



КАРПОВСКИЕ НАУЧНЫЕ ЧТЕНИЯ

Выпуск 8

Часть II

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ



КАРПОВСКИЕ НАУЧНЫЕ ЧТЕНИЯ

Сборник научных статей

Основан в 2007 году

Выпуск 8

В двух частях

Часть 2

Минск
«ИВЦ Минфина»
2014

УДК 80/81 (082)

В сборнике представлены материалы Восьмых Карповских научных чтений, посвященных памяти профессора В.А. Карпова, — ученого, философа, поэта, оставившего заметный след в отечественной культуре.

Адресован филологам, философам, системологам, специалистам по прикладной и компьютерной лингвистике, а также студентам, магистрантам, аспирантам, интересующимся рассматриваемыми проблемами.

Рекомендовано
Ученым советом филологического факультета
Белорусского государственного университета
(протокол № 9 от 19 июня 2014 г.)

Редакционная коллегия:
кандидат филологических наук, доцент А.И. Головня (отв. ред.);
кандидат филологических наук, доцент Н.С. Касюк (зам. ред.);
кандидат технических наук, доцент О.Е. Елисеева

Рецензенты:
кандидат филологических наук, доцент А.В. Лаврененко
кандидат педагогических наук, доцент Т.В. Игнатович

ISBN 978-985-7060-82-5 (ч. 2)
ISBN 978-985-7060-80-1

© БГУ, 2014
© Оформление.
УП «ИВЦ Минфина», 2014

СОДЕРЖАНИЕ

Секция 4: ПРИКЛАДНАЯ ЛИНГВИСТИКА В БЕЛАРУСИ: СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

Савко И.Э. (Минск, БГУ) Обучение студентов использованию интерактивной доски при преподавании русского языка.....	3
Станкевич А.Ю. (Гродно, ГрГУ) Технология сбора и систематизации электронного контента для корпуса русскоязычных СМИ Гродненщины.....	6
Шмыга Д.Л. (Минск, МГЛУ) Подходы к описанию логико-синтаксической и семантико-синтаксической структуры предложения.....	11
Гецэвіч Ю.С., Гюнтар А.В. (Мінск, АПП НАН Беларусі) Распрацоўка лінгвістычных рэсурсаў для тэставання фанемна-алафоннага працэсара сінтэзатара беларускага маўлення.....	15

Секция 5: ПРОБЛЕМЫ ПЕРЕВОДА

Абреу-Фамлюк В.Р. (Минск, МГЛУ) Структурно-синтаксические модели предложений инструкций в системе автоматического франко-русского перевода.....	19
Астапкина Е.С. (Минск, БГУ) Закономерности семантической деривации температурных прилагательных в сопоставительном аспекте (на материале русского, белорусского, английского и немецкого языков)	21
Близнюк-Бискуп Е.В. (Варшавский университет) О способах передачи имплицитной информации в польских переводах «Петербургских повестей» Н. Гоголя.....	25
Брагарник-Станкевич О.С. (БГУ, Минск) Russian bestial verbal lexicon: the problem of translation. Бестиальные глаголы русского языка: Проблема перевода.....	28
Цзин Го (Китай-Беларусь, Минск, БГУ) Особенности фразеологического перевода библейских фразеологизмов на китайский язык.....	30
Иванова Н.А. (Гродно, ГрГУ) Культурный компонент в профессиональном переводе.....	34
Клокова А.Г. (Минск, БГПУ) Лингвистическая интерпретация художественного текста как попытка реконструкции авторской стратегии.....	38

РАСПРАЦОЎКА ЛІНГВІСТЫЧНЫХ РЭСУРСАЎ ДЛЯ ТЭСТАВАННЯ ФАНЕМНА-АЛАФОННАГА ПРАЦЭСАРА СІНТЭЗАТАРА БЕЛАРУСКАГА МАЎЛЕННЯ

На сённяшні дзень задача сінтэза маўлення з'яўляецца не новай, аднак не губляе сваёй актуальнасці з-за імкнення палепшыць працу існуючых сістэм генеравання чалавечага маўлення машынай.

