



КАРПОВСКИЕ НАУЧНЫЕ ЧТЕНИЯ

Выпуск 8

Часть I

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ



КАРПОВСКИЕ НАУЧНЫЕ ЧТЕНИЯ

Сборник научных статей

Основан в 2007 году

Выпуск 8

В двух частях

Часть 1

Минск
«ИВЦ Минфина»
2014

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

Секция 3: ОБЩАЯ ТЕОРИЯ СИСТЕМ КАК МЕТОДОЛОГИЯ НАУКИ

Барбук С.Г. (Минск, БГЭУ) Языковые универсалии.....	221
Ван Цзин (Минск, БГУ) Возможность изучения коннотации в корпусном исследовании.....	225
Ван Цин (Минск, БГУ) Особенности имени существительного в грамматике русского и китайского языков.....	228
Гецэвіч Ю.С., Окрут Т.І., Міхайлова Я.А. (Мінск, НАН РБ, БДУ) Распрацоўка лінгвістычных рэсурсаў для алгарытмаў ідэнтыфікацыі рэшлік дыялогаў у электронных тэкстах мастацкай тэматыкі на беларускай і рускай мовах.....	231
Гецэвіч Ю.С., Скопінава А.М. (Мінск, АПП НАН Беларусі) Лінгвістычныя рэсурсы для пераўтварэння колькасных выразаў з адзінкамі вымярэння тыпу «лічба-сімвал» у словазлучэнні для беларускай і рускай мовы.....	236
Гецэвіч Ю.С. (Мінск, АПП НАН Беларусі), Барадзіна Ю.С. (Мінск, БДУ) Класіфікацыя фразаў дыялогаў па эматыўных прыкметах на матэрыяле рускіх і беларускіх мастацкіх твораў.....	240
Гецэвіч Ю.С., Лысы С.І. (Мінск, АПП НАН Беларусі) Рашэнне прыкладных лінгвістычных задач пры дапамозе сэрвісаў рэсурсу www.corpus.by	243
Глинка Е.В. (Мінск, МГЛУ) Проявление симметрии и асимметрии в условиях русско-белорусского двуязычия	247
Головня А.И. (Мінск, БГУ) Системная симметрично-асимметричная номинация в абревиации.....	251
Иващенко В.П. (Мінск, БГУИР) Пространственно-временные интервальные бинарные отношения на множествах событий и их языковые средства представления.....	255

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

Аскерко Д.С. (Мінск, МГЛУ) Специфика базы данных системы автоматического определения средств выражения вербальной агрессии в текстах англоязычных СМИ.....	259
Бочкова А.Л. (Мінск, МГЛУ) Лингвистическая база данных как основа системы автоматического извлечения мнений участников интернет-коммуникаций.....	263
Гусева Н.Ю. (Мінск, БГУ) Трудности в преподавании курса «Основы информационных технологий» иностранным студентам.....	266

Граматыкі для рускай і беларускай моў, пазітыўнай і негатыўнай эматыўных танальнасцяў былі прыменены да матэрыяла, і праграма знайшла ўсе вызначаныя экспертым сказы, напрыклад, — *Спортсменом будет*, — *улыбается главный врач Михель Тенне*; — *Довлатов*, — *исполненным мужи голосом произнес Туронок*, — *Довлатов, я вас уволю*; — *Што?* — *у жаху закрычалі ваяры* і г.д.

Такім чынам, былі сабраны калекцыі ідэнтыфікатораў эматыўнай палярнасці тэкста для рускай і беларускай моў, вынесены рэкамендацыі па іх фармалізацыі і распрацаўваны тэставыя граматыкі па знаходжанні ў корпусе эматыўна палярных фразаў простай мовы са словамі аўтара.

Тэма патрабуе далейшай распрацоўкі, таму будуць павялічаныя тэставыя корпусы і, адпаведна, спісы ідэнтыфікатораў дзеля паляпшэння якасці працы граматык. Акрамя таго, вызначэнне эмоцыяў будзе весціся не толькі па словах аўтара, але і па простай мове.

