



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

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 978-985-7060-81-8 (ч. 1)
ISBN 978-985-7060-80-1

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

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., Міхайлава Я.А. (Мінск, НАН РБ, БДУ) Распрацоўка лінгвістычных рэсурсаў для алгарытмаў ідэнтыфікацыі рэплік 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 выказаў з адзінкамі вымярэння тыпу «лічба-сімвал» у словазлучэнні для беларускай і рускай моў.....	236
Гецэвіч Ю.С. (Мінск, АПП НАН Беларусі), Барадзіна Ю.С. (Мінск, БДУ) Класіфікацыя фразеў 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. (Минск, МГЛУ) Проявление симметрии и асимметрии в условиях русско-белорусского двуязычия	247
Головня А.И. (Минск, БГУ) Системная симметрично-асимметричная номинация в аббревиации.....	251
Ивашенко В.П. (Минск, 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. (Минск, МГЛУ) Специфика базы данных системы автоматического определения средств выражения вербальной агрессии в текстах англоязычных СМИ.....	259
Бочкова А.Л. (Минск, МГЛУ) Лингвистическая база данных как основа системы автоматического извлечения мнений участников интернет-коммуникации.....	263
Гусева Н.Ю. (Минск, БГУ) Трудности в преподавании курса «Основы информационных технологий» иностранным студентам.....	266

А.М. Скопінава, А.Ф. Есіс // Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2013): доклады XII Международной конференции (Минск, 20 ноября 2013 г.). – Минск: ОИПИ НАН Беларуси, 2013. – С. 282–287.

4. Апісанне СІ на сайце Міжнароднага бюро мер і вагаў [Электронны рэсурс]. – 2006. – Рэжым доступу: http://www.bipm.org/en/si/si_brochure/general.html. – Дата доступу: 15.04.2014.

Ю.С. Гецэвіч (Мінск, АПН НАН Беларусі), Ю.С. Барадзіна (Мінск, БДУ)

КЛАСІФІКАЦЫЯ ФРАЗАЎ ДЫЯЛОГАЎ ПА ЭМАТЫЎНЫХ ПРЫКМЕТАХ НА МАТЭРЫЯЛЕ РУСКІХ І БЕЛАРУСКІХ МАСТАЦКІХ ТВОРАЎ

Сінтэз маўлення знаходзіць прымяненне ў розных сферах, напрыклад, пры стварэнні аўтаадказчыкаў, натуральна-моўных інтэрфейсаў, агучванні інфармацыйных паведамленняў у транспарце, на вакзале, у аэрапорце і г.д. Акрамя таго, тэхналогіі сінтэзу маўлення могуць выкарыстацца для машыннага стварэння аўдыёкніг. Звычайна гэта вымагае сур'езнай працы дыктараў і актораў, але з дапамогай існуючых напрацовак у сферы сінтэзу маўлення працэс можа быць аўтаматызаваны.

У лабараторыі распазнавання і сінтэза маўлення Аб'яднанага інстытута праблем інфарматыкі НАН Беларусі быў распрацаваны сінтэзатар, які тэхнічна ўжо можа “чытаць кнігі” [1, 269]. Тым не менш, працэс якаснага сінтэзу маўлення яшчэ не скончаны, і застаецца некалькі істотных праблемаў, якія дагэтуль не былі вырашаны, і адна з іх — гэта праблема інтанацыі.

У п'есах аўтары спрашчаюць працу актораў з дапамогай рэмарак, якія падказваюць неабходную інтанацыю. У дыялогах праявілі тэжстаў прысутнічаюць словы аўтара, здольныя выконваць гэтую ж функцыю. Так, прыведзены ніжэй тэкст, калі будзе агучаны акторм у аўдыёкнізе, ніколі не будзе прачытаны манатонна: эмоцыі моўцы бачны праз знакі прыпынку, сутнасць самой фразы, і, не ў апошнюю чаргу, праз «падказкі» ў словах аўтара.

— Ад нараджэння вольныя! — люта роў ён. — Вось вам вашы вольнасьці! Усіх іх выразаць!

Так, для якаснага стварэння аўдыёкніг з дапамогай сінтэзатара маўлення трэба ўлічваць эмоцыі, закладзеныя ў рэпліках герояў і словах аўтара.

