ALGORITHMS FOR ACCENTUATION AND PHONEMIC TRANSCRIPTION OF RUSSIAN TEXTS IN SPEECH RECOGNITION SYSTEMS

Yakovenko O.S. (olya.yakovenko@bk.ru)[1], Bondarenko I.Yu. (iv.bondarenko@2gis.ru)[2], Borovikova M.N. (m.borovikova@g.nsu.ru)[1], Vodolazsky D.I. (daniil.vodolazsky@mail.ru)[1]

[1]Novosibirsk State University, Novosibirsk, Russia
[2]2GIS, Novosibirsk, Russia

This paper presents an overview of rule-based system for automatic accentuation and phonemic transcription of Russian texts for speech connected tasks, such as Automatic Speech Recognition (ASR). Two parts of the developed system, accentuation and transcription, use different approaches to achieve correct phonemic representations of input phrases. Accentuation is based on “Grammatical dictionary of the Russian language” of A.A. Zaliznyak and wiktionary corpus. To distinguish homographs, the accentuation system also utilises morphological information of the sentences based on Recurrent Neural Networks (RNN). Transcription algorithms apply the rules presented in the monograph of B.M. Lobanov and L.I. Tsirulnik “Computer Synthesis and Voice Cloning”. The rules described in the present paper are implemented in an open-source module, which can be of use to any scientific study connected to ASR or Speech To Text (STT) tasks. Resulting system has shown 98.3% phone accuracy on a test set of 63 sentences (and 200 phonetic syntagms) which were marked up manually by linguists. The developed toolkit is written in the Python language and is accessible on Github for any researcher interested.

Keywords: automatic speech recognition, corpora, accentuation, rule-based phonemic transcription

РАЗРАБОТКА АЛГОРИТМОВ ДЛЯ АВТОМАТИЧЕСКОЙ РАССТАНОВКИ УДАРЕНИЙ И ФОНЕМАТИЧЕСКОГО ТРАНСКРИБИРОВАНИЯ ТЕКСТОВ НА РУССКОМ ЯЗЫКЕ В СИСТЕМАХ РАСПОЗНАВАНИЯ УСТНОЙ РЕЧИ

Яковенко О.С. (olya.yakovenko@bk.ru)[1], Бондаренко И.Ю. (iv.bondarenko@2gis.ru)[2], Боровикова М.Н. (m.borovikova@g.nsu.ru)[1], Водолазский Д.И. (daniil.vodolazsky@mail.ru)[1]

[1]Новосибирский государственный университет, Новосибирск, Россия
[2]2ГИС, Новосибирск, Россия

Ключевые слова: автоматическое распознавание речи, корпус, расстановка ударений, фонематическое транскрибирование по правилам

1. Introduction

The main difficulty in Natural Language Processing (NLP) tasks is the irregularity of natural language in comparison to artificial language. This irregularity could be met at all levels, including the phonetic level. Moreover, the ambiguity of the phonetic representations in Russian written texts is a lot stronger than that in many other alphabetical languages. This is why, building an automatic speech synthesis or recognition system, it is important to introduce the most accurate phonemic representations of words and phrases.

Complex tasks such as building an acoustic model for ASR, apart from an effective inner architecture (Hidden Markov Models, Neural Networks), require first of all a big corpus with quality mark-up. As mentioned earlier, this quality can be achieved by correct phonemic representation of the spoken data which would serve as a foundation training set for an acoustic model. Modern open-source tools for automatic text accentuation and transcription are based mostly on letter representations of words and too little on possible phonetic features. Our approach is based on real phonetic process modelling in the spoken Russian language.

The goal of this paper is to introduce an approach for quick automatic accentuation and transcription of big texts based on rules, thesaurus and morphological information. We believe our research will be helpful for ASR or TTS projects which do not have the resources to do manual mark-up or build an enormous corpus but still require good results in recognition or synthesis of Russian language.

