



# 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

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

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

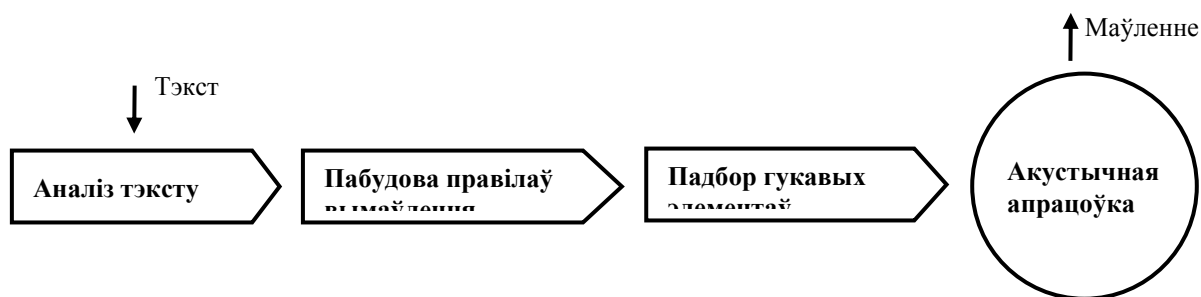
*Секция 4: ПРИКЛАДНАЯ ЛИНГВИСТИКА В БЕЛАРУСИ:  
СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ*

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

Ю.С. Гецэвіч, І.В. Рэянтковіч, С.І. Лысы  
(Мінск, Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі)

## АНАЛІЗ ПАМЫЛАК ВЫХОДНЫХ ДАНЫХ ІНТЭРНЭТ-СІНТЭЗАТАРА БЕЛАРУСКАГА І РУСКАГА МАЎЛЕННЯ ПА ТЭКСЦЕ

Сінтэз маўлення па тэксце з'яўляецца асобнай сферай апрацоўкі натуральнай мовы. Машына агучвае тэкст з правільна размечаным вымаўленнем, інтанацыяй, хуткасцю і г.д. [1]. Сінтэз вуснага маўлення— гэта пераўтварэнне загадзя невядомай тэкставай інфармацыі ў маўленне (малюнак 1) [2].



Мал. 1. Агульная схема працэсу сінтэзу маўлення па тэксце

Тэхналогія сінтэзу вуснага маўлення шырока прымяняецца для людзей, якія маюць праблемы са зрокам. Для ўсіх астатніх яна стварае новыя магчымасці для зручнага карыстання тэхнікай і значна зніжае нагрузку на зрок, на нервовую сістэму, дазваляе задзейнічаць слыхавую памяць.

Любы электронны тэкст складаецца са слоў, раздзеленых прабеламі і знакамі прыпынку. Вымаўленне слоў залежыць ад іх размяшчэння ў сказе, а інтанацыя фразы — ад знакаў прыпынку і, што вельмі важна, ад сэнсу слова. Для таго каб сінтэзаванае маўленне гучала натуральна, неабходна дакладна вырашыць цэлы комплекс задач, звязаных з забеспячэннем натуральнасці голасу на ўзроўні плаўнасці гучання і інтанацыі, з правільнай расстаноўкай націскаў, расшыфроўкай скарачэнняў, лічбаў, абрэвіятур і спецыяльных знакаў з улікам асаблівасцей граматыкі і беларускай, і рускай моў.

У Лабараторыі распазнавання і сінтэзу маўлення Аб'яднанага інстытута праблем інфарматыкі Нацыянальнай акадэміі навук Беларусі былі распрацаваны тры тыпы сінтэзатара беларускага і рускага маўлення па тэксце: сінтэзатар маўлення для стацыянарных платформ, сінтэзатар маўлення для мабільных платформ, сінтэзатар маўлення для Інтэрнэта [3]. Найлепшым (па сваёй функцыянальнай якасці) можна лічыць стацыянарны сінтэзатар, бо ў ім добра рэалізаваны блок тэкставай апрацоўкі. Распрацаваныя алгарытмы лінгвістычнай апрацоўкі лікаў у парадкавыя і колькасныя лічэбнікі і іх мадыфікацыі дазваляюць агучваць абазначэнні

розных тыпаў. Яны, у адрозненне ад існуючых, улічваюць словазмяненне парадкавых лічэбнікаў паводле роду, ліку, склону і дазваляюць павысіць натуральнасць сінтэзаванага маўлення.