ЛІТАРАТУРА

1. Гецэвіч, Ю.С. Аўтаматызацыя шматгласавога стварэння аўдыёўкніг на беларускай мове з дапамогай сінтэзатораў маўлення па тэксце / Ю.С. Гецэвіч, Т.І. Окрут, Б.М. Лабанаў // Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2013): доклады XII Международной конференции (Минск, 20 ноября 2013 г.). – Минск: ОІПІ НАН Беларуси, 2013. – С. 269–276.

2. Hetsevich, Y. Overview of Belarusian And Russian dictionaries and their adaptation for NooJ //Automatic Processing of Various Levels of Linguistic Phenomena: Selected Papers from the NooJ 2011 Intern. Conf. / eds. Vučković Kristina, Bekavac Božo, Silberstein Max. – Newcastle: Cambridge Scholars Publishing, 2012. – P. 29–40.

3. Лінгвістычны працэсар NooJ [Электронны рэсурс]. – 2002. – Рэжым доступу: <http://nooj4nlp.net/pages/nooj.html>. – Дата доступу: 10.03.2014

Ю.С. Гецэвіч, С.І. Лысы (Мінск, АПІ НАН Беларусі)

РАШЭННЕ ПРЫКЛАДНЫХ ЛІНГВІСТЫЧНЫХ ЗАДАЧ ПРЫ ДАПАМОЗЕ СЭРВІСАЎ РЭСУРСУ WWW.CORPUS.BY

Распрацоўка сістэмы сінтэзу маўлення па тэксце (ССМТ) з'яўляецца складанай задачай, якая ахоплівае вялікую колькасць неабходных падзадач, а таксама сутыкаецца з шэрагам пабочных прыкладных лінгвістычных задач, якія патрабуюць решэння. Задачы, якія паўстаюць пры стварэнні ССМТ, можна ўмоўна падзяліць на тры групы:

- перадапрацоўка ўваходных тэкстаў;
- непасрэдная сінтэз маўлення;
- апрацоўка выніковых дадзеных.

Інтэрнэт-рэсурс www.corpus.by [1] мае сваёй мэтай решэнне як непасрэднай задачы сінтэзування маўлення, так і шэрагу пабочных задач перадапрацоўкі тэкstu і апрацоўкі вынікаў працы сінтэзатора маўлення. Дзеля дасягнення гэтай мэты было прынята решэнне распрацаўваць сістэму невялікіх інтэрнэт-сервісаў, кожны з якіх вырашаў бы пэўную ўласную задачу, а ў спалученні з іншымі сервісамі — неабходныя падзадачы ССМТ.

Але перш чым перайсці да разгляду інтэрнэт-сэrvісаў, створаных у падтрымку да сінтэзатора маўлення, звернемся да кароткага агляду самога сінтэзатора маўлення па тэксле (CMT) [2], які ён ёсьць на сённяшні дзень.

Сінтэзатор маўлення па тэксле з'яўляецца прыладай перапрацоўкі тэксту, пададзенага ў звычайнім (графічным) выглядзе, у тэкст у гукавым выглядзе. Знешні інтэрфейс распрацаванага CMT паказаны на малюнку 1. Каб атрымаць тэкст у гукавым выглядзе, карыстальнік павінен увесці гэты тэкст у адпаведнае вакно, абраць мову і націснуць кнопкі “Generate Speech!”. Для лепшага, больш зразумелага гучання тэксту, карыстальнік мусіць пазначыць у кожным слове асноўны націск знакам “+” (пло+с) пасля націскнай галоснай, альбо знакам — “»” (акут) над націскнай галоснай. Адпаведна дадатковыя націскі — знакам “—” альбо “»” (гравіс). Гэтая неабходнасць выклікана тым, што, на сённяшні дзень, да сінтэзатора не далучаны слоўнік. Націснуўшы кнопкі “Generate Speech!”, карыстальнік атрымлівае гукавы файлік у фармаце wav, які можа праслухаць не сыходзячы са старонкі, спампаваць альбо падзяліцца спасылкай, па якой гукавы файлік будзе ўвесь час даступны. Таксама карыстальнік мае магчымасць атрымаць прамежкавую інфармацыю аб выніках працы тэкставага, фанемнага, алафоннага і інш. блокаў CMT, папярэдне пазначыўшы птушачкай “Show log information”.

