



КАРПОВСКИЕ НАУЧНЫЕ ЧТЕНИЯ

Выпуск 10

Часть I



КАРПОВСКИЕ НАУЧНЫЕ ЧТЕНИЯ

Сборник научных статей

Основан в 2007 г.

Выпуск 10

В двух частях

Часть 1

В сборнике представлены материалы Десятых 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 № 6 от 14 апреля 2016 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
кандидат филологических наук, доцент А.В. Лаврененко

Рецензенты:
доктор филологических наук, профессор А.В. Зубов;
кандидат филологических наук, доцент Т.В. Игнатович

Махнач Ю.И. (Минск, БГМУ) Реализация лингвострановедческого преподавания на занятиях по РКИ (начальный этап).....	196
Разводовская Я.В. (Гродно, ГрГМУ) Обучение языку профессиональной коммуникации в области медицины: лингводидактический и прагматический аспекты.....	199
Родина Е.И., Кузьмина Т.В., Малашкина Т.К. (Минск, БГМУ) Национально-культурный компонент в преподавании русского как иностранного.....	203
Сентюрова Н.А. (Минск, БГУ) Реализация сознательно-практического метода в обучении польскоговорящих студентов русскому глагольному управлению.....	205
Серегина Н.А. (Минск, БГМУ), Елисеева О.Е. (Минск, БГУ) Использование сетевого словаря РКИ для студентов-медиков.....	208
Тихоненко Е.В., Лазовская А.И. (Минск, БГМУ) Формирование профессиональной компетенции иностранных студентов-медиков (на примере учебно-методического пособия «Русский язык как иностранный: модуль профессионального владения. Медико-биологический профиль»)...	212
Урбанович И.Г., Подберезская Ю.А. (Минск, МГЛУ) Обучение иностранных студентов морфологическим нормам в курсе «Русский язык и культура речи».....	215
Филюта С.В. (Минск, БГМУ) Использование корпусных технологий при создании общенаучного лексикона для студентов-медиков.....	218
 <i>Секция 3: ОБЩАЯ ТЕОРИЯ СИСТЕМ КАК МЕТОДОЛОГИЯ НАУКИ</i>	
Гецэвіч Ю.С., Рэянтэвіч І.В., Лысы С.І. (Минск, Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі) Аналіз памылак выходных даных інтэрнэт-сінтэзатара беларускага і рускага маўлення па тэксце.....	221
Гладких Р.В. (Минск, БГУ) Об одном из лексических соответствий в аспекте турецкой и русской лингвокультурологии: «Свет звезды».....	225
Головня А.И. (Минск, БГУ) Левизна/правизна и принципы симметрии/асимметрии при создании трехэлементных медицинских терминов..	227
Жук А.А. (Минск, БГУ) О системных аспектах дешифровки древнетюркского рунического письма (в связи с вопросом о докириллической письменности).....	238
Ивашенко В.П. (Минск, БГУИР) Грамматика. Гармония. Геометрия.....	242
Ивашенко В.П. (Минск, БГУИР) От теоретико-множественных моделей к симплициальным моделям языков.....	248
Качан О.Г. (Минск, БГПУ) К проблеме системного описания функционально-семантических категорий (ФСК).....	254

Качан Я.С., Кривальцэвіч А.В., Гецэвіч Ю.С. (Мінск, НАН РБ) Збор мноства маўленчых фраз для тэставання сістэм сінтэзу і распазнавання маўлення па семантычных класах «плошча», «аб’ём».....	258
Станіславенка Г.Р., Лысы С.І., Гецэвіч Ю.С. (Мінск, НАН РБ) Рэдагаванне электронных масіваў тэкстаў на беларускай мове з выкарыстаннем камп’ютарна-лінгвістычных сэрвісаў платформы www.corpus.by.....	262
Станкевич А.Ю. (Гродно, ГрГУ) Применение языка запросов яндексе для поиска белорусскоязычных контекстов к заданному ЛСВ слова.....	268