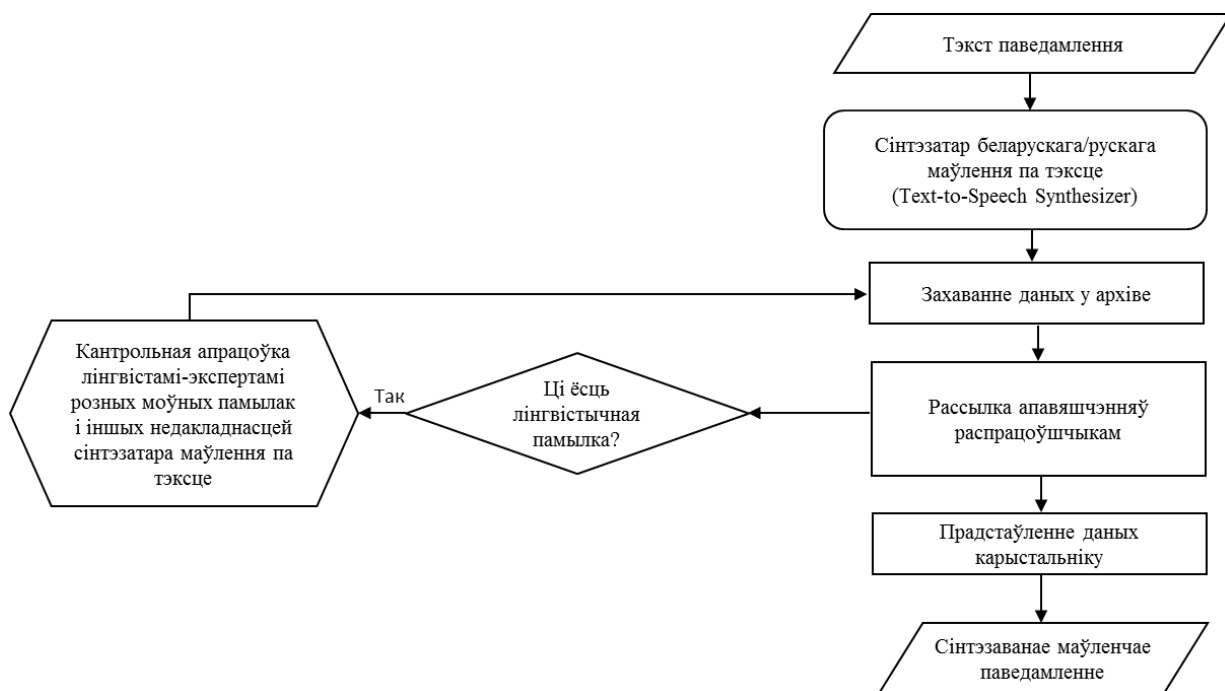
Сінтэзатар маўлення для мабільных платформ з распрацаваным метадам расстаноўкі націскаў слоў, які ўлічвае эўрыстычныя і статыстычныя характарыстыкі націску, можа выкарыстоўвацца на большасці мабільных тэлефонаў, не выключаючы і прыстасаванні на платформах IOS і Android.

Сістэма сінтэзу маўлення для Інтэрнэта рэалізаваная на скрыптавай мове праграмавання PHP. Карыстальнік можа ў любы час зайсці на сэрвіс Сінтэз маўлення па тэксце сайту [www.cogrus.by/tts3i](http://www.cogrus.by/tts3i) сінтэзаваць маўленчае паведамленне для любога свайго тэксту на адпаведнай мове. Сінтэзатар выкарыстоўвае слоўнікавую базу беларускай мовы (паводле публікацый «Слоўнік беларускай мовы», пад рэд. М.В. Бірылы, 1987; «Слоўнік беларускай мовы», навук. рэд. А.А. Лукашанец, В.П. Русак, 2012), рускай мовы (паводле публікацыі «Грамматический словарь русского языка: Словоизменение», А.А. Зализняк, 1980) і, акрамя гэтага, яшчэ адзін асобны слоўнік, што пастаянна папаўняецца новымі словамі з ужо прастаўленымі націскамі і вызначанымі часцінамі мовы ў залежнасці ад таго, калі карыстальнікі генерыруюць беларускае ці рускае маўленне.

