOCAL SELF-GOVERNMENT IN THE REGULATION OF SOCIO-ECONOMIC TRANSFORMATION

Alibaeva Damira *, competitor , KSTU named after I.Razzakov , Kyrgyzstan , 720044, Bishkek c., prospect Mira 66 ,* e-mail:. dk_alibaeva@mail.ru

The purpose of the article зis devoted to the justification of the role of state and local government in the regulation of the reform process in the country and possible ways and mechanisms of state support for the development of socio-economic reforms in the country and regional level.

Keywords: government support, local government, social and economic reforms, growth strategies, social mobilization.

Рыночный механизм экономики в Кыргызской Республике, как и во всем мире, не может быть панацеей от всех социально-экономических бед: безработицы, бедности, инфляции, банкротства предприятия и т.п. Еще со времен Адама Смита классическая политэкономия настаивала на минимизации государственного вмешательства в экономику, а роль государства в этом плане ограничивалась функциями организации общественных работ, обеспечения прав собственника и экономической свободы индивидов, в результате чего, доля государственных расходов в ВВП индустриальных стран к началу XX века составляла

лишь 10 %. Сама по себе рыночная система хозяйствования не может быть целью – она лишь может служить средством для достижения более высокого уровня развития страны. Причем это возможно при строго определенных условиях, прежде всего при организации повсеместного и результативного контроля за управлением общественными финансами и собственностью. Любая модель общественного развития предполагает контроль. Государственный контроль и регулирование – неотъемлемая функция любой системы управления.

Какова же роль государства и местного самоуправления в регулировании социальноэкономических преобразований?

Задача государства состоит не в том, чтобы обеспечивать и поддерживать экономический рост с помощью бюджетных расходов, а в том, чтобы предоставить индивидам, предпринимателям и хозяйствующим субъектам инструменты, посредством которых они смогут получать отдачу от предпринимаемых действий. Роль государства в этом усматривается в повышении конкурентоспособности страны, в создании способствующей этому правовой и институциональной среды, а также в координации усилий хозяйствующих субъектов по достижению ими конкурентных преимуществ. Приоритетной же задачей Кыргызской Республики на сегодняшний день является достижение экономического роста совместно со структурными сдвигами. При этом важным представляется не количественные параметры, а политика, гарантирующая это структурное сближение.

Таким образом, встает вопрос о выработке стратегии экономического роста.

Обсуждение этого вопроса в научных кругах высветило ряд проблем.

Первое: при решении стратегических вопросов социально-экономического роста специалисты опираются на опыт зарубежных стран или прошлый опыт. Такой подход не полностью корректны, поскольку он не всегда учитывает особенности страны, современные реалии, неопределенности и динамику современных рынков. Поэтому часто перенятый успешный где-то опыт оборачивается неуспехом.

Второе: основным показателем при анализе и планирования социальноэкономических результатов считаются цифры, однако они не могут быть абсолютным мерилом жизненного уровня населения и социального прогресса. Конечно, они необходимы в экономических вопросах. Но в современных реалиях нашей страны, когда Нацстатком опирается на цифры, полученные от предприятий без гарантии объективности предоставляемой информации, статистические данные часто могут отражать не реальные значения цифр, а лишь тенденции и динамику их изменения.

Третье: Модель экономической политики Кыргызской Республики должна соответствовать современному состоянию ее экономики. Однако, по мнению В.Мау, большинство республик бывшего СССР реализует стратегию «догоняющего роста» [1, стр. 116] Россия и Казахстан, наши партнеры, осуществляют стратегию «экономического прорыва»; Кыргызстан же решает вопросы роста доли сельского хозяйства и промышленности в народном хозяйстве, о повышении роли индустриального сектора, что присуще для государств с доминированием аграрного сектора.

В работе доктора экономических наук, профессора Мусакожоева Ш.М [2, стр. 21] определен ряд экономических функций, которые оптимальным образом могут быть выполнены именно государством:

- защита и формирование основных принципов рыночной экономики;
- обеспечение товаров и услуг общественного пользования;
- учет побочных последствий;
- помощь отдельным социальным группам населения; стабилизация экономики и обеспечение экономической безопасности.

Помимо этого, надо выделить функцию регулирования отдельных отраслей, сфер и секторов экономики. Пережитый кризис на рынке продовольственных товаров Кыргызстана (рынок зерна и муки) подтвердил необходимость разработки мер государственного регулирования отраслей экономики, в т.ч. реального сектора (и не только в рамках фискальной функции государства) как одной из важнейших целей экономической политики, обеспечивающую экономический рост.

Проведение активной государственной политики должно быть направлено на *недопущение деиндустриализации республики*, на *сохранение промышленного и научнотехнического потенциала*.

По мнению автора, в настоящее время Кыргызстан может использовать следующие три источника экономического развития регионов реальном секторе:

1. ориентир на удовлетворение внутреннего спроса;
2. ориентир на экспорт;
3. развитие высокотехнологичных (наукоемких) секторов.

Первые два направления государство стимулирует уже сегодня, однако влияние третьего еще слишком мало. Целесообразно при выявлении приоритетов в развитии регионов ориентироваться не на более прибыльные и преуспевающие отрасли, имеющие сегодня наилучшее финансовое положение, а на наиболее перспективные в будущем. Сектор, ориентированный на удовлетворение внутреннего спроса, могут представлять строительство, телекоммуникации и ЖКХ. Если говорить об экспортоориентированном реальном секторе, то приоритетными отраслями могут быть электроэнергетика, горнодобывающая, легкая и пищевая промышленности.

Для активизации темпов экономического роста на основе внутреннего спроса, по мнению Кумскова В.И. [3, стр. 55], *необходимо первоначально небольшое, но самоусиливающееся изменение денежных средств*, направленное на совершенствование инфраструктуры экономики. По его расчетам, в республике на сегодня не хватает денежной массы, что затрудняет увеличение внутренних инвестиций. Например, помощь государства требуется в инвестировании в массовое строительство жилья и дорог, модернизацию ЖКХ, а также в развитие нового бизнеса. Выбор перечисленных объектов не случаен. Во-первых, вложения внутренних инвестиций в них создают огромный мультипликативный эффект (появляются обслуживающие отрасли с новыми рабочими местами). Во-вторых, доступное жилье и работающее ЖКХ существенно поднимут качество жизни населения при любом существующем уровне доходов. В-третьих, расширяющаяся дорожная сеть увеличит мобильность капитала и услуг труда, даст дополнительный толчок развитию предпринимательства, превратит малоосвоенные территории, поглощающие дотации и субсидии, в потенциал экономического роста, увеличит эффект масштаба для экономики в целом.

Государственная помощь *требуется в социальной сфере*. Как известно, экономический рост с одновременным улучшением социального положения населения – необходимое условие для обеспечения прогресса в развитии нашей страны. Главная цель социального обеспечения – дать свободу тем, кто стремится и в состоянии эффективно работать.

Важная роль в этом принадлежит *развитию социальной мобилизации населения в сельской местности*. Основная цель при этом – мобилизовать местное население в форме их собственных организаций, продвигать их развитие через их собственные и другие ресурсы и активно участвовать в процессе принятия решений для улучшения их жизни на местах.

Значение социальной мобилизации, которая была включена в Программу развития народонаселения Кыргызской Республики еще на период до 2005 года, заключается в сокращении бедности и обеспечении устойчивого и справедливого экономического роста. К числу важнейших предпосылок развития социальной мобилизации относится *формирование*

общинных организаций на уровне Айыл окмоту, которые бы широко сотрудничали с местными общинами. Здесь необходимо обучение населения, предоставление жителям села консультационных услуг, в том числе на безвозмездной основе по широкому кругу вопросов.

Деятельность таких общинных организаций может охватывать и *выполнение приоритетных проектов на базе микрокредитования* или мини-грантов с внутренним накоплением средств для дальнейшего развития. При этом план развития села согласовывается с Айыл окмоту для обеспечения комплексного подхода и определения доли вклада в финансирование со стороны организации, Айыл Окмоту и Программы децентрализации, если она задействована. Помимо укрепления финансового состояния общинных организаций надо расширять их поле деятельности. Например, создание приоритетных на местах объектов инфраструктуры, улучшение систем ирригации, водоснабжения, реализация социальных проектов. При этом часть финансирования может покрываться за счет грантов международных организаций или Правительства.

Ввиду сказанного, целесообразно *разрабатывать концепцию развития социальной мобилизации областей*, в каждом регионе. Региональную политику в отношении социальной мобилизации надо строить на основе консолидации интересов и ресурсов различных структур, от первичных общинных организаций и органов местного самоуправления до районных и областных.

Государственная политика по социальной мобилизации в регионах должна ставить целью создание условий для развития и эффективного использования потенциала сельских жителей для решения жизненно важных вопросов самостоятельно путем мобилизации и использования социального капитала.

Государственная политика в области социальной мобилизации должна осуществляться на основе таких принципов, как:

- признание социальной значимости общинных организаций;
- открытость приоритетных направлений деятельности общинных организаций;
- интеграция общинных организаций и органов местного самоуправления;
- концентрация сбережений от социальной мобилизации на приоритетных направлениях;
- стимулирование создания и активной деятельности общинных организаций через поощрительные меры органов местного самоуправления различного уровня.

При этом государственная и муниципальная политика по социальной мобилизации может быть направлена на обеспечение их устойчивого функционирования, осуществление системы мер по повышению бедности, выявление проблем, решение которых обеспечит развитие территории, формирование целевых программ преодоления бедности и развития территории и обеспечение информированности населения территории о принципах работы общинных организаций.

Следует коснуться и основных *направлений совершенствования региональной политики*.

Одной из основных проблем в региональной политике является сопряжение экономической политики на государственном и региональном уровне. Ряд специалистов считает, что макроэкономическая политика находится в некотором противоречии с интересами региональных и местных властей, которые вынуждены решать конкретные проблемы жизнеобеспечения, которые не могут быть решены без работы по организации производства и привлечения инвестиций. Причиной этому является *неполное соответствие принципу субсидиарности*, т.е. максимальной реализации на каждом уровне территориальной иерархии тех задач и полномочий, которые могут быть наиболее эффективно реализованы именно на этом уровне.

Надо муниципализировать различные аспекты государственной политики, в частности реализацию социальной политики. *Перенос части функций государственного*

управления на муниципальный уровень с делегированием соответствующих доходных полномочий позволит обеспечить усиление роли государства в экономике без существенного увеличения государственных расходов, обеспечив в большей учет и реализацию общественных интересов.

Государственное регулирование социально-экономических преобразований должно базироваться как на анализе результатов их за предыдущие периоды, так и на тщательном планировании дальнейших шагов. В этой связи важным инструментом государственной политики представляется *организация прогнозных исследований для выявления ключевых проблем социально-экономического развития*, выбора основных направлений политики в этой области как на государственном, так и на региональном уровнях и эффективной реализации ее мероприятий.

Региональный аспект прогноза должен, по мнению автора, разрабатываться на трех уровнях: территориальном (прогноз комплексного социально-экономического развития территории), территориально-отраслевом (прогноз развития отраслей в территориальном разрезе) и в межрегиональном (с выходом на макроэкономические показатели). При этом первый из них должен разрабатываться самими регионами и служить основой для двух последующих. Важна взаимоувязка с прогнозами и программами на государственном уровне.

Таким образом, роль государства и местного самоуправления в регулировании социально-экономических преобразований состоит в том, чтобы предоставить хозяйствующим субъектам и предпринимателям, способным стать двигателем преобразований, инструменты, посредством которых они смогут получать отдачу от предпринимаемых действий. Роль государства в этом усматривается в создании способствующей этому правовой и институциональной среды, а также в координации усилий хозяйствующих субъектов по достижению ими конкурентных преимуществ. Приоритетной же задачей Кыргызской Республики на сегодняшний день является достижение экономического роста совместно со структурными сдвигами.

Список литературы

1. Электронный ресурс: www.president.kg/files/docs/nsur_rus..doc
2. Мау В. Экономическая стратегия: выбор новой модели//Ведомости, 19.01.2013г., www.ane.ru
3. Мусакожоев Ш.М., Нарматова Н.Б. Государственное регулирование национальной экономики (на материале Кыргызстана)//Бишкек, 2006.
4. Кумсков В.И. На пути к рынку//Бишкек: КыргНИИНТИ, 1994.
5. Кузнецова, О. В. Экономическое развитие регионов: теоретические и практические аспекты государственного регулирования. Изд. 4 / О. В. Кузнецова. – М., 2007. – 304 с

УПРАВЛЕНИЕ КАРЬЕРОЙ И ФОРМИРОВАНИЕ КАДРОВОГО РЕЗЕРВА В ОРГАНИЗАЦИИ

Бакытова Нурзат Бакытовна, *магистрант,ФуиБ КНУ им.Ж.Баласагына Бишкек, Кыргызстан ,720000 мкр.Учкун 62, e-mail bakytova.1992, 0707441771*

В статье рассматриваются способы совершенствования системы управления карьерой, а также доказывается их значимость в успешной деятельности организации. В статье также рассматривается обеспечения справедливой, беспристрастной и законной кадровой политики, руководители будут наблюдать и координировать каждый и все аспекты кадровой политики включая комплектование штата, найма на работу, распределение обязанностей, оценку выполняемой работы, компенсацию, дисциплину и увольнение. Никаких действий в

отношении сотрудников не может предприниматься без оформления надлежащих документов, одобренных руководителем. Кроме того, Финансовый Отдел не может начинать какое-либо действие денежного характера по отношению к сотруднику без соответствующей документации, предоставленной сотрудниками УЧР/Офис Менеджером и одобренной руководителем.