Although the problem of automatic text transcription is already thoroughly studied, with numerous complex algorithms introduced on a theoretical level, there are no quality grapheme-to-phoneme systems available in open-source. Thus, our goal is not solely to develop a system of automatic accentuation and transcription based on the research of Lobanov and Tsirulnik, but more importantly to fully develop the following system, reconsider the rules to correctly fit the realities of spoken Russian language and implement reconsidered algorithms in an open-source module, ready to be used by any researcher interested.

2. Related works

Naturally, the investigations of the connection between phonemic and graphicalrepresentations of words, as of phonetic processes in general have a long history. Among the fundamental works dedicated to Russian phonetics there are studies of Avanesov (e.g. [1]), Vorontsova (et al. [2]), Zaliznyak [12]. The following works are used as reference to most automatised grapheme-to-phoneme systems, regarding both presented rules and examples.

One of the first studies, presenting a rule-based approach towards automatic transcription of Russian texts was authored by Krivnova et al. and published in 1970 [13]. Later, Krivnova et al. were working on a rule-based grapheme-to-phoneme system for TTS introduced in [7], which developes the idea of different levels of abstraction and discrimination of different allophones for specifying different types of sounds. Other rule-based transcription system is presented in [6], this paper features distinction of alternative transcriptions for words and phrases. The algorithm is implemented in a module written in C++ language.

Besides rule-based approaches, nowadays there is a scientific tendency for statistical analysis and Machine Learning (ML) to be used in NLP tasks including ASR and TTS. Such methods imply stress or phoneme prediction from the graphical representation of words [3,4,9]. Unfortunately, in application to the following NLP tasks, ML is most effective with big marked up data which researchers rarely have.

Our rule-based system of transcription is generally relied on Lobanov and Tsirulnik [8]. The main focus of this study is the modelling of phonetic processes in Russian language using general computerized assumptions about grapheme to phoneme mapping and basic phonetic processes between groups of phonemes (e.g. assimilation, reduction). The accentuation is founded on the electronic thesaurus[1] compiled from the grammatical dictionary of Zaliznyak [12].

The system presented in this paper is aimed first of all at such open-source products for ASR as Kaldi [10], CMU Sphinx [5] and HTK [11]. Those toolkits simplify and automatize the process of custom acoustic model production, based on both Hidden Markov Models and Neural Networks. Finally, they also feature a decoder for spoken data which uses pre-trained acoustic model to map the acoustic signals to phonemes.

3. Accentuation

To start with, the complexity of accentuation for the Russian language is rather high in comparison to some popular alphabetic languages (English, French). This is connected with a linguistic feature which was named “dynamic stress”. It means that a certain graphical representation of a word does not have a fixed stress: e.g. *зá ногу* or *за нóгу*, *óкна* or *окнá*. This phenomenon illustrated by the examples (when the same graphical representation of the word might represent different pronunciations) is named homography. Thus, homography becomes the key issue behind automatic accentuation. The system described in the present paper has been developed in accordance with homograph resolution techniques based on word morphology.

As mentioned earlier, the developed system consists of two consecutive parts, the accentuation and the transcription. In this chapter the accentuation module is presented.

3.1 Offline corpus

The stress information of the most frequent Russian words has been collected from the electronic resource based on grammatical word analysis presented in [12]. The offline dictionary for the accentuation module was built with the help of the following resource. It includes a simple list of words with stress (which have an unambiguous stress variant) and a more complex data structure for homographs. Homographs in the dictionary are listed without stress and have different morphological information and the corresponding graphical representations with accent for the listed morphological forms. Morphological information is displayed in the Universal Dependencies[2] format and was generated initially by a Python module[3] based on RNN approach. The stress is marked as a “+” symbol after the stressed vowel; below in Table 1 is the sample from the dictionary described.

Table 1. Sample from the dictionary of accents

Homographs	Non-homographs
..., "альплагеря": { *"NOUN Animacy=Inan|Case=Gen|Gender=Masc|Number=Sing": "альпла+геря",* *"NOUN Animacy=Inan|Case=Nom|Gender=Masc|Number=Plur": "альплагеря+"*	*...* *"проистека+л",* *"проистека+ла",*

1 http://www.speakrus.ru/dict/#paradigma
2 http://universaldependencies.org/
3 https://github.com/IlyaGusev/rnnmorph

}, ...	*"проистека+ли",* *"проистека+ло",* ...