Text-to-Speech PHP-Based Synthesizer

Please input a stressed text

Primary stressed vowel must be marked by '+' or '^', a secondary stressed vowel - by '=' or '». Example with '+' and '^': Пойноччно-захо-дні ватры-ско садзьмы-у усё-лі-це навы-спе.
Example with '=' and '»': Пойноччно-захо-дні ватры-ско садзьмы-у усё-лі-це навы-спе.

Ліней.
адзін.
два+.
тын.
чатыры.
пяць.
шасць.

Belarusian (Беларуская мова) Show log information

Listen to the generated speech

Or download the generated speech file. Created during 0 sec. language: bsl

Мал. 1. Знешні інтэрфейс сінтэзатора маўлення па тэксле ў інтэрнэце.

Цяпер пярэйдзем да апісання тых інтэрнэт-сэrvісаў, якія былі распрацаваны ў прылажэнне да CMT.

Адным з важных момантаў у сінтэзаванні маўлення з'яўляецца тое, што ўваходны тэкст мусіць не мець арфаграфічных памылак. Іначай, відавочна, вынікам працы CMT будзе некарэктны гукавы тэкст, які рэжа слых. “Сэрвіс праверкі “у” і “ў” [3] распрацаваны для хуткай праверкі вялікіх электронных тэкстаў на беларускай мове з мэтай пошуку і

выпраўлення адной з найбольш распаўсюджаных памылак — памылкі ў напісанні літар “у” і “ў”. На ўваход сэрвісу падаецца тэкст на беларускай мове ў звычайнym выглядзе. Калі карыстальнік націсне кнопкu “Праверыць!”, сэрвіс абазначыць магчымыя памылкі двух тыпаў: папершае, калі напісана літара “у” ў пазіцыі, дзе верагодна павінна быць літара “ў”, па-другое, у адваротным выпадку, калі напісана літара “ў”, дзе верагодна павінна быць літара “у”.

Сэрвіс праверкі “ў” і “у”

Please input a text / Калі ласка, увядзіце тэкст

Search / Нахадзіць

Напрыклад,

Мама ў краме. {Сказ без памылки}.
Мама у краме. {Сказ з памылкай}.
Кот у ботак. {Сказ без памылки}.
Кот у ботак. {Сказ з памылкай}.
А я тэ ёсць пчолы. {Сказ з памылкай}.

Праверыць

Вынікі праверкі

Магчыма, патрабна пісаць ў ці у:

Сустралася “а” у радку “Мама ў краме. {Сказ з памылкай}.”.
Сустралася “а” у радку “У іх ёсць пчолы. {Сказ з памылкай}.”.
Сустралася “а” у радку “Кот у ботак. {Сказ з памылкай}.”.
Сустралася “а” у радку “На Украіне паўднёвасхідні вечер. {Сказ з памылкам}.”.
Сустралася “а” у радку “Сястра есьць пусянку. {Сказ з памылкай}.”.

Магчыма, патрабна пісаць У ці У:

Сустралася “т” у радку “Кот 9 ботак. {Сказ з памылкай}.”.
Сустралася “т” у радку “Брат” у краме. {Сказ з памылкай}.”.
Сустралася “т” у радку “На Украіне паўднёвасхідні вечер. {Сказ з памылкам}.”.
Сустралася “м” у радку “Усім ўсім пра ўсё распавяждам. {Сказ з памылкай}.”.
Сустралася “б” у радку “Тата любіць бульбі. {Сказ з памылкай}.”.

Мал. 2. Знешні інтэрфейс і прыклад працы “Сэрвісу праверкі “у” і “ў”.