Ключевые слова: персонал; совершенствование; стимулирование; система

CAREER MANAGEMENT AND THE FORMATION OF PERSONNEL RESERVE IN THE ORGANIZATION

Bakytova Nurzat Bakytovna *undergraduate, FuiB KNU im.Zh.Balasagyna Bishkek, Kyrgyzstan, 720000 mkr.Uchkun 62, e-mail bakytova.1992, 0707441771*

This article discusses ways to improve the career management system, and proved their importance in the success of the organization. The article also examines the provision of fair, impartial and legitimate personnel policy, management will monitor and coordinate each and all aspects of personnel policies including staff recruitment, hiring, assignment of responsibilities, evaluation of work, compensation, discipline and dismissal. No action against the staff can not be taken without proper registration documents approved by the head. In addition, the Finance Department can not begin any pecuniary nature action against any employee without proper documentation provided by the staff of HR / Office Manager and approved by the supervisor.

Keywords: staff; enhancement; stimulation; system

Управление карьерой состоит из планирования карьеры и преемственности руководства. Планирование карьеры - это такое продвижение работников в организации, которое соответствует ее потребностям и зависит от показателей труда, потенциала и преимуществ конкретных сотрудников предприятия.

Планирование преемственности руководства осуществляется для того, чтобы гарантировать, что у организации есть менеджеры, которые ей необходимы для удовлетворения ее будущих потребностей.

Планирование карьеры можно рассматривать как формирование одной из частей системы продвижения персонала.

Управление карьерой. Цели.

•Гарантировать, что потребности организации в преемственности управления удовлетворяются

•Обеспечить перспективным работникам обучение и практический опыт, который вооружит их для работы на том уровне ответственности, которого они способны достичь.

•Дать имеющим потенциал работникам рекомендации и оказать поддержку, которые им нужны, если они хотят реализовать свой потенциал и сделать успешную карьеру с помощью организации и своих талантов и стремлений

Кадровый резерв одним из элементов управления карьерой является отбор специалистов в кадровый резерв руководства. В настоящее время способность выявлять и успешно готовить на резервисты должны быть способны:

а. Временно заменить соответствующую должность – при производственной необходимости.

б. Занять соответствующую должность – при появлении вакансии.

Создание кадрового резерва ставит себе целью:

Систематизацию и совершенствование системы замещения, подбора и расстановки кадров на управленческие должности.

Улучшение качественного состава управленцев.

Своевременное удовлетворение потребности в управленческих кадрах.

Повышение уровня мотивации и лояльности сотрудников.

Данная система может быть использована другими операционными отделами, в которых требуется создание кадрового резерва.

Принципы формирования кадрового резерва

При формировании кадрового резерва должны соблюдаться следующие принципы:

а. Объективность. Оценка профессиональных и личностных качеств и результатов профессиональной деятельности кандидатов для зачисления в кадровый резерв должна осуществляться коллегиально на основе объективных критериев оценки.

б. Справедливость. Зачисление в кадровый резерв должна осуществляться в соответствии с личными способностями, уровнем профессиональной подготовки, результатами профессиональной деятельности и на основе равного подхода к кандидатам.

в. Добровольность включения и нахождения в кадровом резерве.

г. Прозрачность. Гласность в формировании и работе с кадровым резервом. Кандидаты в Резервисты, не проходящие этапы отбора, будут информированы о причинах письменно через уведомление по электронной почте.

Порядок формирования Кадрового резерва

Поскольку каждая должность имеет свою специфику, процесс формирования Кадрового резерва не всегда будет одинаков. Формирование Кадрового резерва включает в себя следующие этапы:

1. Объявление о наборе в кадровый резерв
2. Отбор кандидатов
    * Отбор заявлений в соответствие с заданными критериями
    * Тестирование
    * Групповое интервью
    * Собеседование
3. Отбор кандидатов
    * Ранжирование текущих
    * Отбор на кадровый резерв на базе ранжирования
    * Тестирование
4. Теоретический тренинг
5. Практика
6. Замена отсутствующего менеджера или назначение на должность

Отбор кандидатов

5.1. Отбор кандидатов на должность должен проводиться на базе следующих критериев. Переход от одного этапа на следующий должен происходить, только если пройден предыдущий шаг:

а. Отбор заявлений должен проводиться строго по критериям, указанным в Объявлении о наборе.

б. Изучение личного дела заявителя

в. Тестирование:

* Знание кредитного руководства – Он-лайн тест по кредитному руководству и кредитным/депозитным продуктам. Проходной балл – по утвержденному на тот момент в Банке;

* Знание этических норм – Он-лайн тест , Кодекс Проходной балл – по утвержденному на тот момент;

г. Групповое интервью – метод определения у кандидатов способности управлять коммуникациями, отстаивать собственные позиции, принимать или отрицать обоснования других участников, проявлять лидерские качества или быть ведомым.

д. Собеседование – личная беседа с каждым кандидатом, для определения наличия отдельных качеств, необходимых для менеджера (суждение, личные показатели в работе и т.п.).

е. Рекомендация руководителя – мнение о кандидате, прошедшем все этапы.

Отбор кандидатов должен проводиться на базе следующих критериев. Переход от одного этапа на следующий должно происходить только если пройден предыдущий шаг:

а. Ранжирование – классификация среди действующих на соответствие должности и наличие потенциала карьерного роста, которая проводится на базе последних аттестаций и собственных наблюдений при работе, на тренингах, различных собраниях или иных инициативах.

б. Изучение личного дела заявителя, проверка через БД УЧР и устная справка- на наличие дисциплинарных взысканий или иного поведения, неприемлемого для менеджера.

6. Теоретический тренинг

6.1. Теоретический тренинг резервистов на должность будет проходить 5 рабочих дней, и включает в себя следующие темы:

№ Тема

1. Переход от одной должности на другую
2. Эмоциональный интеллект
3. Роли Менеджера в Банке и управленческие навыки
4. Управление рисками
5. Анализ потенциала рынка
6. Управление кредитным портфелем
7. Активное продвижение услуг
9. Базовые компьютерные навыки
10. По требованию времени

7.1. Резервисты на все должности должны пройти практическое обучение по следующей схеме:

8. Замена менеджера или назначение на должность.

8.1. Замена и назначение Резервиста будет происходить по мере возникновения необходимости замены/вакансии в любом филиале. Кроме этого Резервисты будут включаться в рабочие группы по улучшению или внедрению различных инноваций или пилотных проектов в филиале и Банке.

Резервист, назначенный Комитетом по КР на должность, не будет проходить больше других этапов отбора.

Выбор Резервиста на должность на возникшую вакансию будет происходить на базе:

Предыдущих 3-х оценок Резервиста при аттестациях;

Оценки Резервиста во время его прохождения практики;

Различных показателях на момент назначения;

Непосредственный руководитель Резервиста дают свою письменную рекомендацию на момент назначения;

Если мнение Непосредственного руководителя сильно отличается от мнения, высказанного на этапах обучения практики или замены, то окончательное решение по нему

принимается после этого Комитетом по КР. Все исключения по отбору Резервистов должны будут проводиться по решению или согласованию с Комитетом по КР.

Для обеспечения справедливой, беспристрастной и законной кадровой политики, руководители будут наблюдать и координировать каждый и все аспекты кадровой политики включая комплектование штата, найма на работу, распределение обязанностей, оценку выполняемой работы, компенсацию, дисциплину и увольнение. Никаких действий в отношении сотрудников не может предприниматься без оформления надлежащих документов, одобренных руководителем. Кроме того, Финансовый Отдел не может начинать какое-либо действие денежного характера по отношению к сотруднику без соответствующей документации, предоставленной сотрудниками УЧР/Офис Менеджером и одобренной руководителем.

Сотрудник УЧР, на основании предоставленных работником документов вносит в Персональную Базу Данных данные о сотруднике, а также открывает и хранит личное дело каждого сотрудника.Каждое личное дело содержит документы, касающиеся работы служащего и является конфиденциальным. При необходимости, сотрудник может просмотреть в своем личном деле документы, но только те, на которых стоит его её подпись. Руководители отделов также имеют право просматривать личные дела своих сотрудников.

В некоторых случаях руководители может оставить за собой право хранения личных дел своих непосредственных подчиненных. Пред.Правления также может оставить за собой право хранения крайне конфиденциальной информации из личного дела сотрудника. За исключением случаев, упомянутых в руководствах.

Многие предприятия придерживается политики продвижения своих сотрудников по возможности на более высокую, более ответственную должность. При этом учитываются интересы и профессиональные способности сотрудника. После одного года успешной работы в своей должности сотрудник имеет право подать анкету на вакантное место. В редких случаях, для хорошо зарекомендовавшего себя и трудолюбивого сотрудника, рукодители может сделать исключение из этого правила. Кроме того, рукодители поощряет сотрудников, которые обращаются с анкетами на новую должность. При равных профессиональных качествах кандидатов, предпочтение будет отдаваться кандидату от сотрудников предприятия. Однако, должность, занимаемая кандидатом не является гарантией того, что этот сотрудник будет принят на новую должность. Во внимание, обычно, принимаются многие факторы. Рукодители оставляет за собой право нанимать наиболее квалифицированных людей и будет время от времени помещать объявления на вакантные должности и нанимать наиболее квалифицированные кадры со стороны.

Список литературы

1. Бухалков М.И. Управление персоналом: Учебник. - М.: Инфра-М, 200 - 368 с.

2. Кротова Н.В., Клеппер Е.В. Управление персоналом: Учебник. - М.: Финансы и статистика, 2005. - 320 с.

3. Лифшиц А.С. Основы управления персоналом: учеб. пособие, ИвГУ, 1995. - 95 с.

4. Лифшиц А.С. Оценка и развитие управленческого персонала: научное пособие, Иваново: ИвГУ, 1999, 186 с.

5. Макарова И.К. Управление персоналом: учебник - М.: Юриспрюденция, 2002. - 304 с.

6. Музыченко В.В. Управление персоналом. Лекции: Учеб. для студентов высш. учеб. заведений/В.В. Музыченко. - М.: Издат. центр «Академия», 2003. - 528 с.

УДК 330. 101.

РАЗВИТИЕ ПЕНСИОННОЙ СИСТЕМЫ В КР

Бектурганова Карима Арыковна, *к.э.н., доцент, КГТУ им. И. Раззакова. Кыргызстан, 720044. г.Бишкек, пр. Мира 66.*

Цель статьи – раскрытие проблем развития пенсионной системы, необходимости проведения её реформирования в целях повышения уровня системы пенсионного обеспечения, усиления её эффективности и справедливости. Проблема пенсионного обеспечения в условиях политических и экономических реформ явно не были приоритетными в повестке дня. Но дефицит ресурсов пенсионного фонда КР привел к неизбежности совершенствования действующей пенсионной системы. Процесс ее реформирования рассматривается в соответствии с новой Концепцией развития пенсионного обеспечения КР.

Ключевые слова: солидарная, условно-накопительная, накопительная пенсионная система, финансирование пенсионного обеспечения, госбюджет, дотация, валовый внутренний продукт, демографическая ситуация, инвестиционный доход, накопительный компонент, страхование, персонифицированный учет, индексация пенсий, негосударственный пенсионный фонд, социальный фонд, рынок пенсионных услуг.

THE DEVELOPMENT OF THE PENSION SYSTEM IN THE KYRGYZ REPUBLIC

Bekturganova Karima Arykovna, *Ph.D., associate professor, KSTU after name I. Razzakova. Kyrgyzstan, Bishkek 720044. st. Mira 66.*

The purpose of the article is to expose the problems of pension system, the need for reform to improve the pension system , improve efficiency and equity.

The problem of pension systems in terms of political and economic reforms was clearly not a priority on the schema. But the lack of resources of the pension fund of the Kyrgyz Republic has led to the obligatory improvement of the existing pension system. The process of reform is considered in accordance with the new concept of pensions Kyrgyz.

Keywords: solidarity, notional, funded pension system, the financing of pensions, state budget subsidies, gross domestic product, the demographic situation, investment income, funded pillar, insurance, security administration, indexation of pensions, pension fund, the social fund, the market pension services.

В большинстве стран мира преобладают распределительные формы пенсионной системы, которая основана на принципе солидарности поколений. Особенность этой системы заключается в том, что пенсии пенсионерам формируются за счет работающих граждан. Эта система называется «солидарной пенсионной системой», также существуют условно-накопительная пенсионная система (УНС) и накопительные пенсионные системы.