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

Березюк Д.В. (Минск, МГЛУ) Понятие «тональность» сквозь призму понятия «оценка».....	273
Ван Цин (Китай — Минск, БГУ) Этапы становления частеречной систе- мы китайского языка.....	277
Головня А.И. (Минск, БГУ) Симметрично-асимметричные аспекты при образовании одноэлементных медицинских терминов.....	280
Грибач Т.А. (Гомель, УО БелГУТ) К проблеме формализации транс- портных документов.....	286
Гусева Н.Ю., Хайрова А.К. (Минск, БГУ) Роль учебного центра ин- формационных технологий в процессе обучения на филологическом факультете.....	289
Здоронюк Ю.А. (Минск, БНТУ) Технология клонирования интонации... ..	291
Зубова И.И. (Минск, МГЛУ) Семиотические аспекты идентификации личности в виртуальной коммуникации.....	296
Игнатович Л.В. (Минск, БГМУ) Глоссарий терминов и сокращений в сфере промышленного производства лекарственных средств.....	301
Игнатович Т.В. (Минск, БГУ) Словарь речеведческих терминов для курса русского языка в учреждениях общего среднего образования РБ....	305
Кудина Н.Л. (Минск, БГМУ) Рефлексивная деятельность как неотъемле- мая часть познания языковой личности.....	309
Любецкая К.П. (Мінск, БДУ) Беларуская навукова-тэхнічная лексіка XX ст.....	314
Маракулина П.А. (Минск, МГЛУ) Языковые маркеры конфликтности..	317
Матальга С.А. (Минск, БГУИР) Электронные ресурсы как средство повышения эффективности и качества иноязычной подготовки.....	321
Метлицкая Н.А. (Минск, МГЛУ) Психолингвистические аспекты проблемы порождения текстов.....	324

ЗБОР МНОСТВА МАЎЛЕНЧЫХ ФРАЗ ДЛЯ ТЭСТАВАННЯ СІСТЭМ СІНТЭЗУ І РАСПАЗНАВАННЯ МАЎЛЕННЯ ПА СЕМАНТЫЧНЫХ КЛАСАХ «ПЛОШЧА», «АБ'ЁМ»

Сінтэзатар маўлення па тэксце — гэта сістэма, здольная генераваць маўленне па тэксце, прадстаўленым у пісьмовай форме. Пры дапамозе тэхналогіі сінтэзу маўлення электроннае прыстасаванне (камп'ютар, тэлефон і інш.) можа “чытаць” тэкст (электронную кнігу, sms-паведамленне і інш.) голасам, набліжаным да натуральнага. Дзеля дасягнення гэтай мэты сістэма павінна выконваць мноства розных працэсаў: расстаноўку націскаў, сэнсавое акцэнталагічнае выдзяленне вызначаных слоў, дзяленне сказаў на сэнсавыя і інтанацыйна непадзеленыя фрагменты (сінтагмы) і інш.

На сённяшні дзень падобныя сістэмы дастаткова якасна ажыццяўляюць усе пастаўленыя перад імі мэты. Аднак, не гледзячы на ўсе дасягненні ў гэтай галіне, застаюцца некаторыя праблемы, якія не ўдалося вырашыць цалкам. Для павышэння натуральнасці сінтэзаванага маўлення неабходна вырашыць комплекс задач, звязаных як з забеспячэннем натуральнасці голасу на ўзроўні тэмбру, плаўнасці гучання і інтанацыі, так і з правільнай расстаноўкай націскаў, расшыфроўкай скарачэнняў, лікаў, абрэвіятур і спецыяльных знакаў у арфаграфічных словах.

Каб камп'ютар быў здольны размаўляць, як чалавек, распрацоўшчыкам неабходна даследаваць натуральную мову і дакладна фармалізаваць усе тыя заканамернасці, па якіх у пэўнай мове ажыццяўляецца (у дадзеным выпадку) вымаўленне лічбаў і адзінак вымярэння. Натуральна, што такія даследаванні павінны выконваць лінгвіст. Таму аўтары дадзенага артыкула, азнаёміўшыся з даследаваннямі папярэдніх навукоўцаў, лічаць важным напрацаваць правераныя нармалізаваныя мноствы маўленчых фраз і сказаў на прыкладзе семантычных класаў «Плошча», «Аб'ём» для далейшага тэставання сістэм сінтэзу маўлення.