3.2 Architecture

The system waits for a tokenized sentence as input and also can receive morphological tags in the Universal Dependencies format, which could be generated by a Python module mentioned earlier for better homograph disambiguation. Each word of the phrase is processed separately; the steps for processing are the following:

I. Look up in the non-homonyms list;
II. Look up in the homonym list, choosing the exact form that corresponds to the word and the tag;

If the match could not be found in the offline corpus:

III. Search the exact word in wiktionary;
IV. Save the result for future extraction and for rewriting the main offline dictionary.

The system can also handle non-Russian words and symbols, anticipatorily accented words, function words and hyphenated compounds. Examples are in Table 2 below:

Table 2. Examples of automatic accentuation

Input phrase	Morphotags	Result
['подарок', 'для', 'кума']	*['NOUN Animacy=Inan|Case=Nom|Gender=Masc|Number=Sing', 'ADP _', 'NOUN Animacy=Anim|Case=Gen|Gender=Masc|Number=Sing']*	*['пода+рок', 'для', 'ку+ма']*
['фёдор', 'любит', 'кофе']	-	*['фё+дор', 'лю+бит', 'ко+фе']*

The quality of automatic phonemic transcription for ASR and TTS tasks depends highly on the stress placement in the word. This is the reason why automatic accentuation is an important part of the transcription.

4. Transcription

A phoneme is one of the units of sound that distinguish one word from another in a particular language. Briefly speaking, if different words can be distinguished by a substitution of a phone, then that phone is a phoneme (e.g. *free* /fɹiː/ and *three* /θɹiː/). Like in many alphabetical languages with a big history, Russian graphics does not represent fully the phoneme composition of the word; for that cause simple "grapheme to acoustic signal" mapping shows good results only in the cases of huge data amounts. Smaller data require quality mark-up, and it could be achieved with rule-based transcription. As mentioned earlier, the transcription algorithm is based on Lobanov and Tsirulnik rules of phonetic transformations (pp. 35-37, 148 in [8]). Those rules correctly illustrate the correlation between graphemes and phonemes in the modern Russian language and model some of the most frequent phonetic transformations in the Russian spoken discourse.

4.1 Phoneme set

There are several approaches to Russian system phonemes' definition. The most common are the following. Moscow phonological school (MPS) considers that there are 42 phonemes: 6 vowels and 36 consonants. In Leningrad phonological school (LPS) phoneme Ы isn't taken into account. We used the concept of phoneme as such from LPS but we are mindful for the Ы according to MPS. This combination was chosen due to the necessity of specifying most probable and significant acoustic events, since the algorithm is first of all aimed at creating transcriptions for manual speech recognition system training. On the one hand, it is important that this system recognize the Ы phoneme but on the other hand, the LPS's understanding of the phoneme definition is more efficient for such purpose given that there are many errors connected with aphetic phones. For ASR and even more for TTS it is important to distinguish stressed and non-stressed vowels in the word, since the spectral characteristics of the vowels differ greatly depending on stress. Therefore, a new set of "phonemes" has been designed to better illustrate the differences between stressed and non-stressed allophones of a phoneme (stressed are marked by zeroes "0") and allophones which appear only in a certain phonetic context.

One other designation that the following system operates with is the "sil" quaziphoneme that represents the absence of any speech data in the transcription: pauses, endings or beginnings of the recording. For exact "phoneme" designation used see Tables 3 and 4 below.

Table 3. Set of vowels that are used in automatic transcription of Russian texts.