Названные три системы представляют различные формы существующих в мире пенсионных систем. Несмотря на это, они имеют единую цель – обеспечить достойную безбедную старость для значительной части населения. Для успешного достижения этой цели, поступления и выплаты в пенсионной системе должно быть сбалансированными. Поэтому достижения долгосрочной финансовой устойчивости пенсионной системы являются основными целями любой реформы.

В Кыргызской ССР, до распада Советского Союза (1991г), действовала универсальная солидарная система пенсионного обеспечения, как и в других республиках СССР: низкий

пенсионный возраст (60 лет для мужчин и 55 для женщин), универсальность и относительно высокая ставка замещения (не менее 63%). В основе этой системы лежали принципы выслуги лет и финансирования пенсионного обеспечения из средств госбюджета за счет общих налоговых поступлений. В 1991 г. расходы на социальное обеспечение в Кыргызстане составляли 6% ВВП, и республика получала значительные дотации из Москвы. После распада Советского Союза ВВП страны снизился – до 52% от уровня 1990 г., а дотации из центра прекратились. Доля сельского хозяйства, неофициальной торговли и сектора услуг, которые вносили наименьший вклад в пенсионный фонд, возросла в ВВП страны, а доля производства, транспорта и коммуникаций, которые ранее были самыми существенными источниками пополнения пенсионного фонда, сократилась. В результате, несмотря на относительно низкую численность пенсионеров среди населения Кыргызстана - около 10%, пенсионное обеспечение оказалось тяжелым бременем для государства. В 1994 г. из-за дефицита ресурсов пенсионного фонда средний размер пенсии в Кыргызской Республике снизился до 70% от уровня 1990 г. Задержка пенсионных выплат и выдача пенсии в виде товаров и услуг стали широко распространенной практикой.

Государственная система предоставления социальных гарантий населению в целом, системы пенсионного обеспечения, в частности, в истории современного суверенного Кыргызстана стала сложной проблемой. Для выхода из сложившейся критической ситуации в обществе возникла необходимость реформирования практически всех областей жизнедеятельности молодого кыргызского государства.

Кыргызстан стал одним из первых стран СНГ в области проведения реформы пенсионной системы. Предпринимаются определенные меры по восстановлению финансовой устойчивости пенсионной системы. Но из-за слаборазвитости финансового рынка и нехватки денежных средств КР не удалось ввести полноценную пенсионную систему на основе взносов в пенсионный фонд или разработать систему добровольного пенсионного страхования через частные пенсионные фонды. Она выбирает так называемую «условно- накопительную систему» (УНС), согласно которой пенсионные взносы «аккумулируются» на виртуальном индивидуальном пенсионном счете работника, но не сберегаются. Вместо этого они использовались для выплаты пенсий пенсионерам.

УНС значительно ограничивала перераспределение. Это создавало стимулы для того, чтобы работающие делали взносы в пенсионный фонд в течение трудовой жизни. Вместе с тем предпринимаемые усилия по реформированию (переходу на УНС) не изменили пенсионную систему. В рамках данной реформы в 1997 г. в пенсионное законодательство Кыргызстана был внесен ряд законодательных изменений:

- постепенное повышение пенсионного возраста с 60 до 63 лет для мужчин и с 55 до 58 лет для женщин;
- аннулирование ряда льгот, в частности, раннего выхода на пенсию для представителей некоторых профессий и групп населения;
- трехкомпонентная пенсионная система: гарантированная базовая пенсия каждому пенсионеру с достаточным стажем; пенсионное обеспечение в соответствии с трудовым вкладом работника, исчисляемое как процент от средней заработной платы до 1996 г.; индивидуальные пенсионные взносы, накопленные на счету пенсионера.

Предпринятые меры в некоторой степени помогли сократить долги по выплатам пенсий. Расходы пенсионного фонда снизились с 6,9 % от ВВП в 1996 г. до 5% в 2008 г. Многомесячные задержки по выплатам пенсий были устранены. Пенсионная система стала в какой –то мере эффективнее защищать пожилых людей от крайней степени бедности: домохозяйства с пенсионерами на 20% меньше подвержены риску бедности, чем домохозяйства без пенсионеров. Во многом этому способствовало и общее оздоровление экономики страны. Однако пенсии все еще оставались низкими, уровень замещения

(соотношение средней пенсии к средней заработной плате) составлял лишь половину уровня замещения до провозглашения независимости Кыргызстана. Пенсии половины пенсионеров Кыргызстана ниже прожиточного минимума. По информации социального фонда, 52% пенсионеров получает пенсию ниже прожиточного минимума в 4100 сомов (2008г.)

В целях совершенствования пенсионной системы принимаются Законы и постановления Правительства: Закон «О государственном пенсионном социальном страховании», Закон «О пенсионном обеспечении военнослужащих», «О статусе судей», «О прокуратуре КР», Постановление Правительства КР «Положение о пенсиях за особые заслуги».

В настоящее время размеры надбавок к пенсиям установлены в общем от 50 до 1000 сомов. Назначение и выплата надбавок к пенсиям осуществляется в соответствии с действующим Законом «О государственном пенсионном социальном страховании» из средств республиканского бюджета, т.е. за счет средств налогоплательщиков. Следует заметить, что в некоторых из указанных документов, например, Закон «О статусе судей» др., содержатся очевидные перекосы в поддерживаемой государственной политике в области пенсионного обеспечения. Они подрывают основы доверия у населения к пенсионной системе в целом, и сводятся на нет все положительные усилия определенных этапов реформ по обеспечению справедливости, адекватности и эффективности действующей системы пенсионного обеспечения Кыргызстана. В связи с этим предлагаются и альтернативные модели пенсионной системы для Кыргызстана.

Вопрос о том, какая модель пенсионной системы наиболее приемлема для КР, остается открытым. Решая этот вопрос, следует учитывать положительную национальную традицию материальной поддержки пожилых более молодыми членами семьи. Такие семейные «механизмы страхования», дополненные простой пенсионной системой при минимальном участии государства могут стать более эффективным решением для страны, не имеющей пока перспектив превратиться в промышленно развитое государство.

Модели пенсионной системы в нашей республике динамичны, запускаются все новые программы для улучшения жизни граждан. Одна из таких программ обещающая обеспеченную старость-это введение накопительной части в пенсионную систему. Накопительная пенсионная система – это новая модель отношений субъектов пенсионной системы, основанной на накопительных методах финансирования, которая базируется на создании долгосрочных накопительных капитализируемых ресурсов, что связано с долгосрочным по времени нахождением средств после их внесения. Необходимость введения новой модели связана с изменением демографической ситуации в стране.

Введение накопительной части в пенсионную систему направлено на решение проблемы прямой зависимости пенсионной системы от проблем политического и демографического плана, которым подвержена распределительная система. При накопительной системе страховые взносы, аккумулирующиеся в пенсионной системе за счет платежей работника и его работодателя, не расходуются на выплаты сегодняшним пенсионерам, а накапливаются, инвестируются и приносят доход до тех пор, пока плательщик не выйдет на пенсию. Пенсионные отчисления не являются формой налоговых изъятий, все сбережения плательщика и весь его инвестиционный доход, полученный на эти сбережения, являются его личной собственностью, которая обеспечит выплату пенсии. Эта система обязывает задействованное в ней население нести самостоятельную ответственность за уровень своего дохода после выхода на пенсию, так как источником пенсионных выплат станут сформированные ими на индивидуальных пенсионных счетах накопления. Кроме того, каждому гражданину предлагается возможность за счет добровольных пенсионных взносов увеличить и тем самым обеспечить себе более высокий доход после завершения трудовой деятельности.

Обладая рыночными свойствами капитала, накопление пенсионных средств поможет увеличить уровень сбережений в экономике, послужит толчком к развитию финансового рынка, будет способствовать росту долгосрочных национальных сбережений, стимулировать создание финансовых учреждений, которые необходимы для экономического развития. На рынке пенсионных услуг появится институт пенсионного страхования. Пенсионная реформа окажет положительное воздействие на объем инвестиций в стране, пенсионные фонды станут основным источником внутренних инвестиций. Рост рынка капитала способствует появлению значительных свободных средств для новых капиталовложений, что со временем приведет к возрастанию уровня производства и занятости.

Согласно Концепции введения накопительной части в пенсионную систему Кыргызстана, одобренной Указом Президента страны «О мерах по введению накопительной части в пенсионную систему Кыргызской Республики» от 24- сентября 2008 года УП №339, новая пенсионная система состоит из трех компонентов:

1. Государственная обязательная солидарная пенсионная система;
2. Обязательная накопительная пенсионная система;
3. Добровольная профессиональная и индивидуальная накопительная система.

Первый компонент состоит из перераспределительной пенсионной схемы, которая включает базовую пенсию первой и второй страховой частей пенсии.

Второй компонент обязателен для граждан КР старше установленного законом возраста, с которого разрешается трудовая деятельность в КР работающих в организованном секторе (в государственных учреждениях, частных компаниях различной организационно-правовой формы). Первоначально граждане перечисляют страховые взносы на накопление 2% от фонда заработной платы. Тариф для обязательного государственного накопительного пенсионного фонда будет рассматриваться и утверждаться каждый год. В целом будет поддержана тенденция к увеличению ставки тарифов страховых взносов в накопительную часть и постепенному переходу на накопление всех отчислений работника. В целях аккумулирования уплачиваемых страховых взносов в накопительную часть пенсионного фонда для дальнейшего инвестирования Социальным фондом Кыргызской Республики создан Государственный накопительный пенсионный фонд.

Третий компонент состоит из добровольной профессиональной и индивидуальной накопительных схем. Профессиональная накопительная пенсионная схема позволит крупным работодателям или объединениям по отраслевому или профессиональному признаку организовать собственную накопительную пенсионную схему, включающую различные корпоративные, общие или специальные накопительные тарифы. При этом предполагается для работодателя и работника, уплачивающих страховые взносы в добровольную накопительную систему, предусмотреть льготное налогообложение при уплате налога на прибыль и подоходного налога. Индивидуальная накопительная пенсионная схема предполагает индивидуальные накопительные взносы в пределах ставки тарифов страховых взносов для работника. Добровольная уплата страховых взносов в Государственный накопительный пенсионный фонд может осуществляться физическими лицами, юридическими лицами в пользу физических лиц в любом размере, но не менее 2 процентов от среднемесячной заработной платы. Эти взносы дают каждому гражданину возможность увеличить накопления в индивидуальных пенсионных счетах. Участники накопительной пенсионной системы имеют право выбора различных инвестиционных портфелей, представляемых управляющей компанией (организация, которая управляет средствами инвестиционных фондов и обеспечивает инвестирование накопительной части пенсий), а также самих управляющих компаний и негосударственных пенсионных фондов для размещения пенсионных накоплений. Гарантия сохранности пенсионных накоплений закреплена Законом КР «О государственном пенсионном социальном страховании».

Новая «Концепция» определила необходимые шаги по переходу к накопительной пенсионной системе: 1) подготовка соответствующей нормативно-правовой базы; 2) создание ликвидного рынка государственных ценных бумаг и государственного долга; 3) достижения стабилизации финансовой устойчивости пенсионной системы; 4) не снижение пенсионного возраста; 5) снижение ставки взносов работодателей с 19% до 17,25%, с увеличением взносов работника с 8% до 10%, при этом 2% направляются в накопительный фонд; 6) предоставление налоговых льгот в целях стимулирования добровольного пенсионного страхования.

Пенсионная система Кыргызской Республики все еще находится в процессе постепенного реформирования, для успешности которого следует определить общие и специфичные проблемы действующей системы. К общим проблемам относятся 1) низкий уровень пенсий, 2) уклонение от отчисления пенсионных взносов или отсутствие стимулов отчислять взносы, 3) уязвимость пенсионной системы от демографических изменений и экономического развития. Среди специфичных проблем: 1) пенсионные отчисления работников неформального сектора, работающих как в стране, так и за рубежом, 2) недоверие населения Правительству, 3) неразвитость рынков капитала и государственные долги. Все эти проблемы ставят под вопрос возможность создания устойчивой системы пенсионного обеспечения в долгосрочной перспективе. Были предприняты попытки реформирования пенсионной системы, но они не достигли цели- долгосрочной устойчивости. Они только восстановили устойчивость системы в краткосрочном периоде.

Первого октября 2014 г. вице – премьер –министр Кыргызстана Эльвира Сариева отметила, что действующая сегодня в республике пенсионная система страдает серьезными перекосами, однако предполагаемая новая концепция ее реформы поможет устранить недостатки.

В настоящее время пенсии кыргызстанцев формируются из нескольких составляющих. Первая –базовая. Это гарантированная социальная помощь государства, которая выплачивается из госбюджета. Сейчас она составляет 1500 сомов при наличии 25летнего трудового стажа у мужчин, а женщин -20 летнего. На эти цели выделяется внушительная сумма, так как в стране насчитывается 585 тысяч пенсионеров, из них около 100 тыс. человек получают пенсию по инвалидности.

Вторая составляющая - страховая. Она в свою очередь, делится на две части. Одна - начисляется соответственно размеру заработной платы за любые проработанные подряд 60 месяцев до 1996 года. Другая - формируется из суммы страховых взносов на индивидуальном счете гражданина после введения персонифицированного учета, то есть с 1996 года.

