

Министерство образования Республики Беларусь

Учреждение образования  
«Гомельский государственный университет  
имени Франциска Скорины»

Научно-методический Центр русистики

## **КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА:**

### **НАУЧНОЕ НАПРАВЛЕНИЕ И УЧЕБНАЯ ДИСЦИПЛИНА**

**Сборник научных статей**

Основан в 2010 году

**Выпуск 2**

Гомель  
ГГУ им. Ф. Скорины  
2012

УДК 81'322(082)  
ББК 81.1р3  
К 637

**Редколлегия:**

В. И. Коваль (ответств. ред.),  
Н. И. Шабулдаева, А. Ю. Крохмальник

**Рецензенты:**

доктор филологических наук, профессор А. В. Никитевич;  
кандидат филологических наук, доцент П. Е. Ахраменко

**Компьютерная лингвистика:** научное направление и учебная дисциплина : сборник научных статей. Вып. 2 / В. И. Коваль (ответств. ред.) [и др.]; М-во образования РБ, Гомельский гос. ун-т им. Ф. Скорины. – Гомель: ГГУ им. Ф. Скорины, 2012. – 136 с.

ISBN

В сборник включены научные статьи по актуальным вопросам компьютерной лингвистики: корпусная лингвистика, машинный перевод, программное обеспечение лингвистических исследований, прикладная лексикография, использование компьютерных технологий в учебно-методической деятельности.

Для преподавателей вузов, научных работников, аспирантов и студентов-филологов, специалистов в области компьютерных технологий.

**УДК 81'322(082)**  
**ББК 81.1р3**

ISBN

© УО «Гомельский государственный университет им. Ф. Скорины», 2012

## СОДЕРЖАНИЕ

<i>Барковіч А. А.</i> Магчымасці корпусных даследаванняў спалучальнасці адзінак беларускай мовы .....	5
<i>Баркович А. А., Ван Цин</i> Изучение лексического значения многозначных слов на базе лингвистических корпусов .....	8
<i>Важнік С. А.</i> Сабака, малпа ці слімак: да пытання аб трох стыхіях байнэта .....	13
<i>Волчек О. А., Порицкий В. В.</i> Проект корпуса белорусскоязычной периодики и художественной прозы .....	16
<i>Гліннік І. В.</i> Канструкцыі з дзеясловамі <i>быць</i> і <i>мець</i> ў арыгінальных і перакладных тэкстах (на матэрыяле паралельнага беларуска-рускага корпуса) .....	19
<i>Головня А. И.</i> Компьютерные технологии при исследовании и преподавании грамматической омонимии .....	24
<i>Голяк Ю. Д.</i> Концепция разработки электронного словаря для семантических сетей и паутин в компьютерных системах (на базе русского языка) .....	31
<i>Грибач Т. А.</i> Системы машинного перевода, основанные на параллельных корпусах текстов .....	35
<i>Елисеева О. Е.</i> Электронные видеокурсы в преподавании информационных технологий студентам филологических специальностей и повышении квалификации преподавателей .....	42
<i>Зубов А. В.</i> О возможностях белорусско-иноязычных параллельных корпусов текстов .....	46
<i>Коваль В. И.</i> Роль интернет-ресурсов в исследовании фразеологии (на материале устойчивого словосочетания <i>перемывать косточки</i> ) ...	49
<i>Корсак М. А.</i> Структурный анализ зоодемотиваторов (на примере интернет-мема <i>Сова арт-студент</i> , или *OvO* <i>Art Student Owl</i> ) .....	55
<i>Крохмальник А. Ю.</i> Связь классической и академической «тошноты» и «водянистости» текста с его функциональным речевым стилем .....	60
<i>Кураш С. Б.</i> Фрагмент корпусного исследования концепта «метафора»: метафора и её антиподы .....	63
<i>Лашкевич А. С.</i> О субституции компонентов в статусных единицах.....	68
<i>Лашукевич С. А.</i> Критерии и принципы сегментации турецкого предложения в системе машинного перевода деловой документации .	71
<i>Малинина Ю. А.</i> Фреймовое моделирование как способ представления знаний (на материале туристического дискурса) .....	75
<i>Половец М. В.</i> Структура глоссария терминов компьютерной лингводидактики .....	77
<i>Рогалев А. Ф., Рогалев В. А.</i> Материалы сайта «белорусская ономастика и эзотерическая лингвистика» ( <a href="http://bel-onomastika.ucoz.ru">bel-onomastika.ucoz.ru</a> ) в учебной работе .....	80

<i>Рычкова Л. В.</i> Лингвистические корпуса в социокультурном и эколоингвистическом аспектах .....	84
<i>Рычкова Л. В., Станкевич А. Ю., Бодак Ю. А.</i> Принципы создания газетного корпуса СМИ Гродненщины .....	88
<i>Сивакова Н. А.</i> Использование интернет-ресурсов при обучении русскому языку как иностранному .....	92
<i>Сичинава Д. В.</i> Русская конструкция с <i>было</i> и белорусский плюсквамперфект: сопоставительный анализ на материале белорусско-русского и русско-белорусского параллельных корпусов .....	95
<i>Солахаў А. В.</i> Словаўтваральныя і семантычныя асаблівасці новай лексікі ў мове беларускіх вэб-тэкстаў .....	104
<i>Станкевич А. Ю.</i> Использование строковых функций при работе с лингвистической картотекой MS Access .....	108
<i>Трохова Т. А., Холякко Е. И.</i> Экспертная система обучения старославянскому языку .....	113
<i>Хазанова К. Л.</i> Аб выкарыстанні электроннага тэсціравання пры выкладанні лінгвістычных дысцыплін .....	117
<i>Хакимов Б. Э.</i> Проблемы грамматической аннотации в диалектных корпусах текстов (на примере мишарского диалекта татарского языка) .....	121
<i>Шабулдаева Н. И.</i> Языковая норма и национальный корпус русского языка .....	124
<i>Шукайлова А. Г.</i> Использование интернет-ресурсов в процессе преподавания китайского языка в гимназии .....	127
<i>Юдкин-Рипун И. Н.</i> Задачи реферирования драмы для построения корпусного тезауруса перифрастических выражений .....	130
<i>Ядченко Е. И.</i> Потенциал компьютерного программного обеспечения MaxQDA в лингвистическом исследовании .....	134
<i>Асенчик Е. Ф.</i> Использование компьютерных технологий в комплексном анализе лексических единиц .....	138
<i>Иванов Е. Е., Старосельский В. А.</i> Учебная автоматизированная система лексико-грамматического анализа текста (при изучении английского языка на начальном этапе) .....	142
<i>Иванова С. Ф.</i> Трансфармацыі ў маўленні крылатых выразаў з рускамоўных літаратурных крыніц (на матэрыяле антыцытат з тэкстаў Байнэта) .....	145

## МАГЧЫМАСЦІ КОРПУСНЫХ ДАСЛЕДАВАННЯЎ СПАЛУЧАЛЬНАСЦІ АДЗІНАК БЕЛАРУСКАЙ МОВЫ

*У артыкуле асвятляюцца пытанні актуалізацыі магчымасцей корпуснай лінгвістыкі на матэрыяле беларускамоўных тэкстаў у кантэксце з'яўлення беларускіх корпусаў ў Інтэрнэце. Падкрэсліваецца неабходнасць вывучэння спалучальнасці моўных адзінак і іх валентнасці.*

Новыя магчымасці апошнім часам прадстаўляе для даследчыкаў з'яўленне ў Інтэрнэце беларускамоўных корпусных рэсурсаў: «Corpus Albaruthenicum» і Паралельнага беларуска-рускага корпуса ў складзе НКРЯ («Национальный корпус русского языка») [2; 3]. Калі першы з узгаданых корпусаў мае дачыненне да мовы навуковага стыля, то ў складзе НКРЯ прадстаўлены каля двух мільёнаў словаўжыванняў у выглядзе шэрагу твораў беларускай мастацкай літаратуры і некалькіх перакладаў на беларускую мову.

Выкарыстанне корпусаў тэкстаў дазваляе вывучаць пытанні рознай накіраванасці з адпаведнай матэрыялу дакладнасцю.

Безумоўны плюс названых корпусных праектаў – іх даступнасць у Інтэрнэце. Гэта ўласцівасць забяспечвае высокую ступень інфарматыўнасці дадзеных і максімальна шырокае кола іх патэнцыяльных карыстальнікаў. Немалаважныя рэальныя з дапамогай Сеціва перспектывы зваротнай сувязі паміж карыстальнікамі і складальнікамі корпусаў. Хаця інтэрфэйс «Corpus Albaruthenicum» вельмі лаканічны і мінімальна інфарматыўны (малюнак 1).

