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» – «үөнер».
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 «г», «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 «йуу», «йұу»,, depending on the hardness or softness of the other vowels in syllables. For example: «қою» – «қойұу», «түю» - «түйуу»;;
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=үө;

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.

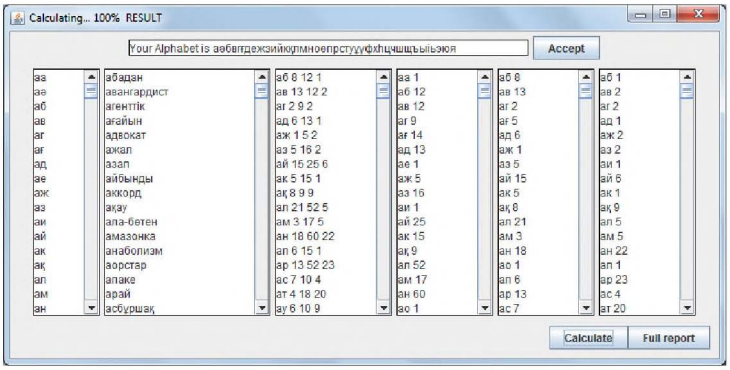

Fig. 2. Text analyzer's window

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.

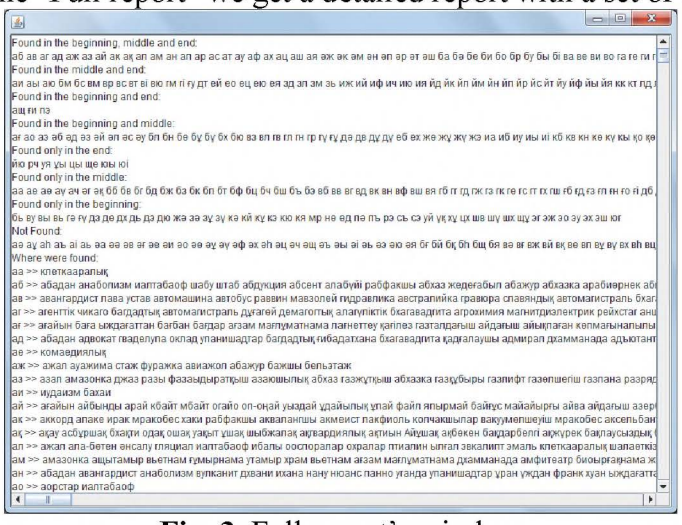

Fig. 3. Full report's window

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 рү 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 ік ім іп іт іу ьб ьд ье ьз ьк ьм ьн ьс ьт ьф ья юд юз юй юс ют юф яж яз яй як ял ям ят яу

Found at the beginning and at the end:
ащ ғи пэ

Found at the beginning and in the middle:
ағ 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 кү 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 ау 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 ек 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 ик 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 қб қг қд қе қж қк ққ қл қм қн қп қр қс қт қф 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 уа үә уе үф үх ұы 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 ыц 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 кұ 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