В будущем будет еще и третья часть- накопительная. Пока ее никому не выплачивают. Она только формируется и инвестируется. Этот компонент- два процента от заработной платы- ввели в 2010 году. Стоит отметить, что такие отчисления делают не все работающие граждане КР. В формировании накопительной части пенсии участвуют мужчины и женщины, которым до пенсии остается более 10 лет. За меньший срок накопить ощутимый капитал на счете невозможно. К тому же национальные фондовые рынки недостаточно развиты, чтобы инвестировать в них, а затем получать ощутимый доход за малый срок. Все-таки пенсионные средства являются долгосрочными активами, доходность которых можно будет оценить только через 15-20 лет.

Определенное значение в развитии накопительной системы имеет рынок. Рынок негосударственных пенсионных услуг в КР пока развит слабо. Закон «О негосударственных пенсионных фондах» приняли еще в 2001 году. Несмотря на это, в настоящее время существует всего один такой фонд -«Кыргызстан». Главная причина – низкие доходы населения и недоверие населения частным финансовым организациям. Масштабы частных пенсионных накоплений в экономике остаются более чем незначительными. Поэтому

реформа пенсионной системы предусматривает и меры по развитию негосударственных фондов. Надо отметить, что в новой пенсионной концепции впервые показано их значение. Полноценная и надежная работа фондов не только повысит общий уровень пенсионного обеспечения в стране, но и поможет внутреннему инвестиционному рынку и повысит финансовую грамотность населения, будет способствовать развитию рынка пенсионных услуг.

Вкладывать денежные средства в пенсионный фонд можно в любом возрасте. Можно начать откладывать средства, будучи совсем молодым и получать пенсию в 30 или 50 лет. Схема такова, что в течение 5 лет вкладываешься, а по истечению 5 лет получаешь свои деньги. Тогда как в обязательном страховании в Соц.фонд, есть возрастные ограничения.

Еще одной важной проблемой для пенсионной системы республики является трудовая миграция. Сегодня от 500 тысяч 1 миллиона наших граждан работают за пределами Кыргызстана. Гарантии прав граждан государств-участников Содружества в области пенсионного обеспечена. предусмотрены соглашением стран СНГ (1992г.). Согласно ему, за пенсионное обеспечение отвечает государство, его представляющее, а взаимные расчеты не производятся, если иное не предусмотрено двусторонними соглашениями.

Доходы граждан, которые работают за рубежом, просто выпадают из пенсионной системы. Сегодняшние мигранты, которые не пополнят пенсионный бюджет, в будущем выйдут на пенсию, и будут требовать минимальную пенсию согласно Конституции Кыргызстана. Это – их право.

Действующей в республике пенсионной системе присущи серьезные недостатки. Необходимо проведение ее реформирования. Главная цель-повышение общего уровня системы пенсионного обеспечения, усиление ее эффективности и справедливости. В качестве основной сохранится накопительная система, доля которой будет преобладать в общем размере пенсии и носить всеобщий и обязательный характер для всех граждан КР. Накопительный же компонент как обязательный, так и добровольный, станет играть роль дополнительного «бонуса» к основному размеру пенсии. Между тем в 2015 г. дотации из бюджета на выплату пенсий возрастут с 15 до 17,5 миллиарда сомов. Социальный фонд сегодня не справляется с функцией сбора страховых отчислений. По словам премьер-министра Дж. Оторбаева «около 52% объема заработных плат выплачивается сегодня «в конвертах». Поэтому в текущем 2014 г. Правительство приняло решение передать функции сбора социальных платежей в налоговые органы - соответствующий законопроект уже разработан. Таким образом, мы избежим дублирования и усилим администрирование сборов».

Пенсионная система в целом имеет динамичную структуру, она постоянно совершенствуется. В 2010 г. в Кыргызстане была проведена пенсионная реформа, главным достижением которой считается появление обязательного накопительного компонента в дополнение к страховой части пенсии. Уровень зависимости пенсионной системы от демографических показателей на данном этапе развития снижен.

Дальнейшее развитие пенсионной системы является предметом пристального внимания руководства страны.

«Национальной Стратегией устойчивого развития КР на период 2013- 2017 годов», утвержденной Указом Президента Кыргызской Республики от 21 января 2013 года №11, определены основные приоритетные направления развития пенсионной системы.

В рамках выполнения программы по переходу Кыргызской Республики к устойчивому развитию на 2013-2017 годы, разработанной Правительством Кыргызской Республики в целях реализации Национальной Стратегии устойчивого развития Кыргызской Республики до 2017 года Социальным фондом разработан проект Концепции развития системы пенсионного обеспечения Кыргызской Республики, который был направлен в

Аппарат Правительства КР и рассмотрен на Наблюдательном совете по управлению государственным социальным страхованием.

Для развития направлений реформирования действующей пенсионной системы Кыргызской Республики разработаны мероприятия, направленные на совершенствование тарифной политики государственного социального страхования, на оптимизацию расходов республиканского бюджета на пенсионное обеспечение, на совершенствование политики индексации пенсий в целях усиления страхового принципа, на дальнейшее развитие накопительного компонента пенсионной системы, на совершенствование персонифицированного учета, на проработку политики по привлечению трудовых мигрантов- граждан Кыргызской Республики в государственную систему социального страхования, на разработку комплекса мероприятий по стимулированию развития негосударственных пенсионных фондов в Кыргызской Республике, и др.

Выводы: В Кыргызстане действует многоуровневая пенсионная система, она имеет динамичную структуру. Проводится пенсионная реформа, главным достижением которой считается появление обязательного компонента в дополнение к страховой части пенсии, а также добровольное частное пенсионное страхование. Благодаря многокомпонентности повышается ее финансовая устойчивость и защищенность от различных рисков, в том числе и демографического характера. На сегодняшний день такая структура признана самой оптимальной в мире всеми международными организациями, которые специализируются на вопросах социального и пенсионного страхования.

Список литературы

1. Закон «О государственном пенсионном социальном страховании» от 30 июня 1997г. поправки 04.11.2003; 06.14.2013г.
2. Указ Президента страны «О мерах по введению накопительной части в пенсионную систему Кыргызской Республики» от 24 сентября 2008 года УП №339
3. Закон «О негосударственных пенсионных фондах» от 07.05.1998г. поправки 21.07. 2014 г.
4. Национальная Стратегия устойчивого развития КР на период 2013-2017 годов 21 января 2013 г.
5. Постановление Правительства КР об утверждении Концепции развития системы развития пенсионного обеспечения КР до 2020года.
6. Продуманное реформа htt://www.rg.ru/2014/10/01 pensiahtml
7. Кыргызстан –на острие пенсионного кризиса: пора радикальных реформ htt://www.parus kg.info/2012/08/21/67535
8. Проблемы в существующей системе http:www.cafmi.kg/ru/analitic/593-pension – reform

УДК 005.95

ТЕХНОЛОГИИ ОТБОРА И НАЙМА ПЕРСОНАЛА

Таалайбекова Эркайым Таалайбековна, *Кыргызский Национальный университет имени Ж. Баласагына, магистрантка 2 – года обучения, г. Бишкек, ул. Боконбаева, 245, кв. 41, e- mail: etaalaybekova@mail.ru, т. 0552382007*

Цель данной научно – исследовательской статьи – подробное изучение технологий профессионального найма и отбора кадров в на предприятии. Автором подробно

рассмотрена и проанализирована сущность технологий, методов, инструментов. Тема научно – исследовательской статьи является актуальной на сегодняшний день. При рыночной конкуренции отбор персонала - важнейший фактор, который определяет экономическое положение и способность к выживанию организаций. Актуальность статьи определяют, с одной стороны, нарастающий интерес к теме в современной науке, с другой стороны, ее недостаточная разработанность. Рассмотрение вопросов с данной тематикой имеет как теоретическую, так и практическую значимость. Кадровая политика в области отбора кадров состоит в определении принципов, методов, критериев отбора работников и последующей их адаптации, необходимых для качественного выполнения заданных функций, методологии закрепления, профессионального развития персонала. Подбор персонала - наиболее ответственный этап в управлении персоналом, так как ошибка обходится слишком дорого.

Ключевые слова: технология, обращение, интервьюирование, отбор, психометрия, кадровая служба, метод, инструмент

TECHNOLOGY SELECTION AND RECRUITMENT

Taalaibekova Erkaiym Taalaibekovna*. Kyrgyz National University is named after Yusuf Balasaghuni, second year master student, 245 Bokonbaeva street, Bishkek, Kyrgyzstan, e – mail: etaalaybekova@mail.ru, t. 0552382007*

The purpose of this scientific - research article - a detailed study of the professional set of technologies and selection of personnel in the organization. The author discussed in detail and analyzed the essence of each technology, methods, tools. The theme of this scientific - research article is the actual nature of today. Recruitment in a competitive market is the most important factor that determines the economic situation and the ability to survive organizations. The relevance of the article is determined, on the one hand, the growing interest in the topic in modern science, on the other hand, it Insufficient developed. Consideration of issues with this subject has both theoretical and practical significance. Personnel policy in the field of personnel selection is to define the principles, methods and criteria for the selection of employees and their subsequent adaptations necessary for quality performance of the specified functions, consolidation methodology, professional development of staff. Personnel - the most important stage in human resource management, since the error is too expensive.

Keywords: technology, treatment, interviews, selection, psychometry, HR Service, a method, a tool.

Основной целью найма и отбора персонала является получение работников, наиболее хорошо подходящих под стандарты качества работы, выполняемой организацией. К этому также добавляются необходимость обеспечения удовлетворенности работников и полного раскрытия и использования их возможностей [1, с.447].

Технология представляет собой определенную последовательность выполнения той или иной деятельности без допущения ошибок.

Все технологии имеют различные инструменты.

Технология обращения использует два метода: резюме и анкета соискателя. Главной целью данная технология преследует осуществление отбора претендентов.

Опрос или анкетирование – это наиболее распространенный метод оценки мотивации и управления карьерой персонала. Он дает возможность за короткий срок получить нужную информацию о мотивации основной части сотрудников.

Кадровая служба заранее подготавливает анкету, состоящей из вопросов, которые должны быть сформулированы согласно законодательным нормам. Заполнение данной анкеты и является основной содержащей процедуры анкетирования. Очень редко анкетирование выступает как единственный метод. Анкеты, которые хорошо продуманы и структурированы, содержат ключевую информацию, которая имеет отношение к работе, и помогают специалистам кадровой службы при проведении первого собеседования. На сегодняшний день все больше используются анкеты, имеющие разделы, которые посвящены способностям претендентов. Эти анкеты содержат не только вопросы, касающиеся биографии соискателя, но и те, которые помогают выявить наиболее значимые для этой должности и компании способности претендента. Как правило, если кандидат претендует на руководящую должность, ему необходимо предоставлять резюме вместе с сопроводительным письмом. Достоинствами анкетирования являются: быстрое получение информации и низкие финансовые затраты. В то же время данный метод приводит к искажению информации, в виде социально желательных ответов, ошибок при разработке анкет и просчетов в процессе проведения опроса. Это может привести к недостаточной достоверности полученной информации. [2, с. 264]

Ознакомление с резюме, хотя и позволяет дать оценку навыкам кандидата излагать информацию в письменном виде и узнать какими коммуникационными навыками он обладает, все же данная технология имеет некоторые недостатки. Одним из таких недостатков является отсутствие установленного формата резюме, в связи с чем кандидаты могут предоставлять разнообразные сведения о своем опыте. Это создает сложности в процессе отсеивания кандидатов по сравнению с использованием стандартной анкеты.

Технология – интервьюирование, то есть собеседование.

Беседа - это наиболее простой и надежный инструмент оценки особенностей мотивации работников. Поговорив с сотрудником, почти всегда можно составить представление об его отношении к работе и факторах, определяющих силу его мотивации. В процессе беседы всю необходимую информацию получают с помощью вопросов. При этом в ходе проведения беседы выделяют закрытые, открытые, косвенные, наводящие и рефлексивные вопросы. [3, с.207]

Наблюдение является самым доступным методом оценки мотивации и управления карьерой с установлением личного контакта с коллегами. Впоследствии возрастает роль мотивов, связанных с потребностями в должностном и профессиональном росте

Собеседование имеет два вида:

Первый вид – дисциплинарное собеседование. Он содержит вопросы, которые касаются условий работы. (график отпусков, командировки и так далее);

Второй вид - квалификационное, которое включает вопросы относительно профессиональной деятельностью.

Интервьюирование представляет собой вид собеседования, для которого характерны непродолжительный отрезок времени и предварительно подготовленный перечень вопросов. Внесение ясности по опросу с целью принятия решения принимать на работу кандидата или нет – это и есть основная цель данной технологии. [4, с. 118]

Различают такие интервью, как традиционное и структурированное.

Сущность традиционного интервью заключается в определении важных для работы характеристик претендента. Данный вид собеседования характеризуется тем, что полагается на интуицию при анализе пригодности кандидата и использует наводящие вопросы, отвечать на которые нужно «да» или «нет».