Тэкстаў: 74  
Абзацаў: 6724  
Сказаў: 18551  
Словаў: 350027

### Corpus Albaruthenicum

#### Пошук:

Слова:

усе формы

Дадаць слова

Ачысціць спіс словаў

Шукаць:  - У сказе  - У абзацы

Шукаць толькі у такой паслядоўнасці

Адлегласць паміж суседнімі словамі не больш:  (0 - любая, 1 - суседнія словы)

Шукаць

Адносна рэпрэзентатыўнасці ў корпуснай лінгвістыцы з сярэдзіны ХХ ст. існуе сваеасаблівы ўзор – Браўнаўскі. У гэтым сэнсе задача развіцця базы дадзеных –«Corpus Albaruthenicum» – надзвычай актуальная, бо «залаты стандарт» ў адзін мільён словаўжыванняў, па ўзору Браўнаўскага корпуса, гэтым рэсурсам не дасягнуты:

–Тэкстаў: 74

Абзацаў: 6724

Сказаў: 18551

Словаў: 350027” – арфаграфія аўтараў корпуса.

НКРЯ прыстасаваны для лінгвістычных даследаванняў у значна большай ступені. Яго аб’ём у некалькі разоў перавышае памеры –«Corpus Albaruthenicum». Для паралельнага корпуса з беларускай мовай арыгінала –найдено 6 документов общим объемом 15 414 предложений, 149 319 слов”. З мовай арыгінала рускай –найдено 14 документов общим объемом 65 078 предложений, 1 419 538 слов”.

Наўрад ці можна лічыць рускамоўны інтэрфэйс НКРЯ нейкім абмежаваннем: беларуская мова тут суседнічае з нямецкай, англійскай, украінскай і шэрагам іншых (малюнак 2).

The screenshot shows the web interface of the National Russian Corpus (НКРЯ). The header includes the logo and text: "НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА". The main content area is titled "Параллельный корпус (белорусский)". On the left, there is a sidebar with navigation links: "главная", "основной", "синтаксический", "газетный", "параллельный", "обучающий", "диалектный", "позитический", "устный", "акцентологический", "мультимедийный". The main search area has a search bar with the text "Поиск точных форм" and a dropdown menu for "Слово или фраза". Below the search bar are buttons "искать" and "очистить". The search configuration panel includes fields for "Слово", "Грамм. признаки", "Доп. признаки", and "Расстояние".

Малюнак 2 – Інтэрфэйс Паралельнага (беларускага) корпуса –НКРЯ”

НКРЯ змяшчае больш за трыста мільёнаў словаўжыванняў і пастаянна расце. На сённяшні дзень аб’ём усяго паралельнага корпуса, часткова нават шматмоўнага: –370 документов, 1 609 609 предложений, 24 022 437 слов”.

Адпаведна павелічэнню аб'ёмаў з цягам часу будзе ўзрастаць дакладнасць дадзеных з любой сукупнасці тэкстаў. Дынаміка развіцця НКРЯ з яго паступовым, але няўхільным нарошчваннем масіва, дае падставы спадзявацца на паглыбленне лінгвістычнай кампетэнтнасці статыстыкі. Запраграмаванасць на паралельнасць значна абмяжоўвае выніковасць і ўніверсалізм даследаванняў на яго базе, але нават такое функцыяніраванне можа быць надзвычай карысным. Ужо сёння шэраг задач, у прыватнасці адносна валентнасці, корпуснымі праектамі можа вырашацца паспяхова.

Недахопам «Corpus Albaruthenicum» безумоўна падаецца дэфіцыт лінгвістычных інструментаў, як мінімум іх непразрыстасць для карыстальнікаў. Адсюль і беднасць пошукавых магчымасцей згаданага корпуса: гэты рэсурс не вырашае праблемы пошуку матэрыялу на падставе граматычных і семантычных прыкмет. Магчыма – пабудова канкардансаў і пошук словазлучэнняў. Аднак па прычыне беднасці корпуса цяжка падабраць варыянт словазлучэння, адлюстраваны ў корпусе.

Іншая справа – Паралельны корпус НКРЯ, дзе выкарыстаны разгалінаваныя сістэмныя сродкі корпуснага менеджэра, які на ўзроўні лепшых сусветных узораў працуе ў шэрагу лінгвістычных напрамкаў – ад семантыкі да сінтаксісу. Зразумела, што першачарговую цікавасць для расійскіх даследчыкаў уяўляе беларуская мова не аднаасобна, а ў кантэксце рускай і іншых моў: вынікі, напрыклад, «пошуку» Паралельны корпус НКРЯ прадстаўляе ў выглядзе дзвюхмоўных ці шматмоўных канкардансаў.

Аналіз паказвае, што ў руска-беларускім паралельным корпусе НКРЯ пошук адзінак паводле іх прыналежнасці да пэўных часцін мовы працуе. Зразумела, хацелася б бачыць больш дакладны пошук слоў па граматычных прыкметах, але канкардансер Паралельнага корпуса НКРЯ можа памыляцца ў «лексіка-граматычным пошуку»: блытае часціны мовы. Такая сітуацыя назіраецца, калі задаецца пошук тыпу «\*вы»: тут будзе і рускае *увы* і беларускае *крывы, дровы* і г. д. Тым не менш, пошук і прыметнікаў, і назоўнікаў з прапанаванай фіналлю выкананы, пры пошуку назоўнікаў знойдзена 480 уваходжаньняў, а прыметнікаў – 51. Няма сумненняў, што распрацоўшчыкі корпуса недахопы выправяць.

Вельмі цікавымі маглі б атрымацца даследаванні марфемнай спалучальнасці ў беларускай мове. Але структура слова недастаткова распрацаваная нават у Паралельным корпусе НКРЯ: праграма шукае спалучэнні тыпу «\*не» і «не\*», але адсутнічае выніковасць з прыкладамі на «\*не\*».

Абодва корпусы, у рознай ступені, на сённяшні дзень дазваляюць ілюстраваць даследаванні па спалучальнасці адзінак беларускай мовы. Даследаванні вызначанага кшталту маюць непасрэднае дачыненне як да эмпірычнай, так і да сістэмнай прадуктыўнасці словаўтварэння.

Сістэмная прадуктыўнасць – паняцце сінхроннага словаўтварэння. Яна залежыць ад унутрымоўных фактараў і абмежаванняў рознага кшталту, якія накладваюцца сістэмай мовы і рэгулююць вытворчасць слоў дадзенага тыпу. На думку Е. А. Земскай, «аб ахарактарызаваць сістэмную прадуктыўнасць

тыпу, трэба ўстанавіць, якія структурныя, семантычныя і стылістычныя абмежаванні існуюць ва ўтварэнні слоў дадзенага тыпу...” [1, с. 330]. Шырокая жанравая і стылявая разнастайнасць тэкстаў корпусаў — «цыянальнага» тыпу дазволіць (у перспектыве) прыцягваць практычна неабмежаваныя стасункі беларускай мовы: ад мовы мастацкай літаратуры да мовы Інтэрнэту.

Устанаўленне, напрыклад, не толькі магчымасцей ўтварэння слоў дадзенага тыпу, але і адсутнасць гэтых магчымасцей («немагчымасці») мае не апошняе значэнне ў лінгвістычных даследаваннях.

Эмпірычная прадуктыўнасць, згодна з М. Докулілам — паняцце дыхранічнага словаўтварэння [4, с. 205]. Яна залежыць ад сацыяльнай запатрабавальнасці ў вытворчасці слоў пэўнага тыпу. Вызначыць, ці характарызуецца словаўтваральны тып эмпірычнай прадуктыўнасцю, — значыць адказаць на пытанне, ці дае ён наватворы, ці актыўны ён як узор у той ці іншы перыяд часу.

У НКРЯ ужо сёння прадстаўлены узоры беларускай мовы як 20-х, так і 80-х гадоў XX ст. Цяжка сумнявацца, што нават пры такіх сціпрых магчымасцях адпаведныя даследаванні будуць з’яўляцца. Корпусная лінгвістыка, як адказ на запатрабаванні часу ў вывучэнні маўлення — надзейны і эфектыўны сродак. Наколькі марудна і паступова развіваецца беларуская корпусная лінгвістыка — неабходна канстатаваць, што першыя крокі ў выглядзе адкрытых корпусаў «Corpus Albaruthenicum» і Паралельнага беларуска-рускага корпуса ў складзе НКРЯ — нарэшце зроблены. Абодва рэсурсы існуюць і прыдатныя для абмежаваных лінгвістычных даследаванняў.