«Phoneme» designation		**Traditional designation**	**Vowel description**
U0	U	*/u/ (/y/)*	Close back/front rounded vowel (п**у́**ля, п**у́**ть, **у**ро́дина, **у**кло́н)
O0		*/o/ (/o/)*	Close-mid back rounded vowel (п**о́**ле, парк**о́**вка)
A0	A	*/a/*	Open front unrounded vowel (трав**а́**, вод**а́**, тр**а**ва́, п**а**рко́вк**а**)
E0		*/ẹ/ (/э/)*	Close-mid/Open-mid front unrounded vowel (ф**е́**рма, м**е́**ра)
Y0	Y	*/ɨ/*	Close central unrounded vowel (с**ы́**р, кор**ы́**то, м**ы**ча́ние, пол**ы**ха́ть)
I0	I	*/i/*	Close front unrounded vowel (п**и́**во, п**и́**цца, в**и**но́, п**и**ла́)

Table 4. Set of consonants that are used in automatic transcription of Russian texts.

«Phoneme» designation		**Traditional designation**		**Consonant description**
K	K0	*/k/*	*/kʲ/*	Voiceless (palatalized) velar stop (**к**итель, Лу**к**ьяненко, **к**у**к**ла, **К**удрявцев)
KH	KH0	*/x/*	*/xʲ/*	Voiceless velar fricative (**х**ек, **х**итрость, бо**г**, у**х**о)
GH	GH0	*/ɣ/*	*/ɣj/*	Voiced velar fricative (мо**х** горит)
G	G0	*/g/*	*/gʲ/*	Voiced velar stop (**г**ен, **г**ибель, у**г**он, **г**азон)
J0		*/j/*		Palatal approximant (**й**огурт, **е**нот)
	TSH0		*/tɕ/*	Voiceless alveolo-palatal affricate (**ч**ерника, **ч**ижик)
SH	SH0	*/ʂ/*	*/ɕː/*	Voiceless retroflex fricative (**щ**авель, мо**щ**ь, ма**ш**ина, ме**ш**ок)
ZH	ZH0	*/ʐ/*	*/ʑː/*	Voiced retroflex fricative (**ж**ена, па**ж**ить)
DZ	DZ0	*[ʣ]*	*[ʣj]*	Voiced alveolar affricate (прин**ц** гуляет, Го**дз**илла)
DZH	DZH0	*[ʥh]*	*[ʥhj]*	Voiced alveolo-palatal affricate (до**ч**ь гуляет, Фи**дж**и)
R	R0	*/r/*	*/rʲ/*	Alveolar and postalveolar trills (**р**инг, т**р**юк, по**р**а, пи**р**синг)
T	T0	*/t/*	*/tʲ/*	Voiceless dental stop (**т**ир, у**т**ёнок, **Т**аня, **т**ор**т**)
TS	TS0	*/ʦ/*	*/ʦj/*	Voiceless alveolar affricate (**ц**арь, во**т** **С**ергей)
S	S0	*/s/*	*/sʲ/*	Voiceless alveolar sibilant (**с**елёдка, би**с**ер, **с**ахар, моро**з**)

D	D0	*/d/*	*/dʲ/*	Voiced dental stop (**Д**има, ка**д**ило, **д**рака, вы**д**ача)
Z	Z0	*/z/*	*/zʲ/*	Voiced alveolar fricative (**з**игзаг, о**з**емь, зиг**з**аг, **з**она)
N	N0	*/n/*	*/nʲ/*	Alveolar nasal (**н**юанс, ко**н**и, нюа**н**с, **н**ос)
L	L0	*/ɫ/*	*/lʲ/*	Velarized alveolar lateral approximant (**л**ев, бо**л**ь, **л**ось, па**л**атка)
P	P0	*/p/*	*/pʲ/*	Voiceless bilabial stop (Ко**п**ерник, **п**иво, **п**ривет, **п**арень)
F	F0	*/f/*	*/fʲ/*	Voiceless labiodental fricative (**ф**ея, по**ф**иг, **ф**лирт, ко**ф**р)
B	B0	*/b/*	*/bʲ/*	Voiced bilabial stop (ря**б**ина, при**б**ой, **б**аржа)
V	V0	*/v/*	*/vʲ/*	Voiced labiodental fricative (**в**ино, на**в**ет, **в**одка, па**в**одок)
M	M0	*/m/*	*/mʲ/*	Bilabial nasal (**м**ир, Пер**м**ь, **м**орж, **м**а**м**а)

4.2 Letter-phoneme transformation rules (TLP)

Rule TLP 0.