Такім чынам, наступны артыкул знаёміць з першаснымі вынікамі працы, мэта якой была ў тым, каб знайсці ў фразях дыялогаў ідэнтыфікатары эматыўнай палярнасці (пазітыўны, негатыўны, нейтральны) і прапанаваць магчымыя сродкі іх фармалізацыі, а таксама працэс іраваць іх на невялікім працоўным корпусе.

Даследаванне вядзецца адначасова для беларускай і рускай моў. Для беларускай мовы ў якасці матэрыяла быў выбраны твор Алеся Рукала «На

службе князя Радзівіла», для рускай — аповесць Сяргея Даўлатава «Компромисс».

Раней супрацоўнікамі лабараторыі распазнавання і сінтэза маўлення быў распрацаваны комплекс граматык па выдзяленню з тэкста прастай мовы разам са словамі аўтара і па вызначэнню ў іх пола моўцы, што таксама накіравана на аптымізацыю працэса стварэння аўдыёкніг з дапамогай сінтэзатара маўлення. З дапамогай адной з такіх граматык з корпусу былі выдзелены ўсе фразы прастай мовы разам са словамі аўтара. Так, агульны памер выніковых канкардансаў склаў 174 рэплікі для беларускай мовы і 551 рэпліка для рускай мовы.

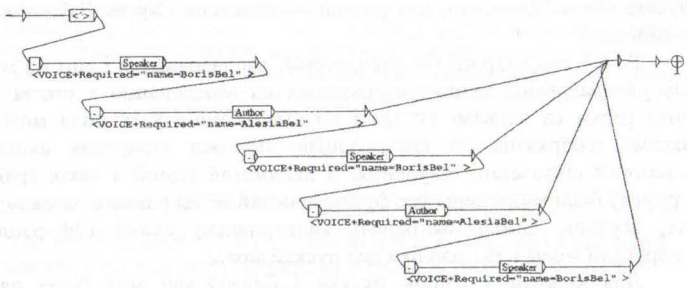
Для кожнай з рэплік рускай і беларускай моў была пазначана эматыўная палярнасць, ідэнтыфікатар, па якім яна вызначаецца і рэкамендацыі для алгарытма.

Напрыклад, фраза: — *Ну вось, — змянін дабрадушна ўсміхнуўся скрозь вусы, — нарэшце пазналі адзін аднаго. Давай абдымаемся.* Эматыўная палярнасць для яе была вызначана як пазітыўная, паводле ідэнтыфікатараў *дабрадушна* і *ўсміхнуўся*, якія ў сваю чаргу з'яўляюцца прыслоўем і дзеясловам. Прапанова для алгарытма апісваецца як ADVERB_pos+VERB_pos і ўяўляе сабой камбінацыю слоў са спісаў AD-VERB_pos (пазітыўна танальныя прыслоўі) і VERB_pos (пазітыўна танальныя дзеясловы).

Пасля апрацоўкі ўсяго матэрыяла вызначылася, што ёсць два асноўных спосабы ўтварэння пазітыўнай ці негатыўнай эматыўнай танальнасці ў словах аўтара: (1) маркіраваныя дзеясловы, якія самі па сабе ствараюць неабходную танальнасць, напрыклад *усміхнуўся, працадзіў, выціснуў, роў*, і (2) ужыванне нейтральных па танальнасці дзеясловаў з эматыўна маркіраванымі прыслоўямі (*беззлобно, счастливо, грустно, сьцюдзена, суха, дабрадушна*) ці спалучэннямі слоў — назоўніка з прыназоўнікамі (з *усмешкай, з цяжкасцю*), дзеяпрыслоўнымі зваротамі (*цяжка дыхаючы, улыбаясь*), і іншымі акалічнасцямі спосабу дзеяння (*ненавидящим шепотом, исполненным муки голосом*). Таксама сустракаюцца выпадкі камбінавання першага і другога спосабаў.