#### СПІС ЛІТАРАТУРЫ

- 1 Земская, Е. А. Словообразование / Е. А. Земская // Современный русский язык. — М. : Высшая школа, 1989. — С. 237–379.
- 2 Национальный корпус русского языка [Электронны рэсурс]. — Рэжым доступу: [www.ruscorpora.ru](http://www.ruscorpora.ru). — Час доступу: 08.10.2012.
- 3 Corpus Albaruthenicum [Электронны рэсурс]. — Рэжым доступу: <http://grid.bntu.by/corpus>. — Час доступу: 08.10.2012.
- 4 Dokulil, M. Tvoření slov v češtině / M. Dokulil. — Praha : Nakladatelství CSAV, 1962. — Dil 1. Teorie odvozování slov. — 219 s.

УДК 81‘33

А. А. Баркович, Ван Цин (Минск)

### **ИЗУЧЕНИЕ ЛЕКСИЧЕСКОГО ЗНАЧЕНИЯ МНОГОЗНАЧНЫХ СЛОВ НА БАЗЕ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ**

*В данной статье рассматривается вопрос использования лингвистического корпуса для изучения лексического значения многозначных слов.*



*Описываются возможности использования лингвистического корпуса в обучении иностранных (китайских) учащихся русскому языку.*

Цель обучения иностранным языкам – формирование личности учащегося, способной и желающей участвовать в общении на межкультурном уровне. Этот процесс подразумевает становление у учащегося основных черт вторичной языковой личности и их закрепление с помощью всех доступных лингводидактических средств. Учёт возможностей компьютерных технологий, с одной стороны, и активное их использование в повседневной жизни, с другой стороны, обусловили вовлечение в учебный процесс как новых методик, так и доступных посредством Интернета материалов. Такими относительно новыми объектами и инструментами изучения в современном лингвистическом контенте являются корпуса.

Корпусный менеджер – своеобразная операционная оболочка лингвистического корпуса – предоставляет целый ряд возможностей для исследователей языка. Наиболее доступна и практична в процессе изучения иностранного языка начинающими пользователями, которыми по определению являются учащиеся – особенно инофоны – программа «конкордансер», строящая конкордансы. В Национальном корпусе русского языка (НКРЯ) в помощь обучающимся предоставляется весьма информативная инструкция, своего рода «руководство пользователя» [3].

Конкорданс – это программа, позволяющая анализировать большие массивы текста на предмет обнаружения закономерностей использования в языке слов или выражений. Конкорданс осуществляет поиск запрашиваемого слова в корпусе и выдает в новом окне несколько фрагментов предложений из разных текстов, в которых используется данное слово или выражение [6].

Тем самым выявляется местоположение конкретных слов и выражений в текстовых массивах. Наиболее широкое распространение в методических целях лингвистический корпус может получить при формировании лексико-грамматических навыков обучающихся.

«Слово – основная единица языка и разносторонний языковой знак. В нём сочетаются:

- 1) означающее (внешняя звуковая сторона, звуковая оболочка);
- 2) означаемое (внутренняя сторона, значение).

Лексическое значение слова – отражение в слове того или иного явления действительности (предмета, события, качества, действия, отношения)» [2].

Лексическое значение определяет все возможные связи и отношения слова с другими словами, поэтому в обучении РКИ, например, китайских учащихся первым этапом, как правило, и является определение лексического значения слова.

Большинство русских слов имеют не одно, а несколько значений и являются многозначными. Многозначность слова реализуется в речи: контекст и проясняет одно из конкретных значений многозначного слова.

Возьмём для примера возможности по сравнению лексического значения русской единицы *горло* и подобной ей – китайской 咽喉 (*yānhóu*).

Словарная статья для слова *горло* выглядит следующим образом. «*Горло* -а, ср. 1. Передняя часть шеи. По г. в воде. Схватить за г. (также перен.: неотступно требовать чего-н.; разг. неодобр.). С ножом к горлу пристать (приступить) (перен.: неотступно просить, требовать; разг. неодобр.). Взять за г. кого-н. (также перен.: грубо принудить к чему-н.; разг. неодобр.). Г. перегрызть кому-н. (также перен.: в ярости уничтожить кого-н.; разг.). Вцепиться в г. кому-н. (перен.: с яростью обрушиться на кого-н.). 2. Общее название зева, глотки (в 1 знач.) и гортани. В горле сохнет. Во все г. кричать (очень громко; разг.). Поперек горла стать кому-н. (перен.: о раздражающей помехе; разг.). Горлом берет кто-н. (добивается чего-н. при помощи крика; разг. неодобр.). Промочить г. (немного выпить; разг.). В три горла есть (очень много; разг. неодобр.). 3. Верхняя суженная часть сосуда. Г. кувшина. 4. Пролив, соединяющий залив с морем, а также рукав, соединяющий устье реки с морем, озером (спец.). \* По горло кого-чего (разг.) – очень много, в избытке. Дел по горло. Сыт по горло (разг.) – 1) совершенно сыт. Пообедал, сыт по горло; 2) чем, чего-н. много до пресыщения (неодобр.). Сыт по горло обещаниями. || уменьш. горлышко, -а, ср. (к 1 и 2 знач.). || прил. горловой, -ая, -ое (к 1 и 2 знач.). Горловое кровотоечение» [4, с. 123].

В НКРЯ найдено 2523 документа, 6098 вхождений (словоупотреблений) единицы *горло*. В соответствующих контекстах представлены указанные в словаре значения во всём их многообразии [3].

*Порубить шею, хребет и крылья на маленькие кусочки. Порезать горло. Приготовление соуса.* [Рецепты национальных кухонь: Франция (2000–2005)].

*Они состоят из таких вот пузырьков, я через горло вдыхаю воздух, пузырьки раздуваются, когда выдыхаю – сдуваются, происходит реакция окисления от кислорода – это нужно, чтоб я жил.* [Евгений Гришковец. ОдноврЕмЕнно (2004)].

*Бутылки наполняют до основания горла, укупоривают пробками и заливают парафином, погружая горло бутылки в расплавленный парафин, или запечатывают сургучом.* [Т. М. Воробьева, Т. А. Гаврилова. Домашнее консервирование (1999)].

*Короткое время спустя голос мотора стал отчетливее, и вот уже лодка, круто повернув с реки, вбежала в горло затона, ткнулась в берег.* [Борис Екимов. Высшая мера (1995)].

Достаточно много вариантов использования рассматриваемой лексики в переносном значении:

*Геометрия души. Судьба, схваченная за горло... Не менее достойны удивления случаи, до которых люди доходят в познании небесных вещей, чем сама природа этих вещей.* [Вячеслав Шевченко. Демон науки: Космический кубок // «Знание – сила», 2003].

А тогда я всё ждал, пока прояснится голова, и, кроме того, вдруг оказался по **горло** занят – у меня появилась новая и непривычная обязанность. [Вера Белоусова. Второй выстрел (2000)].

Есть и подтверждённое словарной статьёй значение, характерное для разговорной речи:

Андрей, забыв о лифте, летел вниз, хохоча во всё **горло**. [Анатолий Азольский. Лопушок // «Новый Мир», 1998].

Утверждение Г. Пауля о том, что производным значением человек нередко овладевает раньше, чем основным, полностью подтверждается корпусными данными, большим количеством словоупотреблений **горла** в переносном значении [5, с. 106].

Слово, имеющее значения передняя часть шеи и общее название зева, глотки и гортани, на китайском языке – это 咽喉 (*yānhóu*). У этой единицы два значения: «1) 咽和喉 (*yān hé hóu* – гортань и глотка); 2) (比喻) 形势险要的交通孔道 ((*bǐyù*) xíngshì xiǎnyào de jiāotōng kǒngdào – (перен.) ключевая позиция; жизненно важный, ключевой)» [8].

В составе «Лингвистического корпуса китайского языка» (Center for Chinese Linguistics PKU – (CCL)), созданного в 2000-м году, функционируют три субкорпуса: «Корпус современного китайского языка», «Корпус древнего китайского языка» и «Параллельный китайско-английский корпус». Объем всего корпуса – 9711 лексических единиц [7]. В «CCL» найдено 1143 контекста, в которых употребляется слово 咽喉 (*yānhóu*).

В значении гортань и глотка лексема использована в следующих контекстах:

蜜饯、果干、果酱和罐头食品。果干入药，能开胃止泄，治疗咽喉痛，是治疗喘咳、吐血和痔疮的良药。

Mìjiàn, guǒgān, guǒjiàng hé guàntóu shípǐng. Guǒgān rùyào, néng kāiwèi zhǐxiè, zhìliáo yānhóutòng, shì zhìliáo xiàochuǎn, tùxuě hé zhìchuāng de liángyào.