Структурированное собеседование, в отличие от традиционного, больше уделяет внимание на факторы, связанные с работой. При таком виде интервьюирования задаются

открытые вопросы, на которые требуются глубокие ответы, к примеру, способность человека представить свое поведение в той или иной ситуации.

Менеджеру по подбору персонала при проведении собеседования стоит обратить особое внимание на физические данные, образование, опыт,

Следующая технология – технология отбора, предполагает пробные задания. В данную технологию включен метод отборочных тестов, которые используются как дополнительный инструмент. Необходимы хорошая организованность данного метода и соответствие его определенной системе.

Робертсон и Кандола предложили систему, которая предусматривала четыре категории:

1.психомоторные задания, которые включали в себя физические манипуляции над объектами;

2.сведения, которые непосредственно связаны с работой и проверяют уровень осведомленности кандидата о работе; [5, с.209]

3.самостоятельное принятие решений – искусственно создаются ситуации, с которыми кандидату придется сталкиваться в процессе работы, и кандидату нужно показать свои способности быстро и самостоятельно принимать решения для разрешения данной ситуации;

4.командное обсуждение и принятие решений – кандидаты разбиваются по отдельным группам для обсуждения и принятия решений, при этом оценивается вклад каждого участника в разрешение проблемы..

Пробные задания имеют большое количество преимуществ, обеспечивая проверку опытности и компетентности претендентов для данной должности еще до приема на работу. Но есть и недостатки: предусмотренный для одного вида деятельности тест может не подойти для других, требует немалых расходов для выполнения тестов.

Психометрия. Метод данной технологии – психологические тесты. Специалистами по отбору персонала используются два основных их типа: тесты, определяющие познавательные умения инструменты для оценки характеристик личности. Использование того или иного метода психометрического тестирования зависит от его технической адекватности и степени соответствия требованиям к персоналу, сформированным по результатам изучения рабочего процесса сотрудников, а не от речей, убедительно произнесенных продавцами тестов.

Тесты обычно представляют собой буклеты с вопросами и отдельные бланки для ответов. При тестировании можно использовать незаконченные предложения, набор фотографий или рисунков. По результатам тестирования делают заключение об особенностях мотивации работника/

Тестирование - это стандартизированное испытание с целью выявления или оценки тех или иных психологических особенностей работника. Разрабатывают тесты для определения особенностей мотивации конкретных работников и степени выраженности у них необходимых характеристик.

Технология альтернативы. Альтернатива представляет собой возможность выбрать одного из нескольких кандидатов на вакантное место. Здесь, как наиболее эффективный, выделяется метод «аквариума», но в силу своей дороговизны используется крайне редко - обычно для отбора специалистов топ-уровня.

Реализация указанного метода требует наличия специального помещения со столом в центре, за которым размещены кандидаты, а по периметру - эксперты. В составе экспертов могут быть:

•Топ - менеджеры, президент, вице-президент, с которыми будет сотрудничать потенциальный сотрудник

•Специалисты по персоналу, юристы, психологи, и пр., которые анализируют способности кандидатов в решении поставленных ими задач.

В качестве задачи может послужить ситуация, с которой компания сталкивалась ранее, либо стоит сейчас. По итогам оценки экспертов делается выбор наиболее достойного претендента. Вся процедура занимает непродолжительный отрезок времени – обычно около получаса.

Метод экспертных оценок заключается в оценке мотивации работников теми людьми, которые достаточно хорошо их знают (например, руководство и коллеги). В качестве экспертов можно привлекать деловых партнеров или клиентов. Обычно экспертная оценка мотивации является элементом комплексной оценки сотрудника.

Для успешного отбора огромное значение имеет определение критериев и принципов, на основании которых будет приниматься решение о преимуществах соискателей. При установлении критериев отбора должны быть соблюдены следующие требования: валидность, полнота, надежность, необходимость и достаточность критериев. Основной принцип отбора и расстановки кадров: "Нужный человек, в нужное время, на нужном месте".

Список литературы

1. Кибанов, А. Я. Основы управления персоналом : учеб. для вузов / А. Я. Кибанов.- 2-е изд., перераб. и доп. - М. : ИНФРА-М, 2011. - 447 с. - ISBN 978-5-16-002854-5.

2. Лукичева Л.И. Управление персоналом: Учеб. пособие. – М.: Омега-Л, 2011. – 264 с.

3. Мордовин С.К. Модульная программа №16 "Управление человеческими ресурсами", 2011. – 207 с.

4. Сидорчук А.П. Методы оценки и принятия решения о приеме кандидата на работу // Московское научное обозрение. – 2011. – № 5(9). – С. 118 – 127.

5. Уткин Э.А.. Профессия – менеджер. – М.: Экономика, 2012. - 209 с.

УДК 81.373.46

ТЕХНИКАЛЫК ТЕРМИНДЕРДИН ЖАСАЛЫШЫ

Исмаилов Асанбек Усөналиевич, *И.Раззаков атындагы КМТУнун Кыргыз тили кафедрасынын башчысы, доцент.Кыргызстан, Бишкек ш. E-mail:*

В каждом направлении науки, в технике, производстве, политике, искусстве и.т.д. должны быть группы терминов принадлежащие к определённым группам включенных в систему. Такая система терминов называется **терминологией**. К примеру, в научном языке используются следующие термины: грамматика, лексика, орфография, звук, буква, согласные, гласные, омоним, синоним, предложение и .т.д. Такие языковые возможности используются в различных отраслях техники, производства, искусства так же физических, математических, химических, биологических, медицинских, философских и экономических науках. Цель статьи - раскрыть некоторые решения происхождения технических терминов и переводов в наше время.

Ключевые слова: терминология, предложение, орфография, наука, язык, звук.

Ар бир илим тармагында, техникада, өндүрүштө, саясатта, искусстводо ж.б. өзүнө таандык терминдердин системага салынып, атайын кабыл алынган белгилүү топтору болууга тийиш. Терминдердин мындай системасы **терминология** деп аталат. Алсак, тил илиминде

төмөнкүдөй терминдер колдонулат: грамматика, лексика, орфография, тыбыш, тамга, үндүү, үнсүз, омоним, синоним, сүйлөм ж. б. Мындай тилдик каражаттар физика, математика, химия, биология, медицина, философия, экономика ж.б. илимдердин, ошондой эле техниканын, өндүрүштүн, искусствонун түрдүү салааларында кеңири колдонулат. Бул баяндаманын максаты - азыркы учурда техникалык багыттагы терминдердин жасалышы жана которулушу боюнча айрым маселелерди ачып берүү болуп саналат.

Негизги сөздөр: терминология, сүйлөм, орфография, илим, тил, тыбыш.

Ismailov Asanbek Usonbekovich, *associate professor of department of kyrgyz language, KSTU named after I.Razzakova.*

In every direction of science, in a technique, production, politics, art and other there must be groups of terms belonging to the certain groups plugged in the system. Such system of terms is named terminology. For example, in a scientific language next terms are used: grammar, vocabulary, orthography, sound, letter, consonants, vowels, homonym, synonym, suggestion and .etc. Such language possibilities are used in different industries of technique, production, arts similarly physical, mathematical, and chemical, biological, medical, philosophical and economic sciences. Aim of the article - to expose some decisions of origin of technical terms and translations in our time.

Keywords: terminology, suggestion, orthography, science, language, sound

Киришүү. Термин — илимдеги же техника багытындагы, өндүрүштөгү белгилүү бир түшүнүктүн официалдуу түрдө кабыл алынган наамын туюндуруу үчүн колдонулган сөз жана сөз айкаштарынын бир түрү. Бул латынча terminus («чек аралык белги», «чек ара», «чек») деген сөздөн алынган. Башка сөздөргө салыштырганда, терминдер өзүнө таандык бир катар мүнөздүү белгилери менен айырмаланат. Жалпысынан алганда, биз күндөлүк турмушта колдонуп жүргөн сөздөрдүн кыйла бөлүгү көп маанилүү келет, алар ар түрдүү эмоционалдык, экспрессивдик кубулушка ээ болуп, көркөм каражат катары да колдонулат.

Мындай көрүнүштөр терминдерге, чынында, ылайык келүүчү белгилерден эмес. Бир маанилүүлүк (однозначность), нейтралдуулук жана белгилүү бир адистикке, илим тармагына тиешелүү болуп колдонулган конкреттүүлүк — булардын бардыгы терминдерге жалпы мүнөздүү болгон башкы белгилерге жатат. Тилдин сөздүк составындагы башка (термин эмес) сөздөр менен салыштыра келгенде, терминдерге мүнөздүү болуп турган бирден-бир өзгөчө касиет деп мына ушуларды эсеп кылууга болот.

Негизги бөлүк. Термин бир маанилүү болуп, белгилүү бир түшүнүктүн гана аты катары колдонулушу керек деген талап, чынында, практикалык зарылдыктан улам келип чыккан. Эгерде термин көп маанилүү сөздөрдүн эсебинен жаралса же ал терминологиялык мааниден тышкары дагы башка мааниге ээ болуп турса, анда чаташуулар пайда болуп, конкреттүүлүк сапатынан ажырашы мүмкүн. Ошондуктан терминдердин ар дайым бир гана мааниге ээ болуп колдонулушу жогоркудай практикалык максатты көздөгөн касиет катары бааланат. Бирок мындай талап бардык эле учурда мүмкүн боло бербегендиктен, термин жана термин эмес сөздөрдүн арасында өз ара алмашуу, биринен экинчисине көчүп туруу сыяктуу карым-катыштарга да жол берилип келе жаткандыгы белгилүү. Ошол себептен улуттук тилдердин жалпы маанидеги айрым сөздөрү термин катары да колдонулуп келүүдө. Буга кыргыз тилинде лингвистикалык термин катары колдонулуп келе жаткан ээ (сүйлөмдүн ээси), *уңгу* (сөздүн уңгусу), *мүчө* (сөздүн же сүйлөмдүн мүчөсү), *муун* (сөздүн мууну), же болбосо математикалык терминдерге айланган *кошуу, алуу, бөлүү, көбөйтүү, бөлчөк* сыяктуу сөздөрдү мисалга келтирсек болот. Булар ушул аталган илим тармагына түздөн-түз

байланыштуу болуп колдонулган учурда жогоркудай чектелген мааниде болуп, ошол тармак жагына тиешеси бар конкреттүү бир түшүнүктүн атын туюндурат.

Адатта сырттан келип кирген сөздөрдөн турган терминдердин көп маанилүү болуп колдонулушу анча мүнөздүү көрүнүш эмес. Ошондуктан булардын ар бири кандай терминологиялык системага тиешелүү болуп турса, ошого ылайыкталган конкреттүү бир түшүнүктүн атын туюнтат. Маселен, кыргыз адабий тилиндеги *атом, молекула, телескоп* сыяктуу физико-техникалык терминдерди, же болбосо *фонема, морфема, артикуляция, орфоэпия, агглютинация* сыяктуу лингвистикалык терминдерди алып көрсөк, булардын ар бири, кандай контекстте турса да, өзүнүн бир гана маанисинде колдонулат. Андан өзгөчөлөнгөн кандайдыр бир кошумча мааниге булардын эч бирөө ээ эмес.

Терминдердин бардыгы тең, жогоруда мисалга келтирилгендей, жеке сөздөрдөн турбайт. Ар бир илим тармагында, техникада, өндүрүштө ж.б. термин катары колдонулуп келе жаткан сөз айкаштары (алар көп составдуу татаал аттардын тобуна кирет) да аз эмес. Буга төмөнкүдөй орусча терминдерди мисал кыла кетсек болот: *точка кипения, сила притяжения, переменный ток, прямой угол, двигатель внутреннего сгорания, совершенный вид* (в глаголе), *производительные силы, прибавочная стоимость, промышленный капитал, финансовый капитал, торговый оборот* ж. б. Мындай структуралык формадагы терминдер кыргыз тилинде да бар: *жарым өткөргүч, тең салмактуулук, тартылуу күчү, тик бурчтук, өндүргүч күчтөр, капиталдык салым, кошумча нарк, географиялык чөйрө* ж. б.

Чыгыш теги боюнча алып караганда, терминологиялык системанын ар бири ар түрдүү булактардан жаралгандыгы, буга улуттук тилдин төл сөздөрү да ички булак катары тартылып, ошондой эле сырттан келип кирген лексикалык кабыл алуулар да кошулгандыгы жөнүндө айта кетүүгө туура келет. Терминдердин жалпы системасы, улуттук тилдин өзүндөгү ички мүмкүнчүлүктөрдү пайдалануу менен бирге, кийинчерээк башка тилдерден оошуп кирген сөздөрдүн эсебинен да толукталып келген.

Кыргыз адабий тилинин терминдик мааниге ээ болуп, белгилүү бир терминологиялык системанын карамагына өткөн төл сөздөрү же ушундай эле милдет аткарып, кийин термин катары да колдонула баштаган араб-иран сөздөрү туурасында биз төмөнкүлөрдү мисалга келтире кетсек болот. Булар термин жаратуунун ички булагы катарында эсептелүүгө тийиш.