У працэсе пошуку магчымых памылак сэрвіс ня толькі вызначае, галосная ці зычная літара знаходзіцца перад “у”, але таксама аналізуе сімвалы, якія не з'яўляюцца літарамі, у выпадку, калі літара “у” знаходзіцца ў пачатку слова. Гэтыя сімвалы могуць мець непасрэдны ўплыв на напісанне слова. На сённяшні дзень сэрвіс правярае напісанне без уліку выключэнняў. Дадаць разгляд выключэнняў у сэрвісе плануецца.

Яшчэ адзін з сэрвісаў перадапрацоўкі тэксту — *“Сэрвіс генераціі назваў сімвалаў”* [4]. Напрыклад, карыстальнік надрукаваў слова “ма+ма”, а СМТ пасля апрацоўкі зачытвае нешта незразумелае накшталт “мма”. У чым праблема? Карыстальнік можа звярнуцца да сэрвісу і даведацца, з якіх сімвалаў складаецца яго запыт да сінтэзатора. У выніку апрацоўкі сэрвісам пазначанага вышэй слова, карыстальнік атрымае інфармацыю аб tym, што другі сімвал мае назыву “Лацінская малая літара эй”. Літара не кірылічная, і СМТ не можа яе апрацаўваць.

Сэрвіс створаны дзеля таго, каб даць магчымасць карыстальніку атрымаць назывы сімвалаў кадыроўкі Windows-1251. На ўваход падаецца любая паслядоўнасць сімвалаў (текст, асобныя слова, выпадковы набор

сімвалаў) дадзенай кадыроўкі. Карыстальнік можа задаць адну з трох моваў (беларускую, рускую ці англійскую), а таксама абраць форму вываду інфармацыі (назвы сімвалаў у слупок альбо праз коску). Мэтай дадзенага сэрвісу з'яўляецца вырашэнне праблемы агучвання тэксту, у якім сустракаюцца незнаёмыя сінтэзатору сімвалы.

Адным з сэрвісаў апрацоўкі выніковых дадзеных СМТ з'яўляецца **“Сэрвіс канвертавання алафоннага тэксту ў розныя транскрыпцыі”** [5]. Сэрвіс прызначаны для генеравання транскрыпций па ўваходным тэксле. На ўваход падаецца тэкст у алафонным выглядзе. Ніжэй прыведзены першыя два радкі знакамітага беларускага верша ў алафонным запісе:

M004,O113,J'013,/R032,O022,D001,N004,Y322,/K001,U032,T000,#C3,
J'002,A142,K004,/T002,Y121,/M001,N'004,E143,/M'002,I041,L004,Y310,#
E2,

(Алафонны запіс слова можна атрымаць пры дапамозе апісанага вышэй сінтэзатора маўлення, які знаходзіцца па адрасе www.corpus.by/tts3, пры ўключанай опцыі Show log information).

Увёўшы тэкст, карыстальнік можа абраць адзін альбо некалькі тыпаў транскрыпціі з трох магчымых.

<p>Transcription in cyrillic letters:</p> <p>[мой] [рóдны] [кút] [йáк] [ты] [мн'é] [м'лы] [забыц'] [ц'аб'è] [н'амáйу] [с'лы]</p>	<p>Transcription in latin letters:</p> <p>[mòj] [ródny] [kút] [ják] [tì] [mn'è] [m'ily] [zabýts'] [tśab'è] [n'amáju] [s'ily]</p>
<p>Transcription in International Phonetic Alphabet:</p> <p>[m,ɔj] [r'odni] [k'ut] [j,ejk] [t,i] [mn'e] [m'ihi] [zeb'itš] [tśeb'e] [n'am'aju] [s'ihi]</p>	

Мал. 3. Прыйклад працы “Сэрвісу канвертавання алафоннага тэксту ў розныя транскрыпцыі”.