Цукаты, сухофрукты, варенье и консервы используются в качестве лекарства, которое помогает против насморка, избавляет от боли в **горле**, а также используется для лечения других болезней (здесь и далее – перевод авторов).

对于因胃酸过多引起胃、十二指肠溃疡病有治疗作用。甘草能维护咽喉部和气管的粘膜，有润喉止咳的功效。

Duìyú yīn wèisuān guòduō yǐngqǐ wèi, shíèrzhǐcháng kuìyángbìng yǒu zhìliáo zuòyòng. Gāncǎo néng wéihù yānhóubù hé qìguǎn de niánmó, yǒu rùnhóu zhǐkéké gōngxiào.

При лечении язвенной болезни двенадцатиперстной кишки, вызванной повышенной кислотностью, солодка поддерживает слизистые оболочки **горла** и трахеи<sup>4</sup>.

В следующих контекстах подразумевается **ключевая позиция**<sup>4</sup>, нечто **жизненно важное, ключевое**<sup>4</sup>:

它是沟通大陆和半岛的中间桥梁，也是交通的**咽喉**

要道。地峡比较狭窄，两边邻水，是开凿运河的良好地段。

Tā shì gōutōng dàlù hé bàndǎo de zhōngjiān qiáoliáng, yěshì jiāotōng de **yānhóu** yàodào. Dìxiá bǐjiǎo xiázhǎi, liǎngbiān línshuǐ, shì kāizáo yùnhé de liánghǎo diduàn.

Для связи с материком и полуостровом используется мост, что очень **важно**, в узкой части которого есть пропускные пункты<sup>4</sup>.

深度310米。对于大西洋和地中海来说,直布罗陀海峡真像它们的**咽喉** – 样重要。

Shēngdù 310 mǐ. Duiyú dàxīyáng láishuǐ, zhíbùluótóu hǎixiá zhēn xiàng tāménde **yānhóu** yíyàng zhòngyào.

Глубина Гибралтарского пролива в 310 метров – для Атлантики и Средиземного моря очень **впечатляюща**<sup>4</sup>.

В «Большом китайско-русском словаре» значение **верхняя суженная часть сосуда**<sup>4</sup> обозначается лексемой – 颈 (jǐng, **горлышко**<sup>4</sup>); значение **пролив, соединяющий залив с морем, а также рукав, соединяющий устье реки с морем, озером**<sup>4</sup> – 海湾狭口 (hǎiwān, xiǎkǒu, **выход из залива**<sup>4</sup>) [1].

Эти слова на китайском языке имеют значения отличные от **咽喉** (yānhóu), и пишутся по-другому. Единицы, соответствующие русским значениям слова **горло**, в китайском языке выражены тремя разными словами: **咽喉** (yānhóu), **颈** (jǐng) и **海湾狭口** (hǎiwān, xiǎkǒu).

Итак, лингвистический корпус может значительно оптимизировать процесс обучения русскому языку как иностранному. На базе лингвистического корпуса можно интенсивно и экстенсивно проанализировать примеры, иллюстрирующие словарные значения слов, и особенности их речевой воспроизводимости. Что немаловажно, данная работа может выполняться учащимися самостоятельно.

#### СПИСОК ЛИТЕРАТУРЫ

1 Большой китайско-русский словарь [Электронный ресурс]. – Режим доступа: <http://bkrs.info> – Дата доступа: 08.10.2012.

2 Жеребило, Т. В. Словарь лингвистических терминов: Изд. 5-е, испр-е и дополн. / Т. В. Жеребило. [Электронный ресурс] – Режим доступа: [http://lingvistics\\_dictionary.academic.ru](http://lingvistics_dictionary.academic.ru) – Дата доступа: 08.10.2012.

3 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: [www.ruscorgo.ru](http://www.ruscorgo.ru). – Дата доступа: 08.10.2012.

4 Ожегов, С. И. Толковый словарь русского языка: Ок. 100 000 слов, терминов и фразеологических выражений / С. И. Ожегов; под ред. проф. Л. И. Скворцова. – 27-е изд., испр. – М.: ООО «Издательство Оникс»: ООО «Издательство «Мир и образование», 2011. – 736 с.

5 Пауль, Г. Принципы истории языка / Г. Пауль. – М.: Изд-во иностр. лит-ры, 1960. – 500 с.

6 Сысоев, П. В. Лингвистический корпус в методике обучения иностранным языкам / П. В. Сысоев. [Электронный ресурс]. – Режим доступа: <http://www.lib.tsu.ru/mminfo/000349304/09/image/09-099.pdf> – Дата доступа: 08.10.2012.

7 Center for Chinese Linguistics PKU (CCL) [Электронный ресурс]. – Режим доступа: [http://ccl.pku.edu.cn:8080/ccl\\_corpus/index.jsp?dir=xiandai](http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xiandai) – Дата доступа: 08.10.2012.

8中国社会科学院语言研究所 现代汉语词典 / 中国社会科学院语言研究所, – 商务印书馆 [Словарь современного китайского языка / Университет китайского языка при академии социальных наук], 2005.

УДК 004.738.5(476)

С. А. Важнік (Мінск)

### **САБАКА, МАЛПА ЦІ СЛІМАК: ДА ПЫТАННЯ АБ ТРОХ СТЫХІЯХ БАЙНЭТА**

*Беларуская мова паступова “адваёўвае” новыя сферы ўжытку. Адной з такіх сфер з’яўляецца Інтэрнэт-камунікацыя, аналізу спецыфічных асаблівасцей якой і будзе прысвечана дадзеная публікацыя. Навідавоку карэляцыя мовы беларускага *netspeak*’а з афіцыйным варыянтам СБЛМ, арыентаваным на рускую мову, з “інавацыйным” варыянтам СБЛМ, арыентаваным на “Беларускую граматыку для школ” (Вільня, 1918) Б. Тарашкевіча, на польскую мову і заходнебеларускія гаворкі, а таксама з пурыстычнай тэндэнцыяй. Усе высновы зроблены адносна маўленчай практыкі 90-х гг. XX ст. – пачатку XXI ст.*

Жывучы ў час “лінгвістычнага постмадэрнізму” на Беларусі, мы не можам не заўважыць новых фарматаў, новых сфер, у якіх выкарыстоўваецца беларуская мова (*Інтэрнэт, мабільная сувязь, рэклама і інш.*), а таксама новых дынамічных працэсаў, якія ў сувязі з гэтым адбываюцца ў сучаснай беларускай літаратурнай мове (далей – СБЛМ).

Для беларускага Інтэрнэт-дыскурсу магчымыя як мінімум чатыры варыянты тлумачэння сімвалу @:

- 1) *at* (*y / na*) – з англійскай мовы;
- 2) *сабака* (*сабачка*) – з рускай мовы;

3) *малпа* – з польскай мовы;

4) *слімак* (*слімачок*) – арыгінальная беларуская назва.

У Байнэце @ тлумачыцца таксама як: “*вітушка*”, “*кручво*”, “*смоўж*”, “*котка*”.

Такім чынам, наша задача зводзіцца да высвятлення таго, чаго ў мове беларускага Сеціва больш – *сабакі*, *малты* ці *слімака*?

Беларускі Інтэрнэт – гэта па сутнасці лакмусавая паперка (люстэрка) той моўнай сітуацыі, што мае месца ў сучаснай Беларусі. Характарызуючы апошнюю, адзначым наяўнасць двух варыянтаў СБЛМ: *афіцыйнага* (*кадыфікаванага*) і *інавацыйнага* (*альтэрнатыўнага*); паралельнае суіснаванне і нават *жанкурэнцыю* [1, с. 131] дзвюх правапісных сістэм: “*тарашкевіцы*” і *наркамаўкі*”; а таксама традыцыйную для сучаснай Беларусі “*трасянку*”.

Мова беларускага *netspeak'a* досыць далёкая ад літаратурнага стандарту. Яна больш жывая нават у параўнанні з мовай такіх выданняў, як “*Наша Ніва*”, “*Arche*”, “*Фрагмэнты*”, “*Студэнцкая Думка*”, “*pARTisan*”, “*Farba*” і інш. Таму напрашваецца выснова адносна таго, што беларуская мова ў Інтэрнэце – гэта адмысловы варыянт СБЛМ.

Найбольш істотнай асаблівасцю беларускамоўных І.-рэсурсаў з’яўляецца выкарыстанне ў іх пераважна *інавацыйнага варыянта* СБЛМ. Правапіс такіх рэсурсаў – у асноўным *тарашкевіца*. У гэтым няма нічога дзіўнага. *Насельнікі*” Байнэта (пераважна маладыя людзі) успрымаюць *наркамаўку* як старую, нягнуткую, ненатуральную сістэму vs. *тарашкевіцу* – як жывую, свабодную ад *умоўнасцей*”, адкрытую сістэму. Такім чынам, у вачах моладзі інавацыйны варыянт СБЛМ – больш прэстыжны ў параўнанні з афіцыйным варыянтам.