Кыргызча терминдердин мындай булактан жаралып чыгышына, чынында, орус тилинин да таасири тийбей койгон жок. Анткени мындай терминдердин көпчүлүк бөлүгү бул тилдеги ушул сыяктуу терминдерди калькалоонун (кыргыз тилине сөзмө-сөз которуунун) натыйжасында жаралып келгендиги талашсыз. Алсак, *баяндооч, аныктооч, толуктооч* деген терминдер орус тилиндеги *сказуемое, определение, дополнение* сыяктуу грамматикалык терминдердин түздөн-түз үлгүсүнө салынып жасалган. Муну аныкташ үчүн ушул терминдердин эки тилдеги материалдык негиздерин, түзүлүш структурасын өз ара салыштырып көрүүнүн өзү эле жетиштүү болуп турат. Дагы: *салмаксыздык, тең салмактуулук, кубаттуулук, кылмыштуулук, үч бурчтук, төрт бурчтук* — булар орусча*невесомость, равновесие, мощность, преступность, треугольник, четыреугольник* сыяктуу терминдердин үлгүсүндө жаралган. Ал эми *амал* (арифметиканын төрт амалы), *менчик* (мамлекеттик менчик, жеке менчик) орусча *действие, собственность* сыяктуу терминдердин семантикалык үлгүсү боюнча терминдик мааниге ээ болгон. Составдык бөлүктөрдөн куралган татаал түзүлүштөгү төмөнкүдөй терминдер да, негизинен, калька аркылуу жаралган: *айыл чарбасы, эл чарбасы, дыйкан чарбасы, кичи ишкана, чекене баа, мамлекеттик камсыздандыруу, жеңил өнөр жай, оор өнөр жай, куралдуу күчтөр, татаал гүлдүүлөр, сүт эмүүчүлөр, өздүк көркөм чыгармачылык, көп чекит, көш чекит, үтүрлүү чекит, эркин күрөш* ж. б.

Көптөгөн адистер, окумуштуулар термин жаратууда башка тилдердин даяр үлгүсүн пайдаланып, ошол аркылуу эне тилдеги ички мүмкүнчүлүктөрдү аракетке келтирип, лексикалык составдын жаңы сөз байлыктары менен толукталышына өз салымдарьш кошуп

келишкен. Алсак, улуу окумуштуу М. В. Ломоносов өз эне тилиндеги бир катар сөздөрдү чет тилдердин үлгүсүнө салып, жаңыча мааниге өткөрүү аркылуу *движение, кислота, наблюдение, опыт, явление* ж. б. терминдерди жаратып, илимий чөйрөгө сунуш кылгандыгы белгилүү. Термин жаратуунун мындай практикасы орус турмушунда кийинки мезгилде да тынымсыз өнүгүп отурган.

Ушундай эле көрүнүш кыргыз тилинен да кеңири орун алган. Эне тилибиздин төл сөздөрүн кылдат пайдаланып, кыргыз тилинин грамматикалык терминдерин элге түшүнүктүү, практикалык жактан ийкемдүү кылып жаратууга белгилүү окумуштуу, мамлекеттик ишмер Касым Тыныстанов көп эмгек сиңирген. Анын колунан жаралган грамматикалык терминдер окуу китептеринде, илимий эмгектерде азырга чейин ийгиликтүү колдонулуп келе жаткандыгы белгилүү. Кыргыздын төл сөздөрүн химиялык терминдерди жаратууга кенири колдонуп, аларды окуу китептерине кийирген алгачкы авторлордун бири белгилүү педагог Султан Арбаев болгон. Мындай терминдер кыргыз тилинде жазылган же орус тилинен кыргызчага которулган башка окуу китептеринен да орун алып келди.

Дүйнө жүзүндөгү көптөгөн тилдерде сырттан кабыл алынган терминдердин саны бир топ эле арбын. Булар туурасында «Кабыл алынган сөздөр» деп аталган буга чейинки параграфта да сөз болгон эле.

Илим менен техникага, искусство менен саясатка ж. б. кеңири тараган терминдердин кыйла белүгү дүйнө жүзүндөгү көптөгөн тилдер үчүн орток экендиги, бардыгында бирдей колдонулуп келе жаткандыгы жалпыбызга белгилүү. Мындай интернационалдык мүнөздөгү терминдердин көпчүлүгүнүн материалдык негизи, тек жагынан алып караганда, грек, латын тилдерине барып такалат. Анткени дүйнөдөгү эң эле көп тилдерге таралган интернационалдык терминдердин жаралышына бул эки тил, чынында, өзгөчө роль ойноп келген. Буга төмөнкүдөй бир канча илимдердин наамын мисалга келтирсек болот: *философия, математика, филология, грамматика, синтаксис, логика, геометрия, психология* ж. б. Төмөнкүдөй илимий терминдер да ушул тилдерден кабыл алынган:*атом, метод, гипотеза, диафрагма, синтез, космос, метафора, термометр, монография, формула, радиус, конус, вакуум, меридиан* ж. б. Коомдук-саясий маанидеги *демократия, монархия, гегемония, диктатура, республика, референдум, меморандум, революция* сыяктуу терминдер да ушул топко кирет. Грек, латын тилдеринен келген терминдерге дагы төмөнкүлөрдү кошо кетүүгө болот: *драма, трагедия, комедия, поэзия, симфония, лирика, фабрика, адвокат, новатор, ребус, курс, субъект, объект, конституция, материализм, коммунизм, социализм* ж. б.

Азыркы кезде кеңири колдонулуп келе жаткан көптөгөн техникалык терминдер, жаңы пайда болгон илим тармактарынын наамдары да тек жагынан алганда грек, латын тилдерине түздөн-түз байланыштуу. Алар: *телефон, телескоп, протон, позитрон, нейтрон, протоплазма, астрофизика, биофизика, агротехника, радиотерапия, зоотехника* ж. б. Латын тили менен байыркы грек тили кийинки доордо өлүү тилдерге айланып кеткен. Бул эки тилде өз доорунда көптөгөн илимий терминдер, ошондой эле техникага, искусствого, саясатка тиешелүү терминдердин да бир канчасы жаралып келгендиги белгилүү. Кыйла жылдарга чейин латын тили илимдин тили болуп колдонулуп келген. Эки тилдин элементтерин (уңгуларын, сөз жасоочу каражаттарын) термин жаратуу жагына пайдалануу кийинчерээк көптөгөн европалык тилдер үчүн салтка айланган көрүнүш болуп кеткен. Мунун натыйжасында небакта өлүү тилдерге айланып калган грек, латын тилдеринин материалдык базасына негизделип жаралган көп сандаган жаңы терминдер пайда боло баштайт. Сырткы булактардын негизинде жаралган мындай терминдер ар бир тилдин өзүндөгү ички бөтөнчөлүктөргө карай тыбыштык түрүн өзгөртүп, ар түрдүү болуп колдонулбай коё алган эмес. Ошонун натыйжасында бир эле термин ар башка тилде ар түрдүү тыбыштык нормага салынып өзгөрүлгөндүгү белгилүү.

Корутунду. Жогоркудай эл аралык мааниге ээ болгон интернационалдык терминдер кийинчерээк башка европалык тилдерде да кабыл алына баштайт. Алсак, англис тилинин спортко байланыштуу төмөнкүдөй терминдери дүйнө жүзүнө кеңири таралган: *старт, финиш, футбол, баскетбол, боксер, спорт, спортсмен, тренер, хоккей, теннис, матч, тайм, аут.* Музыкага же жалпы эле искусство жагына байланыштуу көптөгөн терминдер итальян тилинен кабыл алынды: *соло, солист, соната, тенор, фортепиано, квартет, трио, либретто, новелла, ария, адажио.* Демек, азыркы учурда атоолордун жана терминдердин жасалышы жана которулушу боюнча чечиле элек дагы көп көйгөйлөр бар.

Колдонулган адабияттар

1. Кудайбергенов Т.Л. Кыргызча химиялык терминдер жана алардын табигыйтехникалык предметтер аралык байланышы //Кыргыз тили жана адабияты 14/2008.
2. Сулайманкулов К.С., Кудайбергенов Т.Т. Химиялык терминдердин орусча кыргызча сөздүгү. - Бишкек, 2003.
3. Кыргызстандагы котормо жана анын келечеги. Республикалык илимийпрактикалык конференциянын материалдары. – Б,: КТУ “Манас”, 2009.

УДК 811. 512. 154

КЫРГЫЗЧА СҮЙЛӨШҮҮ ЭТИКЕТИН ҮЙРӨТҮҮДӨ ТЕКСТТИ КАРАЖАТ КАТАРЫ ПАЙДАЛАНУУНУН ЖОЛДОРУ

Исраилова А.М., *И.Раззаков атындагы Кыргыз мамлекеттик университети, ББжИМ КР, Бишкек ш., e-mail:isirailova.akdana@bk.ru*

Макалада кыргыз тилин үйрөтүүдө текстти каражат катары колдонуу жолдору каралат. Колдонууга сунуш кылынган текст дидактикалык, коммуникативдик-маданият таануучулук, лингвистикалык-методикалык принциптер таянуу менен тандалып алынып, адаптацияланып берилиши керек. Текст менен иштөөгө карата түрдүү тапшырмаларды берүүнүн негизги максаты – студенттерге тексттин мазмунун түшүндүрүү, негизги оюн аныктоого үйрөтүү, кыргыз тилиндеги тексттердеги логикалык-грамматикалык байланыштарды аныктоого, тексттин темасын толук түшүнүүгө көнүктүрүү, алар аркылуу өз алдынча текст түзө билүүгө машыктыруу жана лексикалык компетенциясын байытууга багытталат. Кыргызча сүйлөшүү этикетинде колдонулуучу сөз, сөз айкашы жана сүйлөмдөр боюнча студенттердин жекече сөздүк-минимумун түздүрүү тууралуу сөз болот.

Негизги сөздөр: кыргызча сүйлөшүү этикети, маданият, саламдашуу салты, өзгөчөлүктөрү, жаш-курагы, сөздүн мааниси, пайдалануу, колдонуу, сунуш кылынган, дидактикалык, коммуникативдик-маданият таануучулук, лингвистикалык-методикалык принциптер, таянуу, тандала, адаптацияланат, берилет, иштөө, түрдүү, тапшырмалар, берүү, негизги максат, мазмунун түшүндүрүү, негизги оюн аныктоого үйрөтүү, логикалыкграмматикалык байланыштарды аныктоого, тексттин темасын толук түшүнүүгө көнүктүрүү, өз алдынча текст түзө билүүгө машыктыруу, лексикалык компетенциясын байытууга, кыргызча сүйлөшүү этикети, сөз, сөз айкашы, сүйлөмдөр, жекече сөздүк-минимумун, түздүрүү, тууралуу, сөз, болот.

ПУТИ ПРИМЕНЕНИЯ ТЕКСТА КАК СРЕДСТВО ОБУЧЕНИЯ КЫРГЫЗСКОМУ РЕЧЕВОМУ ЭТИКЕТУ НА УРОКАХ КЫРГЫЗСКОГО ЯЗЫКА

Исраилова А.М., *КГТУ им.И.Раззакова*

В статье рассматривается текст как средство обучения кыргызскому языку. Предложенный текст, должен быть адаптированным, он подбираться опираясь на дидактические, коммуникативно-культуроведческие, лингвистико-методические принципы. Основная цель заданий при работе с текстами, как средством обучения – объяснить содержание текста, научить определять основную смысл содержания и уметь составлять самостоятельно текст, а также направлено на обогащения лексического минимума студентов. Здесь также рекомендуется составление слов, словосочетаний, предложений, словаряминимума по кыргызкому речевому этикету

Ключевые слова: кыргызский речевой этикет, культура, традиции приветствия, особенности, возраст, значения слова, использование, предложенный текст, должен быть адаптированным, опирается, на дидактические, коммуникативно-культуроведческие, лингвистико-методические принципы, основная цель, задание,работа, средство обучения, объяснить содержание текста, научить, определять, основную смысл, содержания, уметь, составлять, самостоятельно, текст, направлено, на обогащения, лексический, минимум, здесь, рекомендуется, составление, слова, словосочетания, предложения, словаря-минимум, кыргызкий, речевой, этикет, индивидуально, о, будет,

USIND A TEXT AS A TOOL OF EDUCATION IN THE KYRGYZ SPEECH ETIQUETTE

Isirailova A.M., *The Ministry of the Education and Sciences of Kyrgyz Republic Kyrgyz State Technical University named after I.Razzakov, Bishkek, e-mail: isirailova.akdana@bk.ru*

The article contains an information which could be helpful for people who want to start learning Kyrgyz language as a second one. The following text is built on didactical, communicative-cultural, and linguistic-methodical studies. The main purposes of working with the article are firstly to explain its content, secondly to try to understand essential information given in it, thirdly to know how to make up a new text on your own and finally it would enlarge the students' lexical resources. What's more, there are given several references for students to learn more Kyrgyz language words, phrases and sentences.