Першы тып транскрыпціі — кірылічны, ён можа быць карысны школьнікам, настаўнікам, студэнтам і выкладчыкам ВНУ. Наступныя два — лацінскія: спропечаная лацінская транскрыпцыя і транскрыпцыя ў Міжнародным фанетычным алфавіце (International Phonetic Alphabet — IPA) [6]. Канвертаванне тэксту ў спропечаную лацінскую транскрыпцыю распрацавана паводле працы У.А. Кошчанкі [7]. Транскрыпцыя ў Міжнародным фанетычным алфавіце арыентаваная на інтэрналізацыю беларускай мовы. Яна можа быць карыснай як для навуковых працаў у сувязі з беларускай мовай (прыкладам, стварэнне арфаэпічнага слоўніка), так і

для звычайных карыстальнікаў (асабліва замежнікаў), якія, дзякуючы гэтаму сэрвісу, атрымліваюць магчымасць убачыць гукавы склад беларускіх слоў паводле міжнароднага стандарту.

У выніку паставленай задачы была створаная сістэма інтэрнэт-сэрвісаў у прылажэнне да СМТ, у тым ліку “Сэрвіс праверкі “у” і “ў”, “Сэрвіс генерацыі называў сімвалу (кадыроўка Windows-1251)», “Сэрвіс канвертавання алафоннага тэксту ў розныя транскрыпцыі” і інш. Гэтыя сэрвісы знаходзяцца ў вольным доступе рэсурсу www.corpus.by. У будучыні плануецца праца над удасканаленнем ужо існуючых інтэрнэт-сэрвісаў, інтэграцыя распрацаваных інтэрнэт-сэрвісаў у больш складаныя сістэмы, а таксама стварэнне новых інтэрнэт-сэрвісаў.

ЛІТАРАТУРА

1. Corpus.by [Electronic resource]. – 2013. – Mode of access: <http://corpus.by/>. – Date of access: 06.03.2014.
2. Text-to-Speech PHP-Based Synthesizer [Electronic resource]. – 2013. – Mode of access: <http://corpus.by/tts3/>. – Date of access: 06.03.2014.
3. Сэрвіс праверкі “ў” і “у” [Electronic resource]. – 2013. – Mode of access: http://corpus.by/u_check/. – Date of access: 06.03.2014.
4. Service of generating names of characters (encoding Windows-1251) / Сэрвіс генерацыі называў сімвалу (кадыроўка Windows-1251) [Electronic resource]. – 2014. – Mode of access: <http://corpus.by/getNamesOfCharacters/>. – Date of access: 06.03.2014.
5. Service of converting allophonic texts into different transcriptions / Сэрвіс канвертавання алафоннага тэксту ў розныя транскрыпцыі [Electronic resource]. – 2013. – Mode of access: <http://corpus.by/convertAllophToDifPhonemes/>. – Date of access: 06.03.2014.
6. Wiktionary: International Phonetic Alphabet. [Electronic resource]. – 2008. – Mode of access: http://en.wiktionary.org/wiki/Wiktionary:International_Phonetic_Alphabet. – Date of access: 07.03.2014.
7. Беларуска-англійскі размоўнік / уклад. У. А. Кошчанка. – Мінск.: Артыя Груп, 2010.

Е.В. Глинка (Мінск, МГЛУ)

ПРОЯВЛЕНИЕ СИММЕТРИИ И АСИММЕТРИИ В УСЛОВИЯХ РУССКО-БЕЛОРУССКОГО ДВУЯЗЫЧИЯ

Русский язык в республике Беларусь имеет особый статус второго государственного языка. Поэтому языковая ситуация в Беларуси имеет свою специфику. Русский язык используется повсеместно городским населением в качестве основного языка общения. Белорусский язык, хотя и изучается в школе, такого распространения, к сожалению, ещё не находит. Очевидно, здесь сказывается влияние долговременного вхождения Беларуси в состав республик бывшего Советского Союза и традиции близкородственных отношений наших стран и народов. Но дух любого народа живёт в родном языке, и во многих случаях даже пассивное владение белорусским языком накладывает свой отпечаток на русскую речь белорусов.

Норма русского литературного языка в речи русскоговорящего белорусского населения имеет значительные колебания, которые обусловлены