Сінтэз вуснага маўлення па тэксце ажыццяўляецца на аснове лексіка-граматычнага аналізу ўваходнага тэксту. Арфаграфічны тэкст дакумента паступае на ўваход сінтэзатара і далей падвяргаецца паслядоўнай апрацоўцы спецыялізаваных працэсараў у адпаведнасці з агульнай структурай сінтэзатара маўлення па тэксце. Сінтэзатар уключае чатыры асноўныя модулі: тэкставы працэсар, прасадычны працэсар, фанетычны працэсар і акустычны працэсар [1, с. 111].

Фанетычны працэсар пераўтварае арфаграфічны тэкст у фанемны і генеруе пазіцыйныя і камбінаторныя алафоны. Нагадаем, што алафон – гэта рэалізацыя фанемы, яе варыянт, абумоўлены пэўным фанетычным асяроддзем, які з'яўляецца канкрэтным маўленчым гукам. Нягледзячы на шырокі дыяпазон алафонаў адной фанемы, носьбіт мовы заўсёды ў стане іх распазнаць. Менавіта таму так важна распрацаваць дакладную базу ведаў для тэставання фанемна-алафоннага працэсара сінтэзатара беларускага маўлення. Гэта дазволіць згенераваць усе пазіцыі алафонаў для далейшага тэставання і выпраўлення магчымых памылак у працы сінтэзатара.

Раней ужо былі распрацаваны адзіныя табліцы правілаў пераўтварэння “Фанема-Алафон” для рускай мовы [1, с. 253–258], у якіх алафон пазначаецца імем фанемы, магчымай мяккасцю і наступнымі за ім трыма лікавымі індэксамі, якія абазначаюць: для галосных – націскны, часткова націскны ці ненаціскны гук, а таксама левае і правае асяроддзе; для зычных – адзіночны ці падвоены алафон, левае і правае асяроддзе. Улічваючы гэта, а таксама пэўныя асаблівасці беларускай мовы, былі распрацаваны адпаведныя табліцы правілаў пераўтварэння “Фанема-Алафон” для беларускай мовы. Гэтыя табліцы для беларускай мовы налічваюць 761 алафон, але падлічаная колькасць алафонаў паводле табліц не з'яўляецца дакладнай. Гэта тлумачыцца тым, што ў табліцах прысутнічаюць тэарэтычна спрагназаваныя алафоны, г.зн., што алафоны не былі правераныя на практычнае ўжыванне ў рэальных словах, дзе яны сустракаюцца хаця б адзін раз. Таму на аснове гэтых табліц была распачата праца над стварэннем лінгвістычнага рэсурса для тэставання фанемна-алафоннага працэсара сінтэзатара беларускага маўлення і выяўлення недахопаў і недакладнасцей у вышэй пазначаных табліцах. Дадзены рэсурс уяўляе сабой спіс алафонаў для беларускай мовы колькасцю 761 адзінка, на кожны з якіх неабходна знайсці прыклад слова ці адзначыць, што такога прыклада не знайшлося.

Базай для пакрыцця дадзеных алафонаў з'яўляліся спецыяльна створаныя тэматыкі слоў («прафесіі», «ежа і напоі», «геаграфічныя назвы» і

г.д.). Для пераўтварэння іх у алафонны выгляд быў скарыстаны спецыяльны сэрвіс «Text-to-speech synthesis», які дазваляе атрымліваць алафонны выгляд арфаграфічнага тэксту з расстаўленымі асноўнымі і пабочнымі націскамі (малюнак 1). Дадзены сэрвіс распрацаваны ў межах Інтэрнэт-версіі сінтэзатара беларускага маўлення (www.corpus.by) [2], які ўключае ў сябе і іншыя рэсурсы, якія прызначаны для вырашэння розных лінгвістычных задач (праверка «ў» і «у», пошук і сарціроўка алафонаў, канвертаванне алафоннага тэксту ў розныя транскрыпцыі і некаторыя іншыя).