За сістэму, на якой праводзілася тэставанне, была абрана сістэма сінтэзу маўлення па тэксце, распрацаваная ў лабараторыі распазнавання і сінтэзу маўлення АПП НАН Беларусі, якая з'яўляецца анлайн-рэсурсам для агульнага выкарыстання. Сінтэзатар маўлення па тэксце (*Text-to-SpeechPHP-basedSynthesizer*) ажыццяўляе агучванне тэксту на беларускай і рускай мовах [1]. У адрозненне ад дастаткова якаснага сінтэзатара маўлення па тэксце для беларускай мовы *Multiphone* [2], у ім значна горш распрацаваны агульныя лінгвістычныя алгарытмы, адсутнічае апрацоўка іншых токенаў, апрача натуральнага тэксту і знакаў прыпынку. Мовазалежныя часткі вынесены ў базы дадзеных.

Лінгвістычны працэсар складаецца з наступных модуляў:

– перадапрацоўка тэксту (аналіз тэксту і выдзяленне ў ім слоў і знакаў прыпынку, расстаноўка націскаў паводле слоўніка);

– фанетызацыя пры дапамозе правіл (канвертаванне кожнай тэкставай адзінкі ў паслядоўнасць алафонаў).

Карэктная апрацоўка спецыфічных токенаў (лікаў, дат, скарачэнняў, адзінак вымярэння, абрэвіятур і інш.) на дадзены момант адсутнічае. Карыстальнік мае магчымасць бачыць прамежкавыя вынікі працы сінтэзатара (вынікі працы тэкставага, прасадычнага, фанетычнага і іншых працэсараў).

Адной з галоўных задач паляпшэння якасці працы сістэм сінтэзу маўлення па тэксце (СМТ) з’яўляецца апрацоўка складаных літарна-сімвальных канструкцый, асабліва колькасных выказаў у спалучэнні з адзінкамі вымярэння [3]. Колькасныя выразы з адзінкамі вымярэння (КВАВ) уяўляюць сабой форму адмыслова структураванай інфармацыі, якая выяўляецца спалучэннем колькаснага дэскрыптарна (пераданага на пісьме лінгвістычна з дапамогай слоў ці матэматычна праз лічбы) і пазначэнне адзінкі вымярэння (графічнага ці літарнага).

За аснову даследавання быў выбраны вучэбны дапаможнік «Геаграфія Беларусі» для дзясятага класа ўстаноў агульнай сярэдняй адукацыі з беларускай мовай навучання. Гэты падручнік з’яўляецца найлепшай крыніцай дадзеных, так як ахоплівае сферы жыцця (геаграфічнага, палітычнага, эканамічнага і сацыяльнага), якія ў сваю чаргу апісваюць галоўныя аспекты прадметнай вобласці праз КВАВ.

Сабраны матэрыял быў аформлены як спіс сказаў з ненармалізаванымі выразамі па наступных семантычных катэгорыях: адзінкі вымярэння (АВ) у спалучэнні з наступнымі колькаснымі выразамі (КВ), такімі як (табліца 1):

Табліца 1

Адзінкі вымярэння семантычных класаў «Плошча», «Аб’ём»

Адзінкі вымярэння плошчы	Адзінкі вымярэння аб’ёму
см ²	см ³
м ²	м ³
км ²	тыс. см ³ /м ³
тыс. м ²	млн. см ³ /м ³
м/м ²	млн. см ³ /м ³
тыс. км ²	мм ³

На першым этапе збору мноства маўленчых фраз па прадстаўленых семантычных класах былі выдзелены дадзеныя ў залежнасці ад АВ з прыкладамі семантычных падгруп і прыкладаў да іх (табліца 2). Усяго сабрана 370 сказаў, з якіх 48 датычацца да семантычнага класа «Плошча», і 36 — семантычнага класа «Аб’ём».