У Інтэрнэце беларуская мова пачуваецца досыць вольна і гнутка. Сферы яе ўжытку з кожным годам усё больш пашыраюцца. Яна паступова выходзіць з *дэкаратыўна-рытуальнай* і *побытавай* сфер (мастацкая літаратура, святочныя прамовы, “*жубочная*” псеўда-архаіка на ўсебеларускіх святах з удзелам фальклорных калектываў; сетка беларускамоўных перадач на БТ, ОНТ і беларускім радыё; мясцовыя гаворкі; трасянка).

Больш за тое, у інтэрнэтным маўленні, як і ў паэтычным тэксце, даволі моцная тэндэнцыя да ігнаравання / разбурэння (лібералізацыі) афіцыйных норм і да частковай іх замены інавацыйнымі варыянтамі, што маюць *кантрасны* характар у дачыненні да рускай мовы і да афіцыйнага варыянта СБЛМ, – варыянтамі, узятымі: 1) з моўнай практыкі 20–30-х гадоў ХХ ст.; 2) з дыялектных крыніц (перадусім – заходнебеларускіх гаворак); 3) са славянскіх (польскай, украінскай і чэшскай) і германскіх (перадусім – англійскай) моў (працэсы запазычвання, а таксама калькавання слоў і ўстойлівых канструкцый).

Тэндэнцыя да *адштурхоўвання*” ад усяго рускага, якую можна лічыць у нейкай ступені *праявай пурыстычнай тэндэнцыі* (рэакцыя на моўную сітуацыю ў 50–80-я гады ХХ ст.), выразна праяўляецца і ў Сеціве. Аднак

у сваёй барацьбе з «усім рускім» радыкальная плыня не заўважае, як трапляе ў іншую пастку – пастку паланізацыі. А гэта ўжо чарговая крайнасць.

Прычыны ўзнікнення шматлікіх варыянтаў у СБЛМ, відавочна, звязаны з пошукамі яе носьбітамі (часцей за ўсё маладымі беларусамі) рацыянальных і натуральных крыніц і шляхоў яе развіцця. Адзначаныя вышэй сродкі запускаяць механізм натуральнага развіцця беларускай мовы, які доўгі час —шучна стрымліваўся жорсткімі нарматыўнымі рэгламентацыямі» [2, с. 16].

Нягледзячы на нераспрацаванасць у беларусістыцы такой галіны, як *Інтэрнэт-лінгвістыка*, і на адсутнасць спецыяльных даследаванняў, прысвечаных І.-камунікацыі, ужо зараз відавочна, што Сеціва не можа ў бліжэйшай перспектыве не паўплываць на развіццё натуральных моў. І вельмі важна гэты працэс трымаць пад кантролем, бо мовы «безабаронныя» і моцна паддаюцца негатыўнаму ўплыву размоўнай плыні. Безумоўна, думка аб *маніторынгу* мовы Сеціва натуральна гучыць у дачыненні да любой мовы, а для беларускай – асабліва. Так, беларускія мовазнаўцы мусяць дбаць пра чысціню беларускай мовы. Аднак існуе пэўная спецыфіка *беларускай* моўнай сітуацыі. Для сучаснай беларускай мовы Інтэрнэт – прастора, у якой ніхто і нішто не перашкаджае яе натуральнаму развіццю, не замінае ёй заставацца «жывой», свабоднай ад жорсткіх абцугоў кадыфікаваных норм і правілаў. На нашу думку, задача беларусістаў – не ўмешвацца ў «ворчы працэс», а ўдумліва і крытычна адзначаць і фіксаваць усё тое, што ў мове Байнэта адбываецца. І не рабіць пакуль лёсавызначальных высноў адносна негатыўнага ўплыву Інтэрнэта на чысціню і мілагучнасць беларускай мовы.

Чаго ж у мове беларускага Сеціва больш – *сабакі*, *малы* ці *слімака*?

Мусім канстатаваць наяўнасць усіх элементаў пры пэўным дамінаванні *малы*: паралельнае суіснаванне двух варыянтаў СБЛМ, а таксама дзвюх правапісных традыцый; трасянканасць; моцная пурыстычная традыцыя; шматлікія інавацыі польскага паходжання. Гэта адпавядае агульнай тэндэнцыі быцця СБЛМ: спецыфіка, унікальнасць беларускай мовы ў яе *сінкрэтызме* (параўн., напрыклад, з канцэпцыяй «Беларуская літаратурная мова: паміж *Slavia Orthodoxa* і *Slavia Latina*» ў [3]).

#### СПІС ЛІТАРАТУРЫ

1 Мечковская, Н. Б. Языковая ситуация в Беларуси начала XXI в. Новые черты. Момент истины о белорусском суверенитете летом 2002 г. / Н. Б. Мечковская // Белорусский язык: социолінгвістычныя очеркі. – München, 2003. – С. 126–143.

2 Арцямёнак, Г. Лексічныя інавацыі польскага паходжання ў сучаснай беларускай літаратурнай мове / Г. Арцямёнак // Беларуская-руская-польская супастаўляльная мовазнаўства і літаратуразнаўства: Мат-лы III Міжнароднага навуковага канферэнцыяна. – Т. 1. – Віцебск, 1994. – С. 13–16.

3 Супрун, А. Я. Беларуская літаратурная мова: паміж *Slavia Orthodoxa* і *Slavia Latina* / А. Я. Супрун // Мова – Літаратура – Культура: Мат-лы Міжнароднага навуковага канферэнцыяна. (Мінск, 22–23.09.1998 г.). – Ч. 1. Мінск, 1999. – С. 21–24.

## **ПРОЕКТ КОРПУСА БЕЛОРУССКОЯЗЫЧНОЙ ПЕРИОДИКИ И ХУДОЖЕСТВЕННОЙ ПРОЗЫ**

*Исследуются особенности корпуса белорусскоязычной периодики и художественной прозы. Подробно описываются состав и структура корпуса, процедура сбора и подготовки текстов, поисковые программы. Намечаются основные пути дальнейшего развития проекта.*

Корпусная лингвистика – одно из перспективных и быстро развивающихся направлений в современном языкознании. За последнее десятилетие количество лингвистических корпусов значительно увеличилось, причем, наряду с небольшими узкоспециальными проектами, ведется активная разработка национальных корпусов. Свои национальные корпуса уже имеют многие славянские языки, среди них русский (<http://ruscorpora.ru>), польский (<http://nkjp.pl>) и украинский ([http://lcorp.ulif.org.ua/virt\\_unlc](http://lcorp.ulif.org.ua/virt_unlc)). Создаваемые при государственной поддержке, эти ресурсы характеризуются большим охватом материала и сравнительно глубокой аннотацией: к примеру, «Narodowy Korpus Języka Polskiego» имеет объем 1,5 млрд. словоупотреблений, в том числе 1 млн. – со снятой лексической и грамматической омонимией [4, с. 36].

В Беларуси корпусная лингвистика рассредоточена по нескольким научным центрам. В МГЛУ под руководством А. В. Зубова не один год создаются небольшие параллельные корпуса, доступ к которым ограничен. В 2011 г. в сети был размещен «Corpus Albaruthenicum» (<http://grid.bntu.by/corpus>), разработанный в Институте языка и литературы НАН РБ и содержащий 350 тыс. словоформ (74 научных текста) с неснятой омонимией. Коллектив исследователей из ГрГУ, возглавляемый Л. В. Рычковой, работает над корпусом русскоязычных газет Гродненской области. Наконец, совсем недавно НКРЯ был обогащен белорусско-русским параллельным подкорпусом объемом в 2 млн. словоупотреблений.

Как видно, для белорусского языка остро стоит проблема создания представительного одноязычного корпуса, который бы свободно распространялся и был качественно аннотирован (по крайней мере, морфологически). Эта заметка – первое предварительное сообщение о попытке разработать такой ресурс, предпринятой на филологическом факультете БГУ группой энтузиастов.

### **Состав и структура корпуса**

На сегодняшний день в состав нашего корпуса входят подкорпуса газетных текстов и художественной прозы. Материалом для первого из них стали годовые архивы газет «Звезда» и «Чырвоная змена» за период с августа 2008 г. по июль 2009 г. включительно, а также двухгодичный



архив газеты «Голас Радзімы» за 2008–2009 гг. (всего около 9 тыс. статей, 4,4 млн. словоформ). Статьи из других периодических изданий пока не вошли в корпус по разным причинам, в том числе технического характера (например, отсутствие электронного архива).