If the whole word is in exclusion dictionary, it is converted in a form that will be transformed correctly by following rules. The exclusion dictionary has the following format:

автоби+знесу автоби+знэсу

оттого+ оттово+

здра+вствуй здра+ствуй

This way the algorithm preserves the stictness of the rules, also making it easy to add new exclusions for different contributors.

4.2.1 Vowels transformation

In processing vowels (А, О, У, Э, Ы, И, Я, Ё, Ю, Е), the following transformation rules are used. All the rules are written for letter in the position *i*.

Rule TLP 1.

Letter У always transforms into U and is marked as stressed if necessary.

Rule TLP 2.

If *i-1* symbol is Ь, Ъ, a space, a punctuation mark, a 'silence' or a vowel:

1. Ё, Ю, А, Ы, И go into [J0 O], [J0 U], [A], [Y], [I] and are marked as stressed if necessary;
2. Stressed О, Э, Е, Я go into [O], [E], [J0 E], [J0 A] and are marked as stressed;
3. Unstressed О, Э, Е go into [A], [Y], [J0 I];
4. Unstressed Я goes into [J0 A] if it is at the end of the word, and into [J0 I] if it isn't.

Rule TLP 3.

If *i-1* letter is Й, Ч or Щ, and i+1 symbol is a letter:

1. Unstressed А, Е go into [I];

Other letters are transformed by rule TLP 5.

Rule TLP 4.

If *i-1* letter is Ц, Ж or Ш:

1. И goes into [Y] and is marked as stressed if necessary;
2. Unstressed Е and А go into [Y];
3. Stressed Е goes into [E] and is marked as stressed;

Other letters are transformed by rule TLP 5.

Rule TLP 5.

If *i* letter is in conditions different from above:

1. А, Ё, Ю, Е, И, Ы go into [A], [O], [U], [I], [I], [Y] and are marked as stressed if necessary;
2. Stressed О, Э, Я go into [O], [E], [A] and is marked as stressed;
3. Unstressed О, Э go into [A], [Y];

4. Unstressed Я goes into [A] if it is at the end of the word, and into [I] if it isn't.

Table 5. Examples of application of vowels transformations rules

Word original	пь**я**ный	да**ю**т	**е**ль	**ао**рта	м**о**л**о**к**о**
Phonetic transcription	P0 **J0** **A0** N Y J0	D A **J0** **U0** T	**J0** **E0** L0	**A** **O0** R T A	M **A** L **A** K **O0**

4.2.2 Consonants transformation

4.2.2.1 Transformation of unpaired consonants

In processing vowels (Й, М, Н, Р, Л) next transformation rules are used. All the rules are written for letter in the position *i*.

Rule TLP 6.

Й always transforms into [J0].

Rule TLP 7.

If *i+1* letter is Ь, Я, Ё, Ю, Е, or И, letters М, Н, Р, Л go to [M0], [N0], [R0], [L0].

Otherwise, М, Н, Р, Л go to [M], [N], [R], [L].

Rule TLP 8.

If i letter is Н and *i+1* phoneme is [N0], [D0], [T0] or [S0], letter Н goes to [N0].

Table 6. Examples of application of unpaired consonants transformations rules

Word original	ма**й**ка	ви**н**тик	**р**ека	**н**ебо	мо**л**ь
Phonetic transcription	M A0 **J0** K A	V0 I0 **N0** T0 I K	**R0** I K A0	**N0** E0 B A	M O0 **L0**

4.2.2 Transformation of paired consonants

In processing vowels (Б,П), (Д,Т), (Г,К), (В,Ф), (З,С), (Ж,Ш), и Х, Ц, Ч, Щ, the following transformation rules are used. All the rules are written for letter in the position *i*.

Rule TLP 9.

