

КОНТРАСТИВНЫЕ ИССЛЕДОВАНИЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА

Материалы
Международной научной конференции
г. Минск, 29–30 октября 2014 г.

В двух частях
Часть вторая

**ТЕХНОЛОГИИ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ ИНТЕРНЕТА
И ЕЕ ОБРАБОТКА**

<i>Бочкова А. Л.</i> Автоматическое резюмирование мнений участников интернет-коммуникации об определенном объекте	3
<i>Грибач Т. А.</i> Технология Data Mining для поиска решений	7
<i>Дзубук Т. В.</i> Подходы к автоматическому извлечению фактов из неструктурированного текста	11
<i>Семак А. С.</i> Подходы к автоматическому определению субъективности в интернет-текстах	15
<i>Товпинец К. Г.</i> Основные трудности анализа тональности текстов	21
<i>Чалагаева В. В.</i> Языковая репрезентация категории тональности в новостных статьях англо- и русскоязычной прессы	25
<i>Bilan V., Bobkov A., Gafurov S., Krasnoproshin V., Laar J. van de, Vissia H.</i> <i>An Ontology-based Approach to Opinion Mining</i>	27
<i>Bobkov A., Gafurov S., Krasnoproshin V., Romanchik V., Vissia H.</i> <i>Information Extraction Based on Semantic Patterns</i>	33
<i>Russo Claudio.</i> Extracting linguistic Data from Usenet newsgroups: troubles and challenger	39

АВТОМАТИЧЕСКИЙ АНАЛИЗ И СИНТЕЗ ТЕКСТОВ

<i>Авдеева Н. А., Боярский К. К.</i> Обработка синтаксических групп с числительными в системе автоматического анализа текста SEMSIN.....	43
<i>Анисимович А. В.</i> Задача выделения слова в автоматической обработке текстов на китайском языке	47
<i>Детскина Р. В.</i> Сложные прилагательные в немецком языке и особенности их автоматического перевода на русский язык	50
<i>Елисеева О. Е.</i> Гипертекстовая модель сетевого учебного словаря иностранного языка	55
<i>Зубова И. И.</i> Морфология мифа как основа системы автоматического порождения мифологического текста	59
<i>Кошчанка У. А.</i> Актуальнасць стварэння лексіка-семантычнай анталогії тыпу WordNet для беларускай мовы.....	64
<i>Лысы С. І., Гецэвіч Ю. С.</i> Частотны анализ электроннага тэксту на прадмет выкарыстання слоў і іншых сімвалных паслядоўнасцяў з дапамогай www.corpus.by	68
<i>Метлицкая Н. А.</i> Формализованное представление основного статического содержания текста	72
<i>Нагорнович Э. В.</i> Способы автоматического исправления ошибок в письменном тексте	75

<i>Праблемы амографіі з дапамогай NOOJ для больш чым 50 амографаў рускай мовы</i>	79
<i>Романов Ю. В., Товмач Ю. В.</i> Оценка погрешности определения лингвостатистических параметров	83
<i>Чапля А. И., Чапля С. Г.</i> Опыт статистического разграничения омонимии и полисемии слов на базе триад	87
<i>Черепович А. В.</i> К проблеме использования нормативных словарей при проведении криминалистической экспертизы звукозаписей	88
<i>Шмыга Д. Л.</i> Формальная модель системы автоматической генерации текстов англоязычной контекстной рекламы	92
<i>Шуманская Т. А.</i> Параўнальны аналіз электронных слоўнікаў беларускай мовы	96
<i>Шустова Д. С.</i> Автоматическая экспертиза текстов англоязычных СМИ с элементами вербальной агрессии	100
<i>Яковинин В. С.</i> О методе алгоритмического определения грамматических классов слов	105