Второй подкорпус имеет двухчастную структуру:

1. Подборка текстов первой половины XX в. На данный момент в нее вошла только проза Я. Коласа: трилогия «На ростанях», повесть «Дрыгва» и сборник «Казкі жыцця» (всего 248 тыс. словоупотреблений).

2. Современная белорусская литература. Источником материала послужили журналы «Полымя», «Маладосць» и «Дзеяслоў» за 2009–2011 гг. Из каждого номера отбирались лишь тексты, отнесенные редакцией к рубрике «Проза». Среди авторов, число которых приближается к 200, есть и признанные классики, и начинающие прозаики.

Мы попытались сбалансировать эту часть подкорпуса по двум параметрам:

(1) Жанр. Было принято решение ориентироваться на жанровые пропорции, свойственные подкорпусу художественной прозы XIX в. НКРЯ [1]; как следствие, романы и крупные повести входили в текстовую базу не целиком.

(2) Доля произведений каждого автора. Чтобы избежать преобладания идиолекта нескольких активно печатающихся прозаиков, мы выбирали фрагменты из их текстов таким образом, чтобы общее количество словоформ не превосходило 2% от всей выборки журнальной прозы.

Сейчас размер этой части подкорпуса – около 2,2 млн. словоупотреблений, но перечень текстов не окончателен.

С технической стороны корпус представляет собой иерархию папок. Каждый текст (газетный или художественный) хранится в отдельном текстовом файле с кодировкой ANSI 1251. Художественная проза дополнительно снабжена метатекстовой информацией: автор, заглавие, жанровая принадлежность. Подробнее ознакомиться с устройством ресурса можно на странице <https://github.com/poritski/YABC>, где размещен его демонстрационный фрагмент с подробной документацией.

### **Процедура сбора и подготовки текстов**

Публикации газет «Звязда», «Чырвоная змена», «Голас Радзімы» и журнала «Дзеяслоў» размещены в Интернете в виде HTML-страниц. Для их получения было разработано два скрипта на языке PHP, первый из которых формирует списки требуемых страниц, а второй скачивает эти страницы. При очистке данной части корпусного материала от метатекстовой разметки, как всегда, возник ряд трудностей, связанных с выделением собственно текстового фрагмента, неунифицированной передачей отдельных символов (тире, кавычек, белорусского *i*), спецсимволами HTML и др.

Чтобы справиться с этими сложностями, был написан простой скрипт очистки на языке Perl. В отличие от [5] и других технологий, основанных на принципах машинного обучения, на целевом множестве файлов наш

скрипт обеспечивает качество подготовки текста, близкое к 100 %, но при этом не является универсальным, т. е. не может эффективно очищать от метатекстовой разметки, например, блогговые записи.

Особая проблема – повторяющиеся фрагменты текста внутри одной газетной статьи. Разработан скрипт, способный с хорошей точностью идентифицировать такие дубли, но окончательное решение об удалении мы пока оставляем за пользователем.

Произведения, опубликованные в «Полымі» и «Маладосці», были извлечены из PDF-файлов соответствующих номеров журналов. При этом не обошлось без двух затруднений. Во-первых, конвертер PDF→ТХТ распознает журнальную страницу, сверстанную в несколько колонок, как одну колонку сплошного текста с группами пробелов посередине; дробить текст на колонки по группам пробелов переменной длины технически неудобно. Во-вторых, проблематично разграничить знак переноса и дефис в конце графической строки. Все это потребовало тщательной вычитки сконвертированного текста.

### Поисковый механизм

В нынешней реализации ресурса к каждому из подкорпусов прилагаются две программы на языке Perl. Они обеспечивают достаточно быстрый поиск точных форм слов, в том числе любых наборов форм (например, парадигм склонения), и вывод найденных контекстов в файл.

Скрипт-индексатор *indexer.pl* последовательно читает все файлы подкорпуса, токенизирует их (т. е. явным образом выделяет словоформы и знаки препинания) и строит обратный индекс – таблицу соответствий между словоформами и текстовыми файлами, в которых эти словоформы встречаются. Отдельно индексатор выводит статистику числа графических слов по каждому документу.

В служебном файле *wordlist.txt* пользователь задает свои поисковые запросы. Каждый запрос записывается на одной строке и состоит из регулярного выражения и его идентификатора, разделенных знаком табуляции. Регулярное выражение может определять одну словоизменяющую парадигму, а идентификатор – совпадать с начальной формой слова. Пример запроса:

гарызон(t(a(ŷ|m(i|i)?|x)?|ы|у)?|це) гарызонт

Основной поисковый скрипт *finder.pl*, обращаясь к индексу и текстовой базе, обрабатывает *wordlist.txt*. Результаты поиска скрипт выводит в отдельный текстовый документ, где каждому вхождению соответствует одна строка, содержащая информацию о файле-источнике, идентификатор и точную найденную форму в окружении левого и правого контекстов, ширину которых пользователь может задать самостоятельно.

### Перспективы развития проекта

Совершенствовать наш корпус предстоит в нескольких направлениях. Первоочередной задачей пока остается нанесение морфологической разметки. Для белорусского языка уже существует машинная морфология,

созданная под руководством И. В. Совпеля [2]. С недавнего времени в сети доступен ее вариант [3], с помощью которого мы и надеемся проаннотировать корпус. После этого понадобится много усилий, чтобы снять грамматическую омонимию хотя бы в части корпуса (до 1 млн. словоупотреблений).

Еще одна немаловажная задача – совершенствование текстовой базы: ее расширение (в основном за счет диахронии) и улучшение балансировки. Параллельно будет проводиться очистка текстов от дублей и дооснащение метаинформацией. Кроме того, мы ожидаем, что большинство имеющихся опечаток будет выявлено и исправлено при морфологическом аннотировании.

Наконец, предполагается доработать программную часть, т. к. существующий поисковый механизм не справляется с более чем однословными запросами, не способен учитывать грамматические сведения, а его быстродействие слабо оптимизировано. В долгосрочной перспективе мы также рассчитываем организовать доступ к корпусу через веб-интерфейс.

#### СПИСОК ЛИТЕРАТУРЫ

1 Оскольская, С. А. Корпус письменных текстов XIX века: сферы употребления и жанровое разнообразие / С. А. Оскольская // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы / Отв. ред. В. А. Плунгян. – СПб.: Нестор-История, 2009. – С. 46–51.

2 Совпель, И. В. Компьютерный фонд белорусского языка и его приложения / И. В. Совпель, Н. К. Рубашко, Г. П. Невмержицкая // Информационные системы и технологии (IST 2006) : материалы III междунар. конф. / БГУ, НАН РБ, Науч.-техн. ассоциация «Инфопарк», АУ при Президенте РБ. – Минск, 2006. – Ч. 2. – С. 71–76.

3 Hetsevich, Yu. S. Belarusian and Russian linguistic processing modules for the system NooJ as applied to text-to-speech synthesis / Yu. S. Hetsevich, S. A. Hetsevich, B. M. Lobanov // Диалог: Международная конференция по компьютерной лингвистике [Электронный ресурс]. – Режим доступа: <http://www.dialog-21.ru/digests/dialog2012/materials/pdf/136.pdf>. – Дата доступа : 25.09.2012.

4 Narodowy Korpus Języka Polskiego / pod red. A. Przepiórkowskiego [et al.]. – Warszawa: Wydawnictwo naukowe PWN, 2012.

5 Spousta, M. Victor: The Web-Page Cleaning Tool / M. Spousta, M. Marek, P. Pecina // Proceedings of the 4th Web as Corpus Workshop / ed. by S. Evert [et al.]. – Marrakesh, 2008. – P. 12–17.

УДК 811.163.3‘367.625+811.161.1‘367.625

І. В. Гліннік (Мінск)

### **КАНСТРУКЦЫІ З ДЗЕЯСЛОВАМІ *БЫЦЬ* І *МЕЦЬ* Ў АРЫГІНАЛЬНЫХ І ПЕРАКЛАДНЫХ ТЭКСТАХ**

**(НА МАТЭРЫЯЛЕ ПАРАЛЕЛЬНАГА  
БЕЛАРУСКА-РУСКАГА КОРПУСА)**

*Аб’ектам даследавання сталі канструкцыі з дзеясловамі быць і мець у арыгінальных і перакладных тэкстах мастацкай літаратуры паралельнага*

*беларуска-рускага корпуса. Паралельны корпус тэкстаў дазволіў выявіць поўныя, частковыя эквіваленты, а таксама несінанімічны пераклад.*

Агульнавядома, што слоўнік любой мовы ўключае збор патэнцыяльнай лексікі, якая рэалізуецца толькі часткова. Параўнанне слоўнікаў дзвюх або некалькіх моў дае агульнае ўяўленне пра іх лексічную сістэму, якая аформілася ў працэсе гістарычнага развіцця. На падставе аналізу слоўнікавага матэрыялу даследчыкі не могуць зрабіць высновы пра частотнасць ужывання той ці іншай лексемы, пра значэнні, якія яна набывае ў пэўным кантэксце, пра спалучальнасць з іншымі словамі і г. д. Зусім іншая карціна паўстае пры аналізе тэкстаў той ці іншай мовы, бо ў такім выпадку мы можам гаварыць пра «частку» жывой мовы, таму што кожны тэкст – «яўляецца мікрасістэмай» [1, с. 231] пэўнай мовы. Вывучэнне мікрасістэмы дае веды пра макрасістэму.