Табліца 2

Фрагмент першапачатковага спісу сабранага матэрыяла па семантычных катэгорыях «Плошча», «Аб’ём»

Семантычная катэгорыя	Адзінка вымярэння	Прыклад
Плошча	км ²	Сярод іх і самае буйное возера рэспублікі —Нарач,

		плошча якога складае $79,6\text{км}^2$.
Плошча	тыс.км^2	Заходняя Беларусь з плошчай 108 тыс. км^2 і насельніцтвам каля 4,6 млн чалавек адышла да Польшчы.
Аб'ём	м^3	У бронекамеры аб'ёмам 100 м^3 з тратылу і гексагену атрымліваюць звышцвёрды парашок у выглядзе алмазнага або карундавага пылу.
Аб'ём	тыс.м^3	Агульныя запасы перавышаюць 14 тыс. м^3
...

У дадзеным выпадку большую складанасць выклікаюць лікі, якія павінны быць канвертаваны ў лічэбнікі. У сучаснай беларускай і рускай мовах лічэбнік як часціна мовы мае наступныя лексіка-граматычныя разрады: колькасныя (адзін, два, тры), зборныя (чацвёрта, сямёра), парадкавыя (першы, другі, трэці). Часам да іх дадаюць яшчэ і дробавыя (восем цэлых і сем дзесятых, адна пятая). Колькасныя лічэбнікі, разам з дробавымі з'яўляюцца найбольш цікавымі ў межах дадзенага даследавання, паколькі яны найбольш прадуктыўна спалучаюцца ў скарачанай лічбавай форме з адзінкамі вымярэння (8 км^2 , $76,5 \text{ см}^3$, $28,4 \text{ тыс. м}^3$). Ніжэй прыведзены галоўныя правілы, паводле якіх адбывалася экспертная генерацыя лікаў у колькасныя і парадкавыя лічэбнікі:

1. Складаныя лічэбнікі (і колькасныя, і парадкавыя), якія складаюцца з дзвюх асноў, пішуцца разам (*дзевяцьсот, дзевяцісоты*).

2. Састаўныя лічэбнікі пішуцца асобна і маюць столькі слоў, колькі ў ліку ёсць лічбаў, не лічачы нулёў (*трыста сорок адзін*). Аднак парадкавыя лічэбнікі, якія скончваюцца на «-тысячны», «-мільённы», «-мільярдны», пішуцца разам (*стотысячны*).

3. Дробавыя лічэбнікі пішуцца асобна (*тры цэлых (і) адна другая*). Лічэбнікі «*паўтара*» ужываюцца ў м. і н. склонах (*паўтара метра*), лічэбнік «*паўтары*» — у ж. скл. (*паўтары тысячы*).

4. У састаўных колькасных лічэбніках скланяюцца ўсе словы, якія іх утвараюць (*трохсот пяцідзяці*). Пры скланенні дробавых лічэбнікаў таксама змяняюцца абедзве часткі (*трох чацвёртых*).

5. Пры скланенні састаўнога парадкавага лічэбніка змяняецца толькі канчатак апошняй састаўной часткі (*трыста пяцідзяты*).

6. Дробная частка ліка кіруе адзінкай вымярэння, якая заўсёды ўжываецца ў родным склоне і адзіночным ліку (*чатыры цэлыя дзве дзясятыя метра*).

7. Калі дробная частка складаецца з лічбаў ад 1 да 4, разрад ужываецца ў адзіночным ліку (*сорок тры цэлыя тры дзясятыя*). Калі гэта лікі ад 5 да 9, разрад ужываецца ў множным ліку (*сорок тры цэлыя восем дзясятых*).

8. Адзінкі вымярэння пасля дробных лічэбнікаў, якія азначаюць палову нечага, ужываюцца ў множным ліку (*пяць з паловай метраў*). Калі «палова» ідэнтыфікуецца як «*пяць дзясятых*», адзінка вымярэння ўжываецца ў адзіночным ліку (*пяць цэлых пяць дзясятых метра*).