КОРПУСНЫЕ ИССЛЕДОВАНИЯ

<i>Беридзе М. М., Бакурадзе Л. Д.</i> Грузинский диалектный остров в Иране	108
<i>Беридзе М. М., Киквидзе З. З., Лорткипанидзе Л. Л.</i> Грузинский корпус метаязыка лингвистики: проблемы и решения	111
<i>Бускунбаева Л. А., Сиразитдинов З. А., Ишмухаметова А. Ш.</i> О внедрении башкирского языка в киберпространство	115
<i>Детскина Р. В., Василевская В. А., Василенко Е. С., Жданович А. Е., Филимонова Т. А.</i> Сопоставительный анализ лексики школьных учебников по белорусскому и русскому языкам	119
<i>Детскина Р. В., Шимчук Л. В.</i> Особенности формирования лексики компьютерного сленга	120
<i>Жубанов А. К., Фазылжанова А.</i> О концепции по созданию Национального корпуса казахского языка (НККЯ)	123
<i>Жубанов А. К.</i> Принципы автоматизации морфологической разметки текстов Национального корпуса казахского языка	126
<i>Зыгмантович Н. В., Ильина Е. И., Марковская Е. В., Шалимо Н. В.</i> Об основных задачах изучения лексики школьных учебников по белорусской и русской литературе	131
<i>Камишлова О. Н.</i> Контрактивные исследования в учебном корпусе текстов	132
<i>Кривошечя И. А.</i> Возможности параллельных корпусов текстов при переводе	135
<i>Лорткипанидзе Л. Л., Еремьян Р. А.</i> Разработка менеджера корпуса грузинских литературных текстов	138

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Минский государственный лингвистический университет

**КОНТРАСТИВНЫЕ ИССЛЕДОВАНИЯ
И
ПРИКЛАДНАЯ ЛИНГВИСТИКА**

МАТЕРИАЛЫ
Международной научной конференции
г. Минск, 29–30 октября 2014 г.

В двух частях
Часть вторая

Минск 2015

УДК 81'42 + 81'33

ББК 81.04 + 81.1

K65

Рекомендованы Редакционным советом Минского государственного лингвистического университета. Протокол № 3 (19) от 23.12.2014 г.

Редакционная коллегия: А. В. Зубов (*отв. редактор*), Т. П. Карпилович (*отв. редактор*), Л. Н. Беляева, М. Г. Богова, Р. В. Детскина, Н. Ю. Павловская, И. В. Совпель

Рецензенты: доктор филологических наук, профессор Д. Г. Богушевич (МГЛУ); кандидат филологических наук, доцент А. И. Головня (БГУ)

K65 Контрастивные исследования и прикладная лингвистика : материалы Междунар. науч. конф., Минск, 29–30 окт. 2014 г. В 2 ч. Ч. 2 / отв. ред.: А. В. Зубов, Т. П. Карпилович. – Минск : МГЛУ, 2015. – 227 с.

ISBN 978-985-460-669-9 (Ч. 2).

ISBN 978-985-460-658-3.

В издание, состоящее из двух частей, включены материалы, ракрывающие возможности контрастивного изучения самых различных языков: белорусского, русского, английского, немецкого, башкирского, туркменского, итальянского, китайского, грузинского, казахского, польского. Показано как результаты такого изучения используются при обучении, автоматической обработке текстов, их переводе.

Данная вторая часть книги содержит тексты выступлений участников конференции в секциях «Технологии извлечения информации из Интернета и ее обработка», «Автоматический анализ и синтез текстов», «Корпусные исследования», «Контрастивная лингвистика и вопросы преподавания языков» и «Особенности речевого поведения в разных культурах».

Для научных работников, аспирантов и магистрантов.