Text-to-Speech PHP-Based Synthesizer

Please input a stressed text

Primary stressed vowel must be marked by '^' or 'ˆ', a secondary stressed vowel - by 'ˆ' or 'ˆ'.
To mark two words as one phonetic word use 'b'.

Example with '^', 'ˆ' and 'b': Паўночна-заходні вятры+ска садзьму+ў усё= лі=сце набы+с=не.
Example with 'ˆ', 'ˆ' and 'b': Паўночна-заходні вятры+ска садзьму+ў усё= лі=сце набы+с=не.

Clear / Refresh page

Мой родны кут, дзя ты мне мілы!
Забыць цдзе= няўмаю сілы!
Нільра=з, ую=млени дарогай,
Жыццэ=м вясны+ мае= убо+гай,
Кўтабе= я= ўдв+мках залата+ю.

Belarusian (Беларуская мова) Generate Speech! Show log information

Listen to the generated speech

or download the generated speech file.
Created during 0 sec, language bel.

Log information about text-to-speech processing of:

Text
(change to UTF-8 after downloading)

MO=Й_не_в_д_о_м_а_я_ч_а_с_ц
w PO+ДНы_не_в_д_о_м_а_я_ч_а_с_ц
w КУ+Т_не_в_д_о_м_а_я_ч_а_с_ц

Intonation markers
(change to UTF-8 after downloading)

MO=Й
PO+ДНы
КУ+Т
- - -

Phonemes
(change to UTF-8 after downloading)

M,O,=,J,/,R,O,+,D,N,Y,K,U,+,T,/,R,C3
Г,А,=,К,/,Т,=,/,M,N,E,=,/,M',J,+,L,Y,/,R,E2

Allophones (Latin)
(change to UTF-8 after downloading)

M004,O113,J013,/,R032,O022,D001,N004,Y322,/,K001,U032,T000,/,R,C3.

Transcription in cyrillic letters

Мал. 1. Прыклад работы сінтэзатара маўлення www.corpus.by/tts3.

Наступны крок у падрыхтоўцы матэрыялу для пакрыцця наяўнага спіса алафонаў – іх пошук і сарціроўка. Дадзены працэс таксама аўтаматызаваны: сэрвіс «Service of Searching and Sorting Allophones» [3] дазваляе атрымаць спіс алафонаў з прыкладамі слоў у алфавітным парадку ці па колькасці сустракання ў зададзеным тэксце (малюнак 2). Знойдзеныя алафоны пазначаны колерам, што палягчае іх пошук у кантэксце (слове ці спалучэнні слоў). Пасля атрымання гэтых дадзеных адбываецца іх праверка на правільнасць і ўнясенне ў спецыяльны файл, дзе пазначаны алафон, алафонны выгляд слова, арфаграфічны выгляд слова, а таксама тэматыка слова (малюнак 3).

Service of Searching and Sorting Allophones

Please input a text with allophones

Калі ласка, увардаце тэкст з алафонамі

Clear / Ачысціць

Please input stop-allophones

Калі ласка, увардаце стоп-алафоны

```
D004.A322.A003.G233.M002.E042.I003.T002.Y323.J.F012.D41.  
/M004.A312.T002.E322.R002.Y323.F011.A041.L004.V312./A323.D2004.I42.M000./#P4.  
/M04.T002.B42.G004.A322.L004.A323.S002.S002.E042.D004.A321.V012.A113.N104.E342.  
/A321.F102.R012.Y322.R002.O023.D2004.E343./J012.I141.  
/P002.R012.Y322.C000.L7023.N004.A022.S004./B004.A313.G002.A033.C102.A342.  
/M04.A322.R002.D022.D004.A321.M011.J.B004.Y311.R022.A012.  
/A321.R002.O211.B000.L1004.E042.G004.A323.V011.A032.R004.A321.  
/W012.A112.M001.G004.L1004.I43.I141./W013.T002.Y123.S002.A342.C002.A320./#C3.
```