Эфектыўнасць, якасць і дакладнасць працы сінтэзатара маўлення па тэксце залежыць ад аператыўнага рэагавання на ўзнікаючыя праблемы як з боку праграмістаў-распрацоўшчыкаў, так і экспертаў па апрацоўцы выходных даных — камп'ютарных лінгвістаў. У прыватнасці, праграмісты ўносяць праўкі непасрэдна ў праграмны код сінтэзатара ў адпаведнасці з рэкамендацыямі лінгвістаў.

Працэс працы з Інтэрнэт-сінтэзатарам маўлення па тэксце, а таксама экспертная апрацоўка яго выходных даных прадстаўлены ў выглядзе схемы на малюнку 2.

У табліцы 1 прыведзена апісанне найбольш сур'ёзных моўных памылак і недакладнасцей, якія сустракаюцца падчас апрацоўкі, з непасрэднымі прапановамі спосабаў іх выпраўлення і адпаведнымі прыкладамі, у табліцы 2 — агульная статыстыка па ўжо апрацаваных памылках. Для кожнай знойдзенай памылкі быў вызначаны ятып: лексічная (Л), пунктуацыйная (П), арфаэпічная (АЭ), сінтаксічная (С), акцэнталагічная (А), марфалагічная (М), арфаграфічная (АГ), словаўтваральная (СЛ), фразеалагічная (Ф), стылістычная (СТ) [4].



Малюнак 2. Схема працы з Інтэрнэт-сінтэзатарам маўлення па тэксце і экспертная апрацоўка выходных даных

Табліца 1

**Памылкі і недакладнасці, якія сустракаюцца ў выходных даных працэса генерацыі беларускага і рускага маўлення па тэксце**

Прыклад моўнай памылкі	Апісанне моўнай памылкі	Спосаб выпраўлення	Прыклад выпраўленай моўнай памылкі	Тып памылкі
<i>w пра=па=но=вы=_невядомаяЧасіц  wму=зы=чна=-па=э=ты=чна=e=_невядомаяЧасіц  wно=рве=жски=й_невядомаяЧасіц  wми=ни=-фи=льм_невядомаяЧасіц</i>	Слова з непрастаўленым націскам і/ці нявызначанай часцінай маўлення	Паставіць у слове пасля націскай галоснай знак “+” (адпавядае асноўнаму націску) або “=” (адпавядае дадатковаму націску для састаўнога слова), вызначыць часціну маўлення	<i>w прапано+вы_назоўнік  wмузы=чна-паэты+чнае_прыметнік  wнорве+жский_лічэбнік  wми=ни-фи+льм_назоўнік</i>	<u>А</u>
<i>wе+ўра-_невядомаяЧасіц  w 2_невядомаяЧасіц  w 0_невядомаяЧасіц  w 1_невядомаяЧасіц  wb_невядомаяЧасіц</i>	Сінтэзатар не здольны распазнаць і агучыць складанасастаўныя словы з ужываннем лічбаў (еўра-2016)	Неабходна ўнесці патрэбныя праўкі ў праграмны код, каб усе словы былі прагавораны поўнасьцю	<i>Пасля ўнясення правак у праграмны код:  w е+ўра-2016_назоўнік</i>	<u>Д</u>