УДК 81'42 + 81'33

ББК 81.04 + 81.1

ISBN 978-985-460-669-9 (Ч. 2)
ISBN 978-985-460-658-3

© УО «Минский государственный
лингвистический университет», 2015

Подсказка работает в интерактивном режиме и является консервативным способом коррекции запроса (показывается вместе с выдачей по оригинальному запросу и не искажает результатов поиска). Недостатком подсказки является необходимость совершать дополнительные действия (кликать по ссылке). В случае автозамены пользователю показывается поисковая выдача сразу по исправленному запросу. Под поисковой строкой помещается сообщение об автоисправлении (*В запросе <...> была исправлена опечатка*). Функция работает в автоматическом режиме и экономит время, которое пришлось бы потратить на ручное исправление ошибки или клик на подсказку. На случай неправильного автоисправления предусмотрена ссылка на страницу с результатами поиска по оригинальному написанию. Следует заметить, что каждое неверное автоматическое изменение запроса сильно раздражает пользователя, поэтому требование к точности автозамен намного выше, чем к точности подсказок.

Автоматическое исправление ошибок работает на основе трехуровневой модели: 1) для каждого слова спллчкер генерирует отранжированный список подсказок; 2) классификатор надежности подсказки определяет, является ли слово ошибкой; 3) классификатор надежности автоисправлений решает, можно ли исправить ошибку автоматически. Важно заметить, что исправление ошибок в запросах и текстах имеют ряд существенных отличий: 1) частота опечаток и степень искажения запросов гораздо выше; 2) значительную долю слов в запросах составляют новые слова с неустоявшимся написанием; 3) запросы в среднем состоят из 2–3 слов, поэтому для их исправления невозможно использовать широкий контекст.

Исходя из данных корпорации Yandex [5], можно отметить, что большая часть ошибочных запросов (83,6 %) содержит всего одну ошибку. Для того чтобы проверить предположение о том, что большая часть словарных ошибок в запросах являются тривиальными (имеют однозначное исправление независимо от контекста), корпорация Yandex попыталась оценить влияние словарного окружения на исправление ошибок и провела следующий эксперимент [5]. Были взяты 620 запросов со словарными ошибками, из них выделены все пары «слово–исправление» (всего 754), отфильтрованы пары-орфоварианты¹ и пары-словоформы². Из оставшихся 714 пар были взяты слова, в которых аналитик должен был исправить ошибки, не учитывая контекст запроса. Полученные результаты сравнили с исправлениями системы, сделанными с учетом контекста запроса. Если исправления совпадали, пара «ошибка–исправление» считалась контекстуально независимой (например, аналитик: *рюкз* – *рюкзак*; система: *рюкз* – *рюкзак*). Если исправления не совпадали, пара «ошибка–исправление» считалась контекстуально зависимой (например, аналитик: *крепк* – *скрепка*; система: *крепк* – *крепко*). Контекстуально независимых исправлений оказалось 74 %, что подтвердило

¹ Орфоварианты – слова, имеющие одно и то же значение, но различное, как правило, очень близкое, написание (например, *кэйтлин* – *кейтлин*).

² Неправильная словоформа – слово с опечаткой и его исправление являются разными формами одного и того же слова (например, *лето* – *летом*).

предположение об однозначности изменений для большинства словарных ошибок. Примеры исправлений тривиальных ошибок: *актрисса* – *актриса*, *спаисбо* – *спасибо*, *букенестический* – *буканистический*. Среди контекстно-зависимых исправлений большую часть составили ошибки в коротких словах, допускающие разные, зависящие от контекста, исправления. Например: *сво законов* – *свод законов* и *сво игра* – *своя игра*.

Таким образом, простой эксперимент показал, что для большей части словарных ошибок (74 %) исправления в контексте и без контекста совпадают. Это позволяет провести оценку доли словарных ошибок, которые можно корректировать без учета контекста.