Акрамя лікаў, якія нармалізуюцца ў колькасныя і парадкавыя лічэбнікі, у КВАВ апрацоўваюцца таксама адзінкі вымярэння. Кожная АВ мае сваю скарачаную і разгорнутую форму (м. — метр, см. — сантыметр і г.д.). Пасля нармалізацыі адзінак вымярэння ў разгорнутую форму вызначаецца склон, у якім неабходна ўжываць АВ. Адзінкі вымярэння нармалізуюцца ў залежнасці ад папярэдніх лічэбнікаў: у якім склоне ўжываецца лічэбнік, у тым і ўжываецца АВ. Ужыванне ўсіх лічэбнікаў залежыць ад папярэдняга прыназоўніка ці спалучэння слоў. Так, напрыклад, такія словы як *амаль, прыкладна, складае* і інш. патрабуюць пасля сябе лічэбнікі ў назоўным ці вінавальным склонах.

Такім чынам, пасля нармалізацыі лікаў у колькасныя і парадкавыя лічэбнікі, а скарачэнні ў поўныя назвы адзінак вымярэння, атрымліваецца спіс мноства тэкставых фраз семантычных класаў «Плошча», «Аб'ём» (табліца 3). На дадзены момант спіс неабходны для тэставання якасці працы сістэм сінтэзу маўлення ў звужанай тэматыцы — геаграфія, але ў далейшым ён будзе пакладзены за аснову вырашэння задачы аўтаматызаванага папаўнення базы дадзеных для любога тэматычнага дамена.

Табліца 3

Фрагмент спіса нармалізаваных выказаўсемантычных класаў
«Плошча», «Аб'ём» (усяго сабрана 370 запісаў)

Семантычная катэгорыя	Адзінка вымярэння	Прыклад з тэксту	Нармалізаваны прыклад тэксту
Плошча	км ²	Сярод іх і самае буйное возера рэспублікі — Нарач, плошча якога складае 79,6 км ² .	Сярод іх і самае буйное возера рэспублікі — Нарач, плошча якога складае семдзсятдзесяць цэлых шэсць дзясятых квадратнага кіламетра.
Плошча	тыс. км ²	Заходняя Беларусь з плошчай 108 тыс. км ² і насельніцтвам каля 4,6 млн чалавек адышла да Польшчы.	Заходняя Беларусь з плошчай сто восем тысяч квадратных кіламетраў і насельніцтвам каля чатырох цэлых шасці дзясятых мільёна чалавек адышла да Польшчы.
Аб'ём	м ³	У бронекамеры аб'ёмам 100 м ³ з тратылу і гексагену атрымліваюць звышцвёрды парашок у выглядзе алмазнага або карундавага пылу.	У бронекамеры аб'ёмам сто кубічных метраў з тратылу і гексагену атрымліваюць звышцвёрды парашок у выглядзе алмазнага або карундавага пылу.
Аб'ём	тыс. м ³	Агульныя запасы перавышаюць 14 тыс. м ³	Агульныя запасы перавышаюць чатырнаццаць тысяч кубічных метраў.
...

1. 1. Text-to-Speech PHP-Based Synthesizer [Электронны рэсурс]. — 2016. — Рэжым доступу: <http://www.corpus.by/tts3/>. — Дата доступу : 15.02.2016.
2. Гецевіч, Ю.С., Лобанов, Б.М. Система синтеза беларускай рэчы по тэксце // Рэчывыя тэхналогіі. — 2010. — № 1. — С. 91–100.
3. Гецевіч, Ю.С., Скопінава, А.М., Есіс, А.Ф. Мадэляванне і распрацоўка сістэм пошуку колькасных выказаў з адзінкамі вымярэння ў электронных тэксце на беларускай і рускай мовах // Развитие информатизации и государственной системы научно-технической информации (РИНТИ–2013): доклады XII Международной конференции. — Минск, 2013. — С. 282–287.