Сёння даследаваць праблемы макрасістэмы, якой з'яўляецца мова таго ці іншага народа, дапамагае корпусная лінгвістыка, якая хуткімі тэмпамі развіваецца ў апошні час і дазваляе эфектыўна рашаць самыя розныя лінгвістычныя пытанні.

Ужо многія мовы маюць свае нацыянальныя корпусы (англійская, руская, беларуская, польская, чэшская і г. д.). Вядомы таксама спробы стварэння паралельных корпусаў (руска-ўкраінскі, польска-ўкраінскі і г. д.). Не так даўно ў Інстытуце мовы і літаратуры імя Якуба Коласа і Янкі Купалы НАН Беларусі сумесна з расійскімі калегамі пачалася праца па стварэнні паралельнага руска-беларускага корпуса.

Безумоўна, рускія чытачы разумеюць арыгінальныя беларускія тэксты (а беларускія, у сваю чаргу, – творы рускіх пісьменнікаў), паколькі беларуская і руская мова блізкароднасныя. Аднак большасць чытачоў Расіі знаёміцца з тэкстамі беларускіх пісьменнікаў праз перакладчыка. І тут паралельны корпус будзе найлепшым памочнікам, бо чытач зможа не толькі прачытаць пераклад, але і параўнаць яго з арыгіналам, даведацца, наколькі поўна перададзены змест, мастацкія сродкі.

Асаблівую карысць і дапамогу корпуса, безумоўна, ацэняць лінгвісты, таму што толькі з яго дапамогай можна правесці грунтоўны аналіз тэкстаў. Чаму гэта так? Думаецца, што няма неабходнасці ў XXI стагоддзі даказваць, што камп'ютар дае магчымасць філолагу аналізаваць самыя розныя мадэлі моўнага ўжывання, дазваляе захоўваць вялікую базу дадзеных натуральнай мовы, намнога большую, чым можна стварыць уручную, забяспечвае дакладны і паслядоўны аналіз. Сёння корпусы выкарыстоўваюцца ў лінгвістыцы для аб'ектыўнага апісання мовы як паводле колькасных характарыстык, так і якасных. Гэта, у сваю чаргу, дае магчымасць складаць частотныя, шматмоўныя слоўнікі (у тым ліку і перакладныя), вызначаць эквіваленты перакладу і г. д.

Аб'ектам нашага даследавання сталі канструкцыі з дзеясловамі *быць* і *мець* у арыгінальных і перакладных тэкстах мастацкай літаратуры

беларуска-рускага корпуса. Калі размова ідзе пра пераклад навуковай ці тэхнічнай літаратуры, самае галоўнае для перакладчыка – дакладна перадаць усе нюансы сэнсу, пры гэтым не адступаць ад арыгінала. Але для мастацкіх тэкстаў гэтага недастаткова, бо перад перакладчыкам стаіць задача не проста данесці сюжэт, але і эмацыйныя ўражанні. Вядома, што ўжыванне сінанімічных сродкаў садзейнічае больш глыбокаму раскрыццю рэчаіснасці, стварае яркія мастацкія вобразы, дазваляе перадаць градацыю ў выражэнні думак, паняццяў [2, с. 144]. Выбар дзеяслоўных канструкцый невыпадковы, бо дзеяслоў, як адзначае К. Паляньскі, «займае галоўную пазіцыю ў сказе і сваімі семантыка-граматычнымі ўласцівасцямі ў значнай ступені прадвызначае структуру сказа, у якім выступае» [3, с. 6].

Спосаб даследавання перакладаў тэкстаў мае пэўныя станоўчыя рысы: ён аб'ектывізуе вывады, бо пры перакладзе існуе задача захаваць адзіны семантычны інварыянт. Акрамя таго, мастацкі пераклад не ажыццяўляецца паслоўна, перакладчык улічвае шырокі кантэкст. А больш глыбокаму раскрыццю рэчаіснасці садзейнічае менавіта дзеяслоўная сінаніміка, якая дапамагае стварыць яркія мастацкія вобразы.

Спосабы перадачы сінанімаў можна падзяліць на дзве групы. Да першай адносіцца сінанімічны пераклад, а да другой – несінанімічны. У залежнасці ад ступені эквівалентнасці кожнага кампанента сінанімічнага рада сінанімічныя адпаведнікі падзяляюцца на:

1) поўныя эквіваленты [4], да якіх адносяцца сінанімы з тоеснай семантыкай, эмацыянальна-экспрэсіўнай і стылістычнай афарбоўкай: Тут же в передней *был стул*, и Вера опустилась на него. // Тут жа ў пярэднім пакоі *стаяла крэсла*, і Вера села (А. Купрын). Но *солнце было уже низко*, медленно садилось вдали за широкую багровую тучу над лесом, – в такое время, знала она, служащие в местечке расходились из учреждений и занимались своим хозяйством. // Але *сонца апусцілася ўжо нізка*, марудна садзілася ўдалечыні за шырокую хмару над лесам, – у такую пару, ведала яна, служачыя ў мястэчку разыходзіліся з устаноў і займаліся сваімі справамі (А. Купрын).

Як правіла, пры поўным эквівалентным перакладзе захоўваюцца і сродкі сувязі арыгінальнага твора. Часцей за ўсё паміж членамі сінанімічнага рада выкарыстоўваецца бяззлучнікавая або злучальная сувязь (пры дапамозе злучніка *і*).

2) частковыя эквіваленты: калі адпаведнік перадае толькі частку семантыкі арыгінальных сінанімаў: *Буду пытаць* у Кульшаў. // *Надо спросить* у Кульшей (У. Караткевіч). Значэнне будучага часу ў арыгінальным творы перадаецца як неабходнасць дзеяння пры перакладзе.

Яны выйшлі на двор *і, каб было бліжэй*, хутка пусціліся па мяжы ў поле. // Они вышли во двор *и, сокращая свой путь к деревне*, быстро пустились по меже полем (В. Быкаў). Даданы сказ мэты перакладаецца дзе-епрыслоўным зваротам спосабу дзеяння.

Безумоўна, трэба звярнуць увагу на тое, што пры частковай перадачы эквівалентаў адбываюцца пэўныя семантычныя страты ў межах кантэксту.

Як ужо адзначалася, акрамя сінанімічнага перакладу ёсць і несінанімічны пераклад (прадстаўлены заменай дзеяслоўных сінанімаў адной лексемай). У такіх выпадках перакладчык ужывае толькі адзін сінанім арыгінала або выкарыстоўвае зусім не тоесную лексему, толькі прыблізную па семантыцы, што і робіць тэксты несінанімічнымі: Стараста Бабініч *меўся паслаць* у школу да Лабановіча сваіх дзяцей – сына і дачку. // Староста Бабініч *собираўся* послать в школу к Лобановичу своих детей – сына и дочку (Я. Колас). Видите ли: всюду, где мы останавливались на постоях, городские жители *имели свои исключения* и прибавления, но в Букаресте так коротко обходились с нами жители, что когда однажды я стал играть на скрипке, то девушки тотчас нарядились и пришли танцевать, и такое обыкновение повелось на каждый день. // Бачыце, усюды, дзе мы спыняліся на пастой, гарадскія жыхары *наводзілі сябе па-рознаму*, але ў Букарэсце яны абыходзіліся з намі так проста, што калі аднойчы я пачаў іграць на скрыпцы, дык дзяўчаты хуценька апрануліся па-святочнаму і прыйшлі танцаваць, і гэтак павялося кожны дзень (А. Купрын).