If *i+1* phoneme is [P], [P0], [T], [T0], [F], [F0], [TS], [S], [S0], [SH], [SH0], [KH], [KH0] or [TSH0]:

1. Б, П go to [P] and are marked as palatalized if necessary;
2. Д, Т go to [T];
3. В, Ф go to [F];
4. З, С go to [S];
5. Ж, Ш go to [SH];
6. Г, К go to [K];
7. Х, Ц, Ч go to [KH], [TS], [TSH0];
8. Ш, Ж go to [SH];
9. Щ goes to [SH0].

Rule TLP 10.

If *i+1* phoneme is [B], [D], [G], [Z] or [ZH]:

1. Б, П go to [B];
2. Д, Т go to [D];
3. В, Ф go to [V];
4. З, С go to [Z];
5. Ж, Ш go to [ZH];
6. Г, К go to [G];

7. Х goes to [GH];
8. Ц, Ч, Щ go to [DZ], [DZH0], [ZH0];
9. Ш, Ж go to [ZH].

Rule TLP 11.

If *i+1* letter is Ъ, А, О, У, Э, Ы, Л, М, Н, Р or В, all letters under consideration go to their corresponding phonemes.

Rule TLP 12.

If *i+1* letter is Я, Ё, Ю, Е or И, letters Б, П, Д, Т, Г, К, В, Ф, З, Х, С, go to their corresponding palatalized phonemes, and letters Ц, Ч, Щ, Ш, Ж go to their corresponding phonemes.

Rule TLP 13.

If *i+1* letter is Ь and *i+2* symbol is an unvoiced phoneme, a space, a punctuation mark, or a 'silence'-symbol:

1. Б, П go to [P0];
2. Д, Т go to [T0];
3. Г, К go to [K0];
4. В, Ф go to [F0];
5. З, С go to [S0];
6. Ж, Ш go to [SH];
7. Х goes to [KH0];
8. Ц, Ч, Щ go to [TS], [TSH0], [SH0].

Rule TLP 14.

If *i+1* letter is Ь, *i+2* letter is Я, Ё, Ю, Е И, Л, М, Н, Р or В, letters Б, П, Д,Т, Г, К, В, Ф, З, С, Ж, Ш, Х, Ц, Ч, Щ go to [B0], [P0], [D0], [T0], [G0], [K0], [V0], [F0], [Z0], [S0], [ZH], [SH], [KH0], [TS], [TSH0], [TSH0].

Rule TLP 15.

If *i+1* letter is Ь and *i+2* phoneme is [B], [D], [G], [Z], [ZH], [DZ], [DZH] or their palatalised case:

1. Б, П go to [B0];
2. Д, Т go to [D0];
3. Г, К go to [G0];
4. В, Ф go to [V0];
5. З, С go to [Z0];
6. Ж, Ш go to [ZH];
7. Х goes to [GH0];
8. Ц, Ч, Щ go to [DZ], [DZH0], [ZH0].

Rule TLP 16.

If i letter is Н, Т, С or Д, next phoneme is [M0], [P0], [B0], [V0], [F0], [N0], [T0], [D0], [Z0], [L0] or [S0], i letter goes to the corresponding phoneme.

Table 7. Examples of application of paired consonants transformations rules (Part 1)

Word original	лё**г**	по**д**нять	о**т**дых	**с**бор	**кл**емма
Phonetic transcription	L0 O0 **K**	P A **D0** N0 A0 T0	O0 **D** Y KH	**Z** B O0 R	**K** L0 E0 M A

Table 8. Examples of application of paired consonants transformations rules (Part 2)

Word original	ко**сь**ба	у**с**ни	пу**с**тяк	а**сс**истент	ко**сь**ба
Phonetic transcription	K A Z0 B A0	U **S0** N0 I0	P U **S0** T0 A0 K	A **S0** I S0 T0 E0 N T	K A Z0 B A0

4.2.3 Supplementary transformation rules

Rule TLP 17.

If *i* letter and *i+1* letter are equal and *i+2* letter is Ь, Я, Ё, Ю or И, both letters (i and i+1) go to their corresponding palatalized phonemes.

Rule TLP 18.

In combinations of letters «СТН», «СТЛ» and «НТГ» letter Т is removed.

Rule TLP 19.

In combinations of letters «ЗДН», «ЗДЦ», «НДЦ», «РДЦ», «НДШ» and «ГДТ» letter Д is removed.