ЛИТЕРАТУРА

1. Андреев, А. Автоматизация обнаружения и исправления опечаток в названиях географических объектов для системы семантического контроля документов электронной библиотеки [Электронный ресурс] / А. Андреев // Электронные библиотеки : перспективные методы и технологии, электронные коллекции. – 2007. – Режим доступа : http://www.rcdl2007.pereslavl.ru/papers/paper_25_v1.pdf. – Дата доступа : 20.04.2014.
2. Большаков, И. Проблемы автоматической коррекции текстов на флексивных языках [Электронный ресурс] / И. Большаков // Итоги науки и техники. – 1988. – Режим доступа : http://www.mathnet.ru/links/6daf35dc68ae75956feef114_72f273dc/intv77.pdf. – Дата доступа : 20.01.2014.
3. Гниловская, Л. Автоматическая коррекция орфографических ошибок [Электронный ресурс] / Л. Гниловская // Электронная библиотека. – 2004. – Режим доступа : <http://dspace.npuv.gov.ua/xmlui/bitstream/handle/123456789/6692/25Gnilovskaya.pdf?sequence=1>. – Дата доступа : 14.04.2014.
4. Кузнецова, А. И. Словарь морфем русского языка / А. И. Кузнецова, Т. Ф. Ефремова. – М. : Рус. язык, 1986. – 1 134 с.
5. Рудаков, И. Распознавание текстового изображения с учетом морфологии слова [Электронный ресурс] / И. Рудаков // Электронное научно-техническое издание. – 2012. – Режим доступа : <http://technomag.bmstu.ru/file/505204.html?&s=1>. – Дата доступа : 12.02.2014.
6. Панина, М. Автоматическое исправление опечаток в поисковых запросах без учета контекста [Электронный ресурс] / М. Панина // Диалог : Междунар. конф. по компьютерной лингвистике. – 2013. – Режим доступа : <http://www.dialog-21.ru/digests/dialog2013-materials/pdf/PaninaMF.pdf>. – Дата доступа : 11.04.2014.

Т. И. Окрут, Д. А. Пакладок,
Ю. С. Гецэвіч, Б. М. Лабанаў (Мінск)

ВЫРАШЭННЕ ПРАБЛЕМЫ АМАГРАФІ З ДАПАМОГАЙ НООЈ ДЛЯ БОЛЬШ ЧЫМ 50 АМОГРАФАЎ РУСКАЙ МОВЫ

Лінгвістычнае асяроддзе распрацоўкі НооЈ дазваляе вырашаць розныя задачы на марфалагічным, сінтаксічным і семантычным узроўнях з дапамогай стварэння спецыялізаваных слоўнікаў і граматык (алгарытмай) і іх

наступнай апрабацыі пры работе з тэкставымі корпусамі [1; 2]. У якасці адной з такіх задач можна разглядаць вырашэнне праблемы амографіі на кантэктна-граматычным узоруны.

Вядома, што амографы – гэта слова, якія маюць адноўлькае напісанне, але вымаўляюцца па-разнаму. Яны часцей за ўсё маюць розніцу ў пазіцыі націску або ў напісанні літары «ё» без кропак. Выдзяляюць наступныя класы амографаў: лексемы адной часціны мовы, розныя формы адной лексемы, лексемы розных часцін мовы, варыянты адной лексемы [3]. Пры гэтым у дадзеных групах амографаў на аснове падабенства па граматычных і сінтаксічных параметрах можна выдзеліць падгрупы амографічных пар, элементы якіх адлюстроўваюцца як мінімум адной граматычнай паметай і ад яе залежыць адназначная пазіцыя націску ў слове. Напрыклад, да іх належаньці такія пары, як ГÓДА (адз. лік) – ГÓДÁ (мн. лік), ОЗÉРА (адз. лік) – ОЗÉРА (мн. лік).

Такім чынам, для вырашэння праблемы амографіі з дапамогай NooJ былі абраныя 58 амографаў рускай мовы группы «Розныя варыянты адной лексемы». У іх пазіцыі націску адназначна залежыць ад граматычных памет. Для назоўнікаў мужчынскага або някага роду ў форме роднага склону адзіночнага ліку націск падае на больші блізкія склады да пачатку слова, чым для назоўнікаў у форме назоўнага або вінавальнага склону множнага ліку.