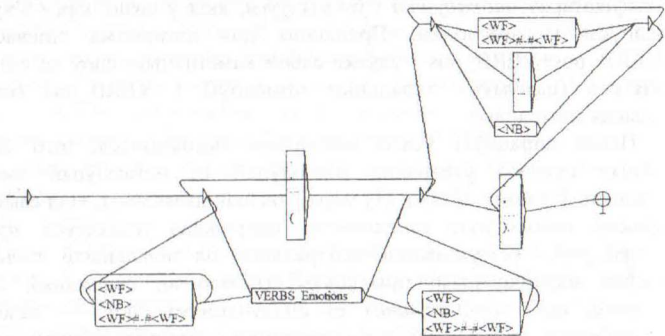
Паводле апісаных спосабаў былі складзеныя наступныя спісы ідэнтыфікатараў: VERBS_neg, VERBS_pos, VERBS_neut, ADVERBS_neg, ADVERBS_pos, COLLOCATION_neg, COLLOCATION_pos.

Каб прагэсціраваць здольнасць кампутара па спісах вызначаць эматыўную танальнасць прастай мовы са словамі аўтара, былі створаны тэставыя сінтаксічныя граматыкі на базе раней распрацаваных у лабараторыі для вызначэння прастай мовы з дапамогай міжнароднай лінгвістычнай праграмы NooJ [2, с. 29; 3]. На малюнку 1 прадстаўлены агульны выгляд граматыкі, дзе блокі Speaker апісваюць простую мову, а Author — словы аўтара.

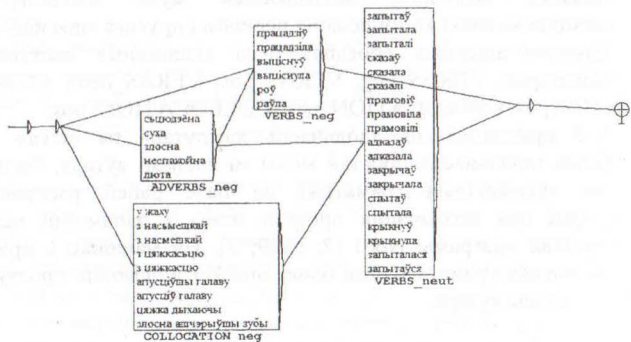


Мал. 1. Граматыка па вызначэнню прастай мовы, агульны выгляд.

У падграфі Author (гл. малонак 2) захоўваецца яшчэ адна граматыка, якая апісвае структуру слоў аўтара і ў якой у блоку VERBS_Emotions размяшчаюцца выпэйапісаныя спісы ідэнтыфікатараў, разам са знакамі прыпынку, уласцівымі для прастай мовы (малонак 3).



Мал. 2. Падграф Author, апісанне слоў аўтара.



Мал. 3. Выгляд падграфу VERBS_Emotions для негатыўнай танальнасці беларускай мовы.

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

Такім чынам, былі сабраны калекцыі ідэнтыфікатараў эматыўнай палярнасці тэкста для рускай і беларускай моў, вынесены рэкамендацыі па іх фармалізацыі і распрацаваны тэставыя граматыкі па знаходжанні ў корпусе эматыўна палярных фразеў простаў мовы са словамі аўтара.

Тэма патрабуе далейшай распрацоўкі, таму будучь павялічаныя тэкставыя корпусы і, адпаведна, спісы ідэнтыфікатараў дзеля паліпшэння якасці працы граматык. Акрамя таго, вызначэнне эмоцыяў будзе весціся не толькі па словах аўтара, але і па простаў мове.

ЛІГАРАТУРА

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

2. Hetsевич, 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, АПП 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 рашэння. Задачы, якія паўстаюць пры стварэнні ССМТ, можна ўмоўна падзяліць на тры групы:

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

Інтэрнэт-рэсурс www.corpus.by [1] мае сваёй мэтай рашэнне як непасрэднай задачы сінтэзавання маўлення, так і шэрагу пабочных задач перадапрацоўкі тэксту і апрацоўкі вынікаў працы сінтэзатара маўлення. Дзеля дасягнення гэтай мэты было прынята рашэнне распрацаваць сістэму невялікіх інтэрнэт-сэрвісаў, кожны з якіх вырашаў бы пэўную ўласную задачу, а ў спалучэнні з іншымі сэрвісамі — неабходныя падзадачы ССМТ.