Rule TLP 20.

In combinations of letters «ЛНЦ» letter Л is removed.

Rule TLP 21.

Combinations of letters «СЧ» и «ЖЧ» go to SH0, «ТС», «ТЬС», «ТЦ», «ДС» and «ДЦ» go to C, «СШ» go to SH, «ЗЖ» go to ZH.

Rule TLP 22.

If word ends by ОГО or ЕГО, ГО goes to VA, except «много», «дорого», «строго» and some other adverbs.

Rule TLP 23.

If *next* phoneme is [TSH0] or [SH0], letter Н goes to [N0].

Table 9. Examples of application of supplementary transformations rules (Part 1)

Word original	поз**дн**о	под у**здц**ы	ла**ндш**афт	со**лнц**е	**сч**астье
Phonetic transcription	P O0 **Z** **N** A	P A D U **S** **TS** Y0	L A **N** **SH** A0 F T	S O0 **N** **TS** Y	**SH0** A0 S0 T0 J0 I

Table 10. Examples of application of supplementary transformations rules (Part 2)

Word original	ра**сш**иб	въе**зж**ать	красн**ого**	о**тц**а	перебе**жч**ик
Phonetic transcription	R A **SH** Y0 P	V J0 I **ZH** A0 T0	K R A0 S N A **V** **A**	A TS A0	P0 I R0 I B0 E0 **SH0** I K

5. Evaluation

The following system was tested on an annotated corpus consisting of 63 sentences, from which 200 phonetic syntagms were generated depending on punctuation. An etalon mark up was verified by linguists and compared to the automatically generated markup of the developed system in derived terms from Word Error Rate (WER) and Word Accuracy: Phone Error Rate (PER) and Phone Accuracy. The result had the following structure:

SYNTAGM	**ETALON**	**RESULT**	**WER**
отыскал ждановскую набережную	A T Y S K A0 L ZH D A0 N A F S K U J0 U N A0 B0 I R0 I ZH N U J0 U	A T Y S K A0 L ZH D A0 N A F S K U J0 U N A0 B0 I R0 I ZH N U J0 U	0
разбил земную кору	R A Z B0 I0 L Z0 I M N U0 J0 U K A R U0	R A Z B0 I0 L Z0 I M N U0 J0 U K A R U0	0
освещённые окна сарая	A S V0 I SH0 O0 N Y I O0 K N A S A R A0 J0 A	A S V0 I SH0 O0 N Y J0 I A K N A S A R A0 J0 A	10.5

Test results were averaged on phonecit syntagms, metrics: PER 1.7%, Phone Accuracy 98.3% accordingly.

The same experiment was conducted with other open-source grapheme-to-phoneme algorithm[4], the following results were achieved: PER 3.3%, Phone Accuracy 96.7% accordingly.

5. Conclusion

The present study shows that the developed system is capable of fast and accurate accentuation as well as transcription of Russian texts. The accent dictionaries complemented by wiktionary parsing and morphological information cover many cases of graphical ambiguity. Generated transcriptions reflect many processes in Russian spoken discourse such as vowel reduction, assimilation and simplification of consonant groups. These grapheme-to-phoneme transformations do not only work within a single word but also at the junctions of words. This correctly reflects natural spoken language processes.

Naturally, there are several things to improve, for example:

1. Implementation of Wikipedia parsing into the accentuation stage that will lead to the expansion of the dictionary with proper names;
2. Possibility of handling the loanwords with non-stressed “ё” for both stages (and also Russian compounds with non-stressed “ё” such as *трёхколёсный*);
3. Flexibility enhancement for the transcription stage: ability to work with double-stressed words, filler instances other than “sil”; correctly handle punctuation;
4. Extension of the rule set towards more precise and complex cases;
5. Featuring of multiple accent variants and transcriptions, portraying different sociolects and dialects.

Finally, it is necessary to mention that this approach has been implemented in a module, written in Python language[5] and is ready for use by any researcher interested. We believe that the present study will boost interest towards open-source development of ASR and TTS systems and hope that it will be of use to the scientific community.