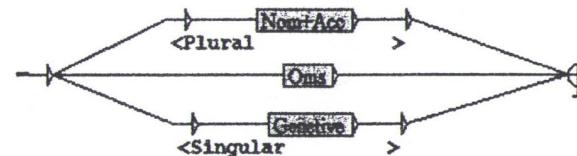
Спачатку на аснове частотнага спіса словаўжыванняў у «Нацыянальным корпусе рускай мовы» (НКРМ) са знятай аманіміяй (каля 6 млн. словаформаў) [4] быў пабудаваны граматычны слоўнік у фармаце NooJ. Для гэтага спатрэбілася адаптаваць граматычныя анататы НКРМ да анататы NooJ і запісаць усе словаформы ў фармаце «словаформа, ЧАСЦІНА МОВЫ+Характарыстыка граматычнай катэгорыі+...». Ніжэй прадстаўлены атрыманы фрагмент граматычнага слоўніка (мал. 1).

пицдали, VERB+Imperfective+Intransitive+Active+Plural+Past+Indicative
пицаль, NOUN+Feminine+Inanimate+Singular+Nominative
пицаль, NOUN+Feminine+Inanimate+Singular+Instrumental
пицат, VERB+Imperfective+Intransitive+Active+Plural+Present+3+Indicative
пицать, VERB+Imperfective+Intransitive+Active+Infinitive
пицце, NOUN+Feminine+Inanimate+Singular+Dative
пицце, NOUN+Feminine+Inanimate+Singular+Prepositional
пиццеварение, NOUN+Neuter+Inanimate+Singular+Nominative

Мал. 1. Фрагмент слоўніка, атрыманага на аснове частотнага спіса словаўжыванняў асноўнага корпуса НКРМ са знятай аманіміяй у фармаце NooJ

Далей на аснове кантэктуальна-граматычнага аналізу акружэння амографаў была распрацавана ў праграме NooJ сінтаксічная граматыка Oms_GO+DA+.nog (мал. 2). Яна ўключае ў сябе 3 ланцужкі пераходаў, два з якіх фармалізуецца правілы для 2-х варыянтаў амографаў і адзін прадугледжвае пошук не ўлічаных у правілах выпадкаў (непасрэдны спіс амографаў, прадстаўлены на мал. 3). Падграф з назвай «Nom+Acc» генерыруе маркер <Plural>, а падграф з назвай «Genetive» – < Singular>. Калі канкрэтны

падграф спрацоўвае, то слова-амограф атрымлівае адзін з маркераў, які адназначна вызначае граматыку слова, што ў нашым выпадку дае – «энтэпіческая амографія». Напрыклад, калі адназначна вызначыць лік слова «АДРЕСА», то адназначна будзе вызначаны і яго націск. Для адзіночнага ліку гэта будзе АДРЕСА, а для множнага – АДРЕСА.



Мал. 2. Галоўны граф граматыкі Oms_GO+DA+.nog

ДЕРЕВЦА	АДРЕСА	РОГА	НОМЕРА
ЖЕМЧУГА	БЕГА	СЕРДЦА	ОБЛАКА
ЗЕРКАЛА	БЕРЕГА	СЛОВА	ОЗЕРА
КОЛОКОЛА	БОКА	СНЕГА	ОРДENA
КОРМА	БОРТА	СОРТА	ОРДЕРА
КОРПУСА	ВНЕРА	СЧЕТА	ОСТРОВА
КРАЯ	ВЕКА	ТЕЛА	ОТПУСКА
КУЗОВА	ВЕТРА	ТОМА	ПАРУСА
ЛАГЕРЯ	ВЕЧЕРА	ТОПОЛЯ	ПАСПОРТА
ЛЕСА	ВОЙСКА	ТОРМОЗА	ПОЕЗДА
МАСЛА	ГЛАЗА	ХЛЕБА	ПОКРОВА
МЕСТА	ГОДА	ХОЛОДА	ПОЛЮСА
МЕХА	ТОНА	ЦВЕТА	ПОЛЯ
МОРЯ	ГОЛОСА	ЦЕХА	ПОЯСА
	ГОРОДА	ЧЕРЕПА	