Keywords: Kyrgyz language, speech etiquette, culture, custom (traditional) greetings, features, age, word meaning, usage, given text, needs to be adapted, leans, on communicativecultural, linguistic-methodical studies, main purpose, exercises, work, source of learning, content, to know how to do smth, make up, on your own, text, directed, on enlargement, lexical, minimum, here, it is recommended, building, words, phrases, sentences, minimum lexicon, Kyrgyz, speech, etiquette, individually(yourself).

Тилди эне тил же экинчи тил катары окутуунун методикасы илиминде соңку мезгилдерде текстке кайрылуу, текстти колдонуу, текст менен иштөө жана текст түздүрүү иштерине өзгөчө көңүл бурула баштады. Ал эми кыргыз тилин экинчи тил катары окутуунун методикасында бул проблемага К.Д.Добаев, С.К.Рысбаев, ЖА.Чыманов, З.Тагаева, Н.Бийгелдиева сыяктуу окумуштуу-методисттер кайрылышып, тилди окутууда текст менен иштөөнүн келечектүүлүгүн жана эффективдүүлүгүн баса белгилешүүдө. Мисалы, методистокумуштуу Ж.А.Чыманов кыргыз тилин окутууда тексттин алган ордун төмөнкүдөй аныктайт: "Текст – кыргыз тилин окутуунун мазмунун негизги өзөгү. Ал сабак процессинде каражат, дидактикалык материал, үлгү, булак ж.б. катары колдонулуучу универсалдуу жана окутуунун максатына жетишүүгө негизги тилдик-кептик татаал бирдик. Текст кыргыз тили мугалиминин колундагы, эгер ыктуу алдын ала ойлонулуп колдонулса, ары таасирдүү, ары кызыктуу лингвистикалык, адабий-эстетикалык, адеп-ахлактык жана башка дидактикалык,

максаттарды ишке ашыруучу курал болуп саналат”. Албетте, көркөм публицистикалык, айрым учурда илимий стилде жазылган тексттерди сүйлөшүү этикетин үйрөтүүдө колдонуу үчүн тексттин дидактикалык маани-маңызын, мүмкүнчүлүктөрүн толук таанып-билип, аны окутуу-үйрөтүү ишинде пайдалануунун ык-жолдорун “максаттуу тандап алуунун” мазмунун толук билүү зарыл.

Максат белгилүү болуп, аныкталгандан кийин аны ишке ашыруунун (үйрөтүүнүн) жагдай-шарттарын тактап алуу зарылдыгы туулат. Бул маселеде кыргыз тилин экинчи тил катары окутуунун методикасы боюнча белгилүү адис-окумуштуу С.К.Рысбаев: ”... сүйлөшүү маданияттуу жүзөгө ашышы үчүн адамдардан көп нерсе талап кылат”, – деп белгилеп, маданияттуу сүйлөшүүнүн он шартын ачык көрсөткөн.

Студенттер кыргызча сүйлөшүү этикетин толук өздөштүрүү, аны өз жашооишмердүүлүгүндө такай колдонууларына жетишүү үчүн алар кыргыз элинин саламдашуу салттарын билүү менен бирге кыргыз маданиятын, салтын, үрп-адатын ж.б. терең билүүлөрү керек. Албетте, бул сыяктуу кыргыз таанууга тиешелүү маалыматтардын өзүн окутупүйрөтүүнү кыргыз тили сабагы өз алдына максат катары койбойт. Бирок, ошол эле учурда маданияттуу, кыргызча сүйлөшүү этикетинин нормаларына ылайык пикир алышууга көнүгүү үчүн студенттер саламдашуунун камтыган маанилерин, аткарган кызматмилдеттерин да билүүгө тийиш. Бул сыяктуу татаал жана көлөмдүү материалды окутупүйрөтүүдө төмөнкү текст менен иштөөнү сунуш кылабыз.

Кыргызча сүйлөшүү этикетин үйрөтүүдө каражат катары колдонууга сунуш кылынган текст төмөнкүдөй принципке таянуу менен тандалып алынып, адаптацияланат: дидактикалык принциптер: сүйлөшүү этикетин үйрөтүүнүн максаты менен мазмунуна, кеп ишмердүүлүгүндөгү сүйлөшүү этикетинин алган ордуна студенттердин кызыгуу, өздөштүрүү деңгээлине, практикалык принциптерине карай;

коммуникативдик-маданият таануучулук принциптер кыргыз тилинде баарлашууда кандай учурда жана кандай максатка сүйлөшүү этикетине тиешелүү тил каражаттарын колдонула тургандыгына, сүйлөшүү этикетинин пикир алышууга тийгизген таасирине карай;

лингвистикалык-методикалык принциптер: кыргыз тилинин лексико-грамматикалык өзгөчөлүктөрүнө жана жаш-курагына карай тандалат.

Текст боюнча биринчи тапшырма лексикалык компетенцияны байытууга багытталат. Башкача айтканда 1) текстти окуп, түшүнгөнүн кыскача айтып берүү; 2) тексттеги маанисин түшүнбөгөн сөздөрдү дептерге көчүрүп жазуу; 3) өз ара жардамдашуу менен түшүнүксүз сөздөрдүн маанилерин чечмелөө; 4) түшүнүксүз болгон сөздөрдүн түшүндүрмөлөрүн окутуучу менен бирдикте тактап, сөздүктү түзүп чыгуу; 5) 2-3 студенттин текстти үн чыгарып, мүмкүн болушунча көркөм окуу.

Экинчи тапшырма: 1) тексттин темасын ачып көрсөткөн 2-3 сөздү тексттен көчүрүп жазгыла; 2) түшүнүксүз сөздөрдүн маанисин орус тилинде түшүндүрүп бергиле: салам, алик, арыбаңыз, бар болуңуз, эсенсизби, мал-жан аманбы.

Текст менен иштөөгө карата түрдүү тапшырмаларды берүүнүн негизги максаты – студенттерге тексттин мазмунун түшүндүрүү, негизги оюн аныктоого үйрөтүү, кыргыз тилиндеги тексттердеги логикалык-грамматикалык байланыштарды аныктоого, тексттин темасын толук түшүнүүгө көнүктүрүү, алар аркылуу өз алдынча текст түзө билүүгө машыктыруу.

Биринчи тапшырма. Тексттен саламдашуу жөнүндө маалыматтарды камтыган сүйлөмдөрдү тапкыла.

Экинчи тапшырма. Текстти маанилик бөлүктөргө ажыраткыла, ар бир бөлүккө ат койгула.

Үчүнчү тапшырма. Текст боюнча суроолорго жооп бергиле.

Төртүнчү тапшырма. Тексттин мазмунунун негизинде өз элиңде саламдашуу менен кыргыз элиндеги саламдашуунун ортосундагы жалпылыкты жана айырмачылыктарды аныктап бер. Алгылыктуу жактарын аныктап көрсөт.

Бешинчи тапшырма. Сол тараптагы сүйлөмдөрдүн уландысын оң тараптагы колонкалардан таап, көчүрүп жазгыла.

Текст боюнча жогорку тапшырмалардын бардыгы аткарылып бүткөндөн кийин “Кыргызча саламдашуу этикетинин өзгөчөлүктөрү эмнеде?” деген темада талкуу уюштурулуп, жыйынтыктары чыгарылат. Андан соң 2-3 студент текстти кайрадан үн чыгарып окушат. Башка тилдүү студенттер кыргызча тексттерди шар, уккулуктуу жана ***ү,ө,ң*** тыбыштарын туура айтып окууга да көнүгүүлөрүнө дайыма көңүл бурулууга тийиш.

Кыргызча сүйлөшүү этикетинде колдонулуучу сөз, сөз айкашы жана сүйлөмдөр боюнча студенттердин жекече сөздүк-минимумун түздүрүү да жакшы натыйжа берет.

Кыргызча саламдашуу, көрүшүү, учурашуу жөрөлгөлөрү, аларда колдонулуучу тил каражаттары тууралуу студенттер толук маалыматтарды алып, өз сүйлөөлөрүндө активдүү колдоно башташкандан кийин, бул теманы кеңейтип ар түрдүү улуттардагы саламдашуу этикетин, жакшы сөздөр боюнча, маалыматтарын берүү үчүн эки тилде жазылган тексттерди пайдалануу сунуш кылса болот.

Колдонулган адабияттар

1. Чыманов Ж.А. Байланыштуу кепти окутуунун негиздери [Текст] / Ж.А.Чыманов – Б.: Эркин-Тоо, 1997. – 100 б.

2. Култаева Ү.Б. Кыргыз тилин бөтөн тил катары окутуунун методикасы Текст] / Ү.Б.Култаева.– Б.: 2003. – 226 б.

3. Рысбаев С.К. Окуу орус тилинде жүргүзүлгөн мектептерде кыргыз тилин окутуу маселелери [Текст]/ С.К.Рысбаев.– Б.: 2007. – 131 б.

УДК 334.784(1.712.1/.2)

ИНТЕГРАЦИЯ СТРАН СНГ: ЭВОЛЮЦИЯ И ПРОБЛЕМЫ

Коджомуратова Росия Назарбековна, *к.э.н., доцент, КГТУ им. И. Раззакова,Кыргызстан, 720044, г.Бишкек, пр. Мира 66, e-mail:mika_1979@mail.ru*

Асанакунова Гулжан Букарбаевна, *доцент, КГТУ им. И. Раззакова, Кыргызстан, 720044, г.Бишкек, пр. Мира 66, e-mail: gasanakunova@mail.ru*

Развитие интеграционных процессов является важнейшей характеристикой современного мирового хозяйства. Процессы международной экономической интеграции заметно активизировались во второй половине XX в. в различных регионах земного шара, в частности, на постсоветском пространстве.

Ключевые слова: Интеграция, интернационализация, глобализация, международное разделение труда, зона свободной торговли, таможенный союз, мировой рынок, ЦАЭС, ВТО, Единое экономическое пространство.

THE INTEGRATION OF THE CIS COUNTRIES: EVOLUTION AND PROBLEMS

Kodzhomuratova Rosiya Nazarbekovna, *PhD, Associate Professor, Kyrgyzstan, 720044, c.Bishkek, KSTU named after I.Razzakov e-mail: mika_1979@mail.ru*

Asanakunova Gulzhan Bukarbaevna, *Associate Professor, Kyrgyzstan, 720044, c.Bishkek, KSTU named after I.Razzakov e-mail: gasanakunova@mail.ru*

The development of integration processes is the most important characteristic of the modern world economy. The process of international economic integration has intensified in the second half of the twentieth century in different regions of the globe, particularly in the post-Soviet space.

Keywords: Integration, internationalization, globalization, international division of labor, free trade area, customs union, global market, CAPS, the WTO, the Common Economic Space.

На пороге XXIв. возник так называемый «новый регионализм». Резкое усиление межфирменной и межгосударственной конкурентной борьбы, новые сферы конкуренции и более жесткое соперничество на традиционных рынках обуславливает необходимость кооперации как материально-финансовых, так и производственных усилий территориально сопряженных стран, позволяя укрепить свои позиции в глобализирующейся экономике, использовать потенциал крупного экономического пространства, наконец, выступать единой силой против общих конкурентов на мировом рынке. В результате имеет место не просто определенная увязка национально – государственных интересов, но и их возвышение до уровня региональных интересов. Таким образом, процессы глобализации в мировом хозяйстве сопровождается регионализацией- хозяйственным сближением стран на региональной основе, принимающей форму экономической интеграции (economic integration)

Основоположники экономической науки (А. Смит, Д. Риккардо, К. Маркс) и их последователи (современные ученые-экономисты) выводили международную торговлю, мирохозяйственные связи, международные экономические отношение, а вместе с тем и международную экономическую интеграцию из разделения труда в обществе между странами и народами. Концентрация труда и других ресурсов на изготовлении определенных видов продукции для продажи на внешнем рынке и ввоз необходимых при этом товаров предполагают зависимую от спроса на международном рынке специализацию производства. Это означает то или иное объединение усилий в удовлетворении потребностей отдельных стран, создание условий для увеличения количества и расширении набора товаров и услуг за счет их импорта, углубление международного разделения труда, количественное и качественное развитие мирохозяйственных связей, обеспеченных экономическими интересами их участников.

Международное разделение труда, порождаемая им внешняя торговля, а также более высокие формы мирохозяйственных связей, позволяют получить существенный экономический выигрыш, сократить затраты на производство продукции. Согласно классической экономической теории, это связанно с абсолютными (по А. Смиту) и относительными (по Д. Риккардо) преимуществами , возникающими в результате разделения труда между странами.

В современных условиях технологические и конкурентные факторы формируют преимущества в мирохозяйственных связях, способствуют увеличению их масштабов, диверсификации, переходу к более высоким формам интеграции. Всё большее значение приобретают близость экономических уровней и структур, технологическое подобие стран партнеров.

Предпосылки для более тесных хозяйственных связей интеграционного характера создаются между странами и группами стран в отдельных регионах мира. Дополнительные маркетинговые преимущества возникают в связи с модификацией жизненного цикла продукции и отдельных его этапов, которая вызвана участием в международной экономической интеграции и включением во внешний рынок. И в этом случае близость структур, потребительских склонностей и предпочтений, технологическое подобие играют

важную роль. С теоретической точки зрения вся совокупность названных факторов продвигает мирохозяйственные связи к интеграционному типу развития.