Прыклад моўнай памылкі	Апісанне моўнай памылкі	Спосаб выпраўлення	Прыклад выпраўленай моўнай памылкі	Тып памылкі
<i>w a=дзн_невядомаяЧасц</i>	Нераспазнанае слова — “адзін”	Слова запісваецца правільна, пазначаецца часціна маўлення, прастаўляецца націск	<i>w адзі+н_лічэбнік</i>	<b>АГ</b>
<i>w да=ъа=по=шня=га=_невядомаяЧасц</i>	Знак “Ъ” у сінтэзатары служыць для аб’яднання двух слоў, што ў працэсе вымаўлення прагаворваюцца як адно, распазнаваўся ў сістэме як рускамоўны цвёрды знак	1. Словы запісваюцца правільна ў асобным радку, пазначаюцца часціны маўлення, прастаўляюцца націскі 2. Знак “Ъ” змяняецца на нейтральны, напрыклад “^” (цыркумфлекс)	<i>w да+_прыназ оўнік</i> <i>r ^ w апо+шняга_прыметнік</i>	<b>АЭ</b>
<i>w«_невядомаяЧасц</i> <i>w _невядомаяЧасц</i> <i>w/_невядомаяЧасц</i> <i>w‘_невядомаяЧасц</i> <i>w@_невядомаяЧасц</i> <i>w©_невядомаяЧасц</i> <i>w---_невядомаяЧасц</i> <i>w[_невядомаяЧасц</i> <i>w&lt;_невядомаяЧасц</i>	Недакладная ідэнтыфікацыя яўных і няяўных знакаў прыпынку, незнаёмых сімвалаў	Патрэбна замяніць параметр класіфікацыі “w” (словаформа)нераспазнанага знаку на патрэбны параметр “p” (паўза)	<i>r«</i> <i>r /</i> <i>r /</i> <i>r ‘</i> <i>r @</i> <i>r©</i> <i>r ---</i> <i>r [</i> <i>r&lt;</i>	<b>П</b>

Табліца 2

### Агульныя статыстычныя даныя па памылках Інтэрнэт-сінтэзатара беларускага і рускага маўлення па тэксце

Апісанне	Даныя
Колькасць апрацаваных інфармацыйных паведамленняў аб памылках (мэйлы)	каля 115
Згрупавана памылак	10 тыпаў
Прыкладная колькасць усіх непасрэдна выяўленых памылак:	каля 8500
• <i>паўзы (p)</i>	каля 160
• <i>выпраўленыя моўныя памылкі і недакладнасці</i>	каля 7500
• <i>нявыпраўленыя моўныя памылкі і недакладнасці</i>	каля 840

Неабходна адзначыць, што для апісання і прымянення якасных метадаў і спосабаў эфектыўнага вырашэння праблем генерацыі маўленчых паведамленняў па тэксце можна паспяхова выкарыстоўваць праграмны

лінгвістычны сродак NooJ [5] або іншыя праграмы аналізу і лінгвістычнай апрацоўкі тэкстаў [6].

Такім чынам, працэс пастаяннага дэталёвага аналізу моўных памылак у выходных даных сінтэзатара беларускага і рускага маўлення па тэксце, іх паслядоўнае выпраўленне праз унясенне правак у праграмны код і лінгвістычныя рэсурсы дае магчымасць значнага павышэння эфектыўнасці працы дадзенага сінтэзатара.

#### ЛІТАРАТУРА

1. Усё пра апрацоўку натуральнай мовы | Лабараторыя распазнавання і сінтэзу маўлення [Электронны рэсурс]. — 2016. — Рэжым доступу: <http://ssrlab.by/archives/2708>. — Дата доступу: 12.03.2016.

2. Синтез речи (TTS –Text-to-Speech) | Речевые Технологии [Электронный ресурс]. — 2013. — Режим доступа: <http://www.speetech.by/technologies/tts#fragment-1>. — Дата доступа: 12.03.2016.

3. Сінтэзатар беларускага і рускага маўлення па тэксце для ста цыянарных, мабільных і інтэрнэт платформ [Электронны рэсурс]. — 2016. — Рэжым доступу: [http://ssrlab.by/wp-content/uploads/2014/09/Belarusian-And-Russian-Speaking-TTS\\_-bel.pdf](http://ssrlab.by/wp-content/uploads/2014/09/Belarusian-And-Russian-Speaking-TTS_-bel.pdf). — Дата доступу: 13.03.2016.

4. Асіпчук, А.М. Маршэўская, В.В., Садоўская, А.С. Беларуская мова. Прафесійная лексіка [Текст]: дапаможнік па курсе «Сучасная беларуская мова» для студэнтаў нефілалагічных спецыяльнасцяў; Гродзенскі дзяржаўны ўніверсітэт імя Янкі Купалы. — Гродна, 2009.

5. NooJ: A Linguistic Development Environment [Electronic resource]. — 2016. — Mode of access: <http://www.nooj4nlp.net>. — Date of access: 19.03.2016.

6. Программы анализа и лингвистической обработки текстов [Электронный ресурс]. — 2013. — Режим доступа: <http://rvb.ru/soft/catalogue/c01.html>. — Дата доступа: 19.03.2016.