Acknowledgements

1. This paper was supported by Botan Investments[6].
2. Special thanks to Zamir Ostroukhov[7] for providing the script in open-source and being able to use the algorithm for evaluation.

References:

1. *R.I. Avanesov* (1984): “Literary pronunciation in Russian language”, Moscow, Prosvyashenie [“Russkoe literaturnoe proiznoshenie”].
2. *S.N. Borunova, V.L. Vorontsova, N.A. Eskova (1989): “Ortho-epic dictionary of the Russian language: Pronunciation, stress, grammatical forms: About 65 000 words”, Russian language [“Orfoepicheskiy slovar' russkogo yazyka: Proiznosheniye, udareniye, grammaticheskiye formy: Okolo 65 000 slov”, Russkiy yazyk].*
3. *A. Deri, K. Knight* (2016): “Grapheme-to-Phoneme Models for Almost Any Language”, Proc. ACL, available at: http://aclweb.org/anthology/P/P16/P16-1038.pdf.

4 https://github.com/zamiron/ru4sphinx/tree/master/text2dict
5 https://github.com/nsu-ai/russian_g2p
6 http://botaninvestments.com/ml-program/
7 https://github.com/zamiron

4. *K.B. Hall, R. Sproat* (2013): “Russian Stress Prediction using Maximum Entropy Ranking”, EMNLP, available at: https://static.googleusercontent.com/media/research.google.com/ru//pubs/archive/41392.pdf.
5. *D. Huggins-daines, M. Kumar, A. Chan, A.W. Black, M. Ravishankar , A. I. Rudnicky* (2006): “PocketSphinx: A free, real-time continuous speech recognition system for hand-held devices”, in Proceedings of ICASSP, available at: http://www.cs.cmu.edu/~awb/papers/ICASSP2006/0100185.pdf.
6. *I.S. Kipyatkova, A.A. Karpov* (2008): “The Phonemic Transcription Module for a System of Conversational Russian Speech Recognition”, Artificial Intelligence [“Modul’ fonematicheskogo transcribirovania dlya sistemy raspoznavaniya razgovornoi russkoi rechi”, Shtuchnii intellekt] available at: http://dspace.nbuv.gov.ua/bitstream/handle/123456789/7670/093-Kipyatkova.pdf.
7. O.F. Krivnova, L.M. Zakharov, G.S. Strokin (2001): “Automatic Transcriber of Russian Texts: Problems, Structure and Application”, Proc. of intern. workshop "Speech and computer".
8. *B.M. Lobanov, L.I. Tsirulnik* (2008): “Computer Synthesis and Voice Cloning”, Minsk, Belarusskaya Nauka [“Kompjuternyi Sintez i Klonirovanie Rechi”].
9. *F. Mana, P. Massimino, A. Pacchiotti* (2001): "Using machine learning techniques for grapheme to phoneme transcription", EUROSPEECH-2001, available at: https://pdfs.semanticscholar.org/ce0e/7ca7c745c2a65b6f3ac7be5df5a8e72065fe.pdf.
10. *Povey, Daniel, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlíček, Y. Qian, P. Schwarz, J. Silovský, G. Stemmer, K. Veselý* (2011): “The Kaldi Speech Recognition Toolkit”, IEEE Signal Processing Society, available at: https://homepages.inf.ed.ac.uk/aghoshal/pubs/asru11-kaldi.pdf.
11. *S. Young, G. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland* (2006): “The HTK Book”, Cambridge University Engineering Department.
12. *A.A. Zaliznyak* (1977): “Grammatical Dictionary of the Russian Language”, Moscow, Russia [Grammaticheskii Slovar’ Russkogo Iazyka. Russki Iazyk].
13. L.V. Zlatoustova, S.V. Kodzasov, O.F. Krivnova, I.G. Frolova (1970): “Algorithms for converting Russian orthographic texts into a phonetic record”, Moscow, MSU Publishing House [“Algoritmy preobrazovaniya russkikh orfograficheskikh tekstov v foneticheskuyu zapis'”].