В процессе интеграции создаются новые товарные потоки между странами – членами интеграционной группировки, которые устраняют производство более дорогих аналогичных товаров внутри данной страны, затем товары, изготовляемые в интегрирующихся странах, постепенно замещают импорт соответствующих товаров из третьих стран.

Таким образом, «чистым результатом» новых товарных потоков в рамках интеграции является рост производства и, следовательно, благосостояния в странах-членах, возрастает уровень международной специализации. Все это способствует повышению эффективности производства в целом и в каждой стране.

Создание интеграционной системы позволяет участниками поставить общую цель и совместно ее достигнуть (рост производства и занятости, социальная стабильность и т.п.). В этом случае явный упор делается на увеличении значения государства в решении задач экономической интеграции, когда именно его усилиями создается общий рынок, принимаются оптимальные меры, обеспечивающие производство товаров и услуг.

Страны стремятся к интеграции для определения «фактора ограниченности» (имеются в виду природные ресурсы, сырье, энергоносители и т.д.). Тем самым возникают условия для увеличения производства, лучшего использования статических (размеры предприятий) и динамических (учиться производить) факторов, обеспечивающих рост емкости рынка, лучшую организацию производства. Это обычно иллюстрируется данными роста торговли между странами, в том числе внутрифирменной торговли, широкой диверсификации товаров одинакового назначения по внешнему виду, дизайну, качественными характеристикам. В других случаях акцент делается на особое значение технологических факторов экономической интеграции. Возрастание их значение на современном этапе заставляет страны резко увеличивать затраты на научно-исследовательские и опытно – конструкторские работы (НИОКР). Уменьшить их долю позволяет объединение усилий и ресурсов в данной сфере путем интеграции. Еще одним важным преимуществом МЭН является обострение конкуренции - эффективный стимул повышения качества и обновление ассортимента товаров и услуг.

Процесс интеграции обычно начинается с либерализации взаимной торговли, устранения ограничений в движении товаров, услуг, капиталов и постепенно при соответствующих условиях и заинтересованности стран-партнеров ведет к единому экономическому, правовому, информационному пространству в рамках региона. Формируется новое качество международных экономических отношений.

Современное развитие мирохозяйственных связей характеризуется интенсификацией международных экономических отношений как на двух- и многосторонней основе, так и на региональном и межрегиональном уровне.

Актуальность теоретического осмысления мирового опята экономической интеграции обуславливается тем, что еще многие развивающиеся страны не нашли оптимального пути развития.

Стремление к интеграции, тесному взаимовыгодному сотрудничеству отмечается и среди арабских государств Персидского залива. С 1981 г. создан и функционирует Совет по сотрудничеству ряда арабских государств, включающий Саудовскую Аравию, Кувейт, Катар, Бахрейн, Объединенные Арабские Эмираты и Оман («Нефтяная шестерка»). В 1992 г. было объявлено о создании Организации экономического сотрудничества центральноазиатских государств (ОЭС-ЭКО), которая, по замыслу учредителей, должна стать прообразом будущего Центрально-азиатского общего рынка, который должен включать в мусульманские республики СНГ- среднеазиатские, Казахстан, Азербайджан.

Тому примеры, попытки и проблемы формирования центрально-азиатского союза, Евразийского союза, Содружества независимых государств. Объединение национальных

интересов стран СНГ для более эффективного участия на мировом рынке требуют создания интеграционных объединений, в рамках которых взаимодействие самостоятельных экономических субъектов дают им реальные экономические выгоды.

СНГ в настоящее время не имеет оптимального механизма регионального сотрудничества и это ведет к нерациональному использованию природных и экономических ресурсов, а совокупность страновых территорий не используется как единый рынок, отсутствует разделение труда, кооперация.

Страны СНГ обладают большим природным и экономическим потенциалом, который даст им значительные конкурентные преимущества и позволяет занять достойное место в международном разделении труда. Они располагают 16,4% мировой территории, 5% численности населения, 20% мировых запасов нефти, 40% природного газа, 25% угля, 10% производства электроэнергии, 25% мировых запасов леса, почти 11% мировых возобновляемых водных ресурсов,13% пахотных земель, транспортно коммуникационные системы государств – участников СНГ играют все большую роль в мировых транспортных связях.

Только разведанные месторождения нефти в России составляют 13% мировых, в Азербайджане – более 10, в Казахстане и Туркменистане- около 10. В России сосредоточено около 3,5% мировых запасов природного газа, в Азербайджане, Туркменистане, Казахстане, и Узбекистане – 20%. По суммарной добыче каменного и бурого угля Россия, Украина и Казахстан занимают второе место в мире. Основные запасы алмазов, бокситов, медных, никелевых, кобальтовых и оловянных руд расположены на Украине и в Казахстане. В России и Белоруссии находятся крупнейшие в мире зеленые массивы (1/4 лесов земного шара) и запасы калийных солей.

Сегодня СНГ-это геополитическая реальность, играющая важную роль в обеспечении стабильности и безопасности на евразийском пространстве. Эффективно объединяя усилия и конкурентные преимущества, а также развивая мировой опят интеграции страны региона в состоянии достичь желаемых результатов.

Важнейшим результатом сотрудничества стран СНГ в области экономической интеграции является формирования Евразийского экономического сообщества (ЕврАЗЭС). Это- наиболее удачное и реально работающее интеграционное объединение на постсоветском пространстве: -в октябре 2000 года в Астане президентами Белоруссии, Казахстана, Киргизии, России и Таджикистана был подписан Договор об учреждении Евразийского экономического сообщества (ЕврАзЭС). Изначально ЕврАзЭС рассматривалось как основа для экономической интеграции постсоветских республик, формирования Таможенного союза и Единого экономического пространства, а также попытка усиления военно-политической интеграции (в рамках ОДКБ).

Реальная возможность форсированного укрепления и реализации экономического и военно-стратегического потенциала постсоветского пространства столкнувшегося с глобальными вызовами, может быть реализована только коллективными методами, т.е. через реальную экономическую и военно-политическую интеграцию. Хотелось бы отметить, что именно к 2010 году на постсоветском пространстве сформировались ряд предпосылок и факторов, способствующих реальной интеграции. Однако отсутствие полноценной интеграционной стратегии препятствует успешной реализации на постсоветском пространстве потенциала для модернизации экономики и обеспечения национальной безопасности СНГ.

В связи с этим, 5 октября 2007 г. решением Совета глав государств Содружества принята Концепция дальнейшего развития СНГ и План по ее реализации, которые имеют стратегическое значение и призваны адаптировать Содружество к современным условиям, повысить результативность его деятельности. При этом Концепция не только закрепляет сложившиеся формы сотрудничества, но и определяет новые ориентиры.

Так, в Концепции отмечается, что страны будут взаимодействовать в соответствии со своими реальными практическими потребностями и эффективно обеспечивать национальные интересы каждого государства.

Ограниченное участие отдельных государств участников СНГ в деятельности его органов и принимаемых документах, обусловленное особенностями национальных интересов, воспринимается в Содружестве с пониманием и уважением. Наряду с этим в СНГ реализуются разноуровневые и разноформатные модели взаимодействия, учитывающие специфику национальных интересов и внешнеполитического курса государств – членов СНГ.

В контексте рассматриваемой проблемы определение приоритетных отраслей развития экономики Кыргызской Республики до настоящего времени остается нерешенной. В разные годы приоритетными считались то электроэнергетика, что проявилось в принятии Программы перевода на электроснабжение (1991-1996 гг.), чтобы заменить потребление газа на электроэнергию собственного производства, то туризм, то стремление превратить Кыргызстан в транзитное государство согласно доктрине Великого Шелкового пути. Но ни электроэнергетика, ни туризм, ни транспорт не стали ведущими отраслями экономики страны. Хотя и неоднократно эта проблема поднималась в различных программных документах, в частности, в стратегии Развития КР на 2007-2010гг.

Приоритетом СНГ сегодня является экономическое сотрудничество , предусматривающее комплексные действия в данной сфере и выработку Стратегии экономического развития СНГ с выделением приоритетов, направленных на содействие развитию национальных экономик.

Наконец, 14 ноября 2008 г. решением Совета глав правительств СНГ была принята Стратегия экономического развития СНГ на период до 2020 года. И она призвана придать дополнительные импульсы разностороннему взаимодействию в интересах экономического роста стран Содружества.

Так в Стратегии развития отмечается, что разработка и обоснование долгосрочного экономического сотрудничества выполненно на основе развития взаимовыгодных торговых отношений и осуществления совместных инвестиционно-инновационных проектов и программ за счет наиболее полного использования совокупного потенциала с учетом международного опыта.

Общая стратегическая цель стран СНГ – это поиск ключевых внутренних источников развития национальных экономик для реализации преимуществ межгосударственного разделения труда, специализации и кооперирования производства, и взаимовыгодной торговли.

Этапы реализации Стратегии:

-первый этап- 2009-2011гг. предусматривает завершение создания зоны свободной торговли с нормами и правилами ВТО;

-второй этап- 2012-2015гг. – формирования межгосударственного инновационного пространства на основе межгосударственных отраслевых научно-технических программ;

-третий этап – 2016-2020гг. – формирование регионального рынка нано- и пикоиндустрии в целях сохранения и развития наукоемких отраслей экономики.

Осенью 2010г. утверждена Концепция дальнейшей экономической интеграции СНГ, дополняющая Стратегию экономического развития СНГ на период до 2020 года. Документ определяет цели, принципы направления и этапы формирования общего (интегрированного) экономического пространства СНГ с учетом современных политических и экономических реалий и в увязке с прогрессом в создании Таможенного союза (ТС) и Единого экономического пространства (ЕЭП) России, Белоруссии, Казахстана и Кыргызстана.

Главная цель Концепции – дать новый импульс интеграционным прогрессам в СНГ на долгосрочную перспективу через создание Многосторонней зоны свободной торговли в

течение 2011г., затем общего (интегрированного) экономического пространства к 2015г., что в перспективе должно привести к формированию широкой евразийской зоны экономического сотрудничества (или Большого евразийского экономического пространства) с участием СНГ и ЕС и с выходом на АТР.

Таким образом, за минувшие 25 лет интеграция на постсоветском пространстве обеспечила единство целей и средств, регулирующих поступательное развитие стран СНГ опирающихся на торговые и производственные связи и доказывает необходимость дальнейшего развития процесса высокими темпами.

Изучение опыта функционирования интеграционных пространств показывает, что набирающие силу процессы интеграции по-разному идут в различных регионах, отличаются по моделям, глубине и масштабам, а это ставит как перед политиками, так и экономистами практическую задачу не только признания целесообразности интеграции в экономическом взаимодействии стран, но и поиска форм их движения и методов разрешения.

Изложенные выше обстоятельства предопределяют необходимость извлечения уроков из опыта формирования интеграционных союзов в мире, анализа факторов, способствующих усилению интеграционных процессов, оценки моделей экономического взаимодействия стран, входящих в тот или иной интеграционный блок, а также управления интеграционными процессами в мире и т.д.

Список литературы

1. В.В. Путин. Новый интеграционный проект для Евразии - будещее, которое рождается сегодня/Известия от 4 октября 2011г.

2. Декларация о Евразийской экономической интеграции/Официальный сайт президента Российской Федерации//http://newskremlin.ru/ref notes/ 109) 3. Авдокушкин О.Ф. Международные Экономические отношения. – М. 1999.

4. Концепция дальнейшего развития Содружества Независимых государств. Испольный комитет Содружества Независимых Государств. Душанбе, 5 октября 2007г.

5. Стратегия Экономического развития Содружества Независимых Государств на период до 2020 года- Исполнительный комитет Содружества независимых государств. Кишинев, 14 ноября 2008г.

6. Декларация глав государств-участников Содружества Независимых Государств о дальнейшем развитии всестороннего сотрудничества Решение Совета глав государств СНГ, Ашхабад. 5 декабря 2012 г.

7. Кыргызская Республика в Евразийском экономическом союзе. – Бишкек, 2016

ИЗВЕСТИЯ

КЫРГЫЗСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ им. И. РАЗЗАКОВА

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

2016

№ 2 (38)

JOURNAL

OF KYRGYZ STATE TECHNICAL UNIVERSITY NAMED AFTER I.RAZZAKOV

THEORETICAL AND APPLIED SCIENTIFIC TECHNICAL JOURNAL

2016 № 2 (38)

Ответственный за выпуск	**Курманалиев Б.К.**
Редакторы языковой редакции	Дмитриенко К.М.,Турдукулова А.К. Эркинбек к. Ж.
Корректор	
Технический редактор и компьютерная верстка	Турдукулова А.К., Эркинбек к. Ж.

Подписано к печати 15.08.2016. Формат бумаги $70_{x100}{}^{1}/_{16}$ Бумага офс. .
Печать офс. Объем 13,375 п.л. Тираж 200 экз. Заказ 252.
Издательский центр “Текник”
Кыргызского государственного технического университета им. И.Раззакова
720044, Бишкек, ул. Сухомлинова, 20.
Тел.: 54-29-43, e-mail: beknur@mail.ru