U131

I332

Y320

U110

E140

A008

A001

A002

Get allophones / Знайсці алафоны

Get allophones only (Толькі алафоны)

The total number of elements is 20.
Different elements are 6:

Allophones sorted by alphabet

A320, occurs 8 times) in

```
U302.T002.Y123.S002.A342.C002./#C3.  
W002.O313.S002.E341.M003.S002.O122.T002./P002.A343.D2104.E343.S002.A142.T002./S002.E141.M003.T002.Y123.S002.A342.C002./#C3.  
U302.T002.Y123.S002.A342.C002./#C3.  
/M02.I42.R001.P002.I142.N003./P002.I42.S002.E122.R003./T002.Y123.S002.A142.C002.A320./#C3.  
W002.O313.S002.E341.M003.S002.O122.T002./S002.E322.S002.D2104.E343.S002.A142.T002./S002.E141.M003.T002.Y123.S002.A342.C002./#C3.  
W002.O313.S002.E341.M003.S002.O122.D001./V002.O113.S002.E341.M001.D2004.E343.S002.A342.T002./T002.R022.Y122.  
/T002.Y123.S002.A342.C002./#C3.  
W002.O313.S002.E341.M003.S002.O122.D001./V002.O113.S002.E341.M001.D2004.E343.S002.A342.T002./T002.R022.Y122.  
/T002.Y123.S002.A342.C002./#C3.  
G004.E132.T002.A320./#P1.
```

Мал. 2. Прыклад работы сэрвіса пошуку і сартыроўкі алафонаў
<http://corpus.by/allophones-get/>

Пасля апрацоўкі сабраных тэматык стала зразумела, што дадзенага матэрыялу недастаткова для поўнага пакрыцця алафонаў беларускай мовы. Таму з электроннай версіі арфаграфічнага слоўніка беларускай мовы былі ўзяты ўсе пачатковыя формы назоўнікаў і апрацаваны пры дапамозе таго ж сэрвіса. Але і гэтага матэрыялу не хапіла. Многія пазіцыі алафонаў сустракаюцца на сутыку слоў, таму неабходна было ўлічваць не толькі адзінкавыя словы, але і розныя спалучэнні слоў. З гэтай мэтай для апрацоўкі быў узяты падручнік «Грамадазнаўства» за 10 клас, які папярэдне быў апрацаваны праграмай VocEditor. Гэта праграма дазволіла знайсці і апрацаваць усе словы з невядомай пазіцыяй націску, а таксама словы-амографы.

Для складання прыкладаў таксама былі выкарыстаныя дадзеныя пры выкарыстанні сэрвісаў «Text-to-speech synthesis», «Service of converting allophonic texts into different transcriptions» і «Orthoexic Dictionary Generator» карыстальнікамі. Вынікі працы алгарытмаў гэтых рэсурсаў выдаюцца карыстальніку ў выглядзе мэтавых транскрыпцый, а эксперту-лінгвісту і інжынеру-праграмісту — ў выглядзе ўсіх уведзеных уваходных, выходных і аналітычных (велічыня тэксту, IP-адрас, час і т.д.) дадзеных для аперацыўнага выпраўлення памылак. Пасля аналізу згенераваных транскрыпцый некаторыя алафоны атрымлівалі прыклад выкарыстання.

Паколькі пасля ўсёй праведзенай працы пакрыццё алафонаў не перавышала мяжу 604 адзінкі, было вырашана шукаць алафоны, для якіх яшчэ не знойдзены адпаведнікі ўручную.

У выніку выкананай працы пакрыццё алафонаў складала 704 адзінкі, пры гэтым колькасць алафонаў для беларускай мовы будзе прапанавана па-

меншыць, бо створаныя адзіныя табліцы правілаў пераўтварэння «Фанема-Алафон» улічваюць абедзве мовы (рускую і беларускую), а таму ў іх сустракаюцца пэўныя недакладнасці. Напрыклад, паводле дадзеных табліц існуе пазіцыя г-фрыкатыўнага, калі яму папярэднічае любая фанема ці дапаможны сімвал, а пасля яго ідзе паўза, іншымі словамі г-фрыкатыўны на канцы слова. Па правілах жа беларускай мовы ён аглушаецца, таму гэты алафон можна выдаліць з агульнага спіса алафонаў для беларускай мовы.