Мал. 3. Падграф граматыкі Oms_GO+DA+.nog

Прыведзём для азнямлення некаторыя правілы вырашэння праблемы амографіі для абранных амографаў (былі ў тым ліку ўключаныя правілы, сформуляваныя ў працы [5]).

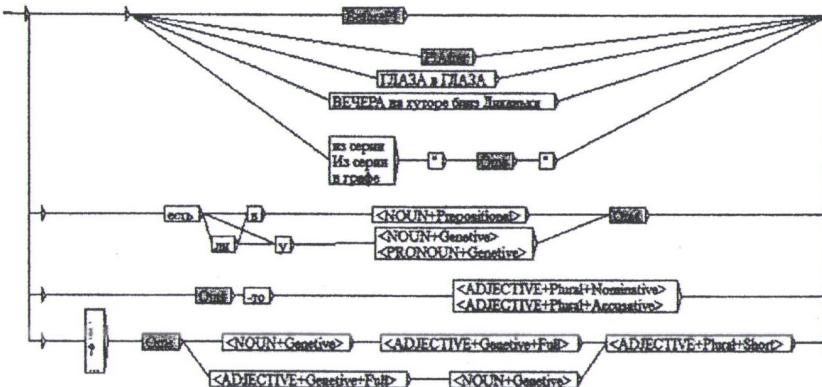
Разгледзім правілы кантэктуальна-залежнасці для назоўнікаў множнага ліку ў форме назоўнага або вінавальнага склону (мал. 4). Амографу прыпісваецца множны лік, калі спрацоўвае адно з наступных 5 правіл (усяго іх больш за 15):

- 1) пасля назоўніка або перад ім стаіць дзеяслоў, прыметнік, кароткі прыметнік, назоўнік, прыналежны або зваротны займеннік у форме множнага ліку;
- 2) назоўнік з'яўляецца аднародным членам сказа;
- 3) дзеепрыметнік множнага ліку ў вінавальнім або назоўным склоне, ці перашнягі або прошлагага часу;

4) пераходныя дзеясловы ў загадным або абвесным ладзе, або ў форме інфінітыва;

5) перад назоўнікам стаяць пераходныя дзеепрыметнікі і дзеепрыслоўі

Nom+Aj



Мал. 4. Граф Nom+Acc граматыкі Oms_GO+DA+.nog

Разгледзім правілы кантэкстуальнай залежнасці для назоўнікаў мужчынскага обо сярэдняга роду ў форме роднага склону адзіночнага ліку. Амографу прыпісваецца адзіночны лік, калі спрацоўвае адно з наступных 6 правіл (усяго іх больш за 20). Такім чынам, перад назоўнікам стаіць:

- 1) прыметнік мужчынскага або ніякага роду ў форме роднага склона адзіночнага ліку;
 - 2) займеннік-прыметнік мужчынскага або ніякага роду ў форме роднага склону адзіночнага ліку (*твоего, вашего*);
 - 3) колькасны лічэбнік;
 - 4) парадкавы лічэбнік у родным склоне;
 - 5) лічэбнік-прыметнік мужчынскага або ніякага роду адзіночнага ліку ў родным склоне;
 - 6) прыназоўнік, які кіруе родным склонам (*вокруг, до, из-за, у, от, из, около, среди, в течение, в продолжение, без, для, ради, после, с, со, внутрь, напротив, против, вдоль, вне, вместо, кроме, близ, относительно*).

Дадзеная граматыка была выкарбыстана для тэкставага корпуса памерам 1 млн словаформаў, які таксама быў выдадзены аўтарам прадстаўнікам НКРМ. Корпус змяшчае тэксты розных жанраў: публіцыстычныя, мастацкія, афіцыйна-справавыя, вучбна-навуковыя і рэкламныя. У ім налічваецца 2 967 абраных амографаў, з іх 2 017 – назоўнікі мужчынскага або ніякага роду ў форме роднага склону адзіночнага ліку з націскам, 950 – назоўнікі у форме назоўнага або вінавальнага склону множнага ліку.

Пасля праверкі лінгвістам-экспертам вынікаў апрацоўкі тэкстаў былі выяўлена, што гэтая граматыка вызначыла правільны варыянт націску

у амографе для 2 756 слоў-амографаў (больш за 92 %). Было налічана 211 памылак: 49 – памылкова выбраныя як элементы першай пазіцыі у амаграфічных парах; 6 – памылкова выбраныя як элементы другой пазіцыі у амаграфічных парах; 156 – навызначаныя граматыкай.

Асноўнымі прычынамі памылак сталі складаныя выпадкі, калі для вызначэння правільнага элемента амографічнай пары патрабуецца аналіз ўцаленага кантэксту, напрыклад як у сказе *МЕСТА в своих отделах Аверину по старой дружбе предложили знакомые по институту Борисевич и Багаев*, дзе вызначальнае для амографа слова заходзіцца на адлегласці восьмі слоў.

Далей аўтары плануюць дапрацаўваць граматыку і правесці яе тэсці-праванне на іншых тэкстах. Цяпер таксама адбываецца фарміраванне правілаў, і мэтай вырашэння праблемы амаграфіі для іншых амографаў, у першую чаргу для самай частотнай амаграфічнай пары ВСЕ – ВСЁ.

ЛІТЕРАТУРА

- Лінгвістычны працэсар NooJ [Электронны рэсурс]. – 2002. – Рэжым звароту : <http://www.nooj4nlp.net/pages/nooj.html>. – Дата звароту : 01.10.2013.
 - Гецэвіч, Ю. С. Метад пабудовы кампанентаў сінтэзу маўлення па тэксе для натуральна-маўленчага інтэрфейса пры дапамозе NooJ / Ю. С. Гецэвіч, А. М. Скопінава, Т. І. Окрут // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS–2014) : материалы IV Междунар. науч.-техн. конф., Минск, 20–22 февраля 2014 г. / редкол. : В. В. Голенков (отв. ред.) [и др.]. – Минск : БГУИР, 2014. – С. 333–338.
 - Венцов, А. В. Словарь омографов русского языка / А. В. Венцов [и др.]. – СПб. : Филологический факультет СПбГУ, 2004.
 - Нацыянальны корпус рускай мовы [Электронны рэсурс]. – Рэжым звароту : <http://www.ruscorpora.ru/>. – Дата звароту : 01.10.2013.
 - Цирульник, Л. И. Статистический анализ и контекстуальные правила разрешения графической омонимии при синтезе речи по тексту / Л. И. Цирульник, С. Г. Барбук, Б. М. Лобанов // Труды Междунар. конф. «Компьютерная лингвистика и интеллектуальные технологии» (Диалог'2009), Москва, 27–31 мая 2009 г. – С. 530–536.

Ю. В. Романов, Ю. В. Товмач (Санкт-Петербург, Россия)

ОЦЕНКА ПОГРЕШНОСТИ ОПРЕДЕЛЕНИЯ ЛИНГВОСТАТИСТИЧЕСКИХ ПАРАМЕТРОВ

Лингвостатистический анализ речи (письменной и устной) широко применяется в современной лингвистике для определения (в том числе) психического состояния автора текста [1]. Давно признана эффективность использования для этих целей закона Ципфа–Мандельбрата, отражающего линг-частотную зависимость параметров словоформ в частотном списке, полученном по определенному тексту и чутко реагирующего на изменения