	A	B	C	D	E
1	Алафон	Алафонны выгляд	Слова	Тэматыка	
2	A000	A000, #E2,	А+!	Выклічнік.	
3	A001	W013,S002,T002,R022,Y323,J012,A340, #P4,	А+Устрыя	Геаграфічныя назвы	
4	A002	A002,N001,GH004,I304,I343,J012,A340, #P4,	А+нглія	Геаграфічныя назвы	
5	A003	A003,Z004,I343,J012,A340, #P4,	А+зія	Геаграфічныя назвы	
6	A010	T002,U222,R001,M002,A010, #P4,	Турызм	Месцы і ўстанова	
7	A011	P002,L002,A011,W013,K002,I340, #P4,	Плаўкі	Воаграпа, яе элементы	
8	A012	S002,P002,A241,V011,A012,K000, #P4,	Слова к.	Прафесіі	
9	A013	P002,R012,A322,T002,P002,R012,Y221,M002,A013,L001,N004,I342,K000, #P4,	Прадпрыемствы	Прафесіі	
10	A020	M004,A242,ZH002,A020, #P4,	Мажэ-	Месцы і ўстанова	
11	A021	K004,A332,S002,M004,A212,N002,M021,W013,T000, #P4,	Касман-ўт.	Прафесіі	
12	A022	R022,E221,A022,L000, #P4,	Рэ-л.	Грошы	
13	A023	D004,Y222,Z002,A023,J013,N004,E342,R000, #P4,	Дыя-бнер	Прафесіі	
14	A030	R022,U222,K002,A030, #P4,	Рука-	Часткі цела	
15	A031	K002,A031,V012,A310, #P4,	Катва	Ежа і напоі	
16	A032	M004,I312,Z004,Y222,K002,A032,N003,T000, #P4,	Музыкант.	Прафесіі	

Мал. 3. Фрагмент лінгвістычнай рэсурсаў для пакрыцця алафонаў беларускай мовы ў прымяненні да сінтэзатара маўлення па гэксце

Другі лінгвістычны рэсурс з'яўляецца яшчэ адной базай ведаў, у якой пазначаецца арфаграфічны выгляд слова/спалучэння слоў, іх трансфармацыя ў алафонны выгляд стацыянарным і інтэрнэт-версіяй сінтэзатара маўлення, а таксама паметы лінгвіста наконт знойдзеных памылак у такой трансфармацыі. Для апісання памылак вынікі генерацыі сінтэзатараў правяраюцца, знойдзеныя памылкі пазначаюцца чырвоным колерам без унясення зменаў, а ў асобную графу ўносіцца правільны алафонны выгляд слова/спалучэння слоў. На аснове дадзенай базы ведаў праца сінтэзатараў будзе паляпшацца шляхам унясення ў іх змен з улікам пазначаных заўваг.

Такім чынам, распрацоўка лінгвістычных рэсурсаў для тэставання фанемна-алафоннага працэсара сінтэзатара беларускага маўлення з'яўляецца важнай і разам з тым складанай задачай. Магчымасць аўтаматычнай апрацоўкі некаторых дадзеных дазваляе значна палегчыць працу лінгвіста па вырашэнні дадзенай праблемы.

ЛІТАРАТУРА

1. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи. Минск: Беларуская наука, 2008.
2. Text-to-Speech PHP-Based Synthesizer [Electronic resource]. – 2013. – Mode of access : <http://corpus.by/tts/>. – Date of access: 09.03.2014.
3. Service of Searching and Sorting Allophones / Сэрвіс пошуку і сартыроўкі алафонаў [Electronic resource]. – 2013. – Mode of access : <http://corpus.by/allophones-get/>. – Date of access: 08.12.2013.