Аналіз мастацкіх тэкстаў паказаў, што асабліва часта пры перакладзе замяняецца дзеяслоў *быць* і яго формы. Чаму так адбываецца і што абумоўлівае змену дзеяслова? Агульнапрынята, што асноўнае значэнне дзеяслова *быць* – экзістэнцыяльнае. Аднак блізкія да дзеяслова *быць* у гэтым значэнні сінанімы: бел. *існаваць, мець месца*, рус. *существовать, иметься* таксама выражаюць стан экзістэнцыі, г. зн. інфармуюць аб наяўнасці якога-небудзь рэальнага або нерэальнага быту, які ўтварае пэўны элемент акаляючага нас свету. Таму сустракаюцца выпадкі, калі перакладчык замяняе дзеяслоў *быць* або яго часавыя формы сінанімічнай лексемай, якая перадае экзістэнцыяльнае значэнне: Я прачынаюся і думаю, што не прайшоў ягоны час, пакуль *ёсць* цемра, голад, нераўнапраўе і цёмны жах на зямлі. // Я просыпаюся и думаю, что не прошло ее время, пока *существуют* мрак, голод, неравноправие и тёмный ужас на земле (Я. Колас).

Акрамя таго, рускія спалучэнні *может быть, должно быть* перакладаюцца беларускімі словамі *магчыма, можа, мусіць*: Князь Шейн, несмотря на своё видное положение в обществе, а *может быть*, и благодаря ему, едва сводил концы с концами. // Князь Шэйн, нягледзячы на сваё віднае месца ў грамадстве, а *можа*, і дзякуючы яму, ледзьве зводзіў канцы з канцамі (А. Купрын). – Вы, *должно быть*, смолоду очень красивы были. // – Вы, *мусіць*, у маладосці вельмі прыгожыя былі (А. Купрын).

Нельга не звярнуць увагі на тыя выпадкі, калі дзеяслоў *быць* перакладаецца дзеясловам *мець* (*тіе́с*). Случалось, што какой-нибудь заблудший подпоручик, присланный для долговременной отсидки из такого захолустья, где даже *не имелось* собственной гауптвахты, признавался, что он, по безденежью, довольствуется из солдатского котла. // Здаралася, што які-небудзь заблудны падпаручнік, якога прыслалі для доўгачасовай адседкі з такой глухамані, дзе нават *не было* сваёй гаўптвахты, прызнаваўся, што ён, з прычыны безграшоўя, забяспечваецца з салдацкага катла (А. Купрын).

Хоць дзед і *не меў асаблівай прыхільнасці* да гэтага занятку, але паколькі ён быў вольны і хацелася яму крыху прайсціся, дык і далучыўся да рыбаловаў. // *Хотя дед и не был особым любителем*, но поскольку делать ему было нечего и хотелось поразмяться, присоединился он к рыбакам (М. Лынькоў). Як правіла, дзеяслоў *мець* выкарыстоўваецца для выражэння адносін пасесіўнасці: Дзед Астап быў старым мікалаеўскім салдатам, службыў вартаўніком у дэпо, *меў два срэбныя медалі* за турэцкую вайну. // Дед Астап был из старых николаевских солдат. Служил он сторожем в депо. И *были у деда две серебряные медали* за турецкую войну (М. Лынькоў). Акрамя таго, даволі распаўсюджаны сказы, у якіх дзеяслоў *мець* утварае не толькі ўласна пасесіўныя структуры, якія звязаны з выражэннем адчужальнай прыналежнасці, але службыць і для ўтварэння апасесіўных структур, увогуле не звязаных з выражэннем пасесіўнасці: *Помяни моё слово, что лет через тридцать женщины займут в мире неслыханную власть*. // Успомніш маё слова, што год праз трыццаць жанчыны *будуць мець* нечуваную ўладу ў свеце. Дело, о котором мы *будем иметь честь* говорить с вами, одинаково касается и князя и меня, или, вернее, супруги князя, а моей сестры. // *Справа, пра якую мы будзем мець гонар* гаварыць з вамі, аднолькава тычыцца і князя і мяне ці, дакладней кажучы, князевай жонкі, а маёй сястры (А. Купрын).

Замена дзеяслова *быць* дзеясловам *мець* пры перакладзе абумоўлена тым, што пасесіўная канструкцыя ў рускай мове часцей за ўсё будзеца пры дапамозе пераходнага дзеяслова *быць*. А таму справядліва адносіць рускую мову да *esse-* моў. У беларускай мове паралельна выкарыстоўваюцца мадэлі з дзеясловам *мець*, *быць*, што дазваляе характарызаваць яе як мову змешанага тыпу, бо ў ёй прысутнічаюць тыпалагічныя рысы *habeo-* і *esse-* моў.

Трэба заўважыць, што сказаў, у якіх дзеяслоў *быць* перакладаецца на рускую мову сінанімічным дзеясловам або ўвогуле апускаецца, ці перадаецца апісальна, даволі шмат. Гэта адбываецца асабліва тады, калі дзеяслоў *быць* ужываецца ў спалучэнні з безасабова-прэдыкатыўнымі словамі. Улічваючы, што гэта не даслоўны, а мастацкі пераклад, можна дапусціць такую пазіцыю перакладчыка, бо сэнс захоўваецца: Анна вся *состояла* из весёлой безалаберности и милых, иногда странных противоречий. // Ганна ўся *была поўна* вясёлай бязладнасці і мілых, іншы раз дзіўных супярэчнасцей. Я *нашла* только один переплёт, остальное всё *пришлось* придумывать – листочки, застёжки, карандаш. // Праўда, я *знайшла* толькі адну вокладку, астатняе ўсё *трэба было* прыдумаць – лісточкі, засцежкі, аловак.

Нельга пакінуць па-за ўвагай перадачу фразеалагізмаў, якія адлюстроўваюць спецыфіку кожнай мовы. Менавіта пры іх перакладзе «*вскрывается* свойственная данному языку специфичность сочетаний, которая иначе могла бы и не быть замечена» [5, 192]. У большасці выпадкаў мы назіраем эквівалентную перадачу фразем: Мне хацелася б і павучыцца ў вас сямутаму, бо вы ўжо *набілі*, як кажуць, *руку і маеце практыку*. // Мне хотелось бы и поучиться у вас кое-чему, ведь вы уже *набили*, как говорится, *руку и имеете опыт...* (Я. Колас).

Але, нягледзячы на блізкасць лексічнага і фразеалагічнага саставу славянскіх моў, такія якасці ўстойлівых выразаў, як вобразнасць і эмацыянальнасць, нацыянальны каларыт, асаблівасці спалучэнняў слоў складаюць значныя цяжкасці пры перакладзе [6].

Такім чынам, папярэдні аналіз перакладных мастацкіх тэкстаў на базе паралельнага беларуска-рускага корпуса (бо праца ў гэтым кірунку толькі пачалася) паказаў, што ў большасці выпадкаў пры перакладзе на рускую мову і наадварот дзеясловы перадаюцца эквівалентнымі сінанімічнымі адзінкамі, якія маюць тоесную семантыку. Такую заканамернасць можна растлумачыць тыпалагічнай блізкасцю дзвюх моў. Аднак сустракаюцца выпадкі значнага разыходжання з арыгіналам як з боку зместу, так і формы, што, магчыма, абумоўлена асабістай ініцыятывай перакладчыка. А гэта вядзе да даволі значных страт у семантыцы.

#### СПІС ЛІТАРАТУРЫ

1 Гюлумянц, К. М. К вопросу о переводе польских фразеологизмов на русский язык / К. М. Гюлумянц // Веснік Беларускага дзяржаўнага ўніверсітэта імя У. І. Леніна. – Мінск: Выдавецтва БДУ імя У. І. Леніна, 1972. – № 2. – С. 49–53.

2 Лапцёнак, А. М. Надзённыя праблемы лексікалогіі і анамастыкі славянскіх моў / А. М. Лапцёнак // Матэрыялы II Міжнар. навук. канф., г. Мазыр, 22–23 красавіка 2010 г. / рэдкал.: В. В. Шур (адк. рэд.) [і інш.]. – Мазыр: УА «МДРУ імя І. П. Шамякіна», 2010. – С. 144–146.

3 Сцяцко, П. У. Слоўнік лінгвістычных тэрмінаў / П. У. Сцяцко, М. Ф. Гуліцкі, Л. А. Антанюк; пад рэд. М. В. Бірылы. – Мінск: Выш. школа, 1990.

4 Фёдоров, А. Ф. Основы общей теории перевода / А. Ф. Федоров. – М., 1968. – С. 192.

5 Siatkowska, E. Perspektywy leksykalnych badań porównawczych w oparciu o paralelne teksty / E. Siatkowska // Z polskich studiów slawistycznych. Seria X. Językoznawstwo. – Warszawa, 2002.

6 Słownik syntaktyczno-generatywny czasowników polskich. T. 1. A-M / Oprac. Zespołowo pod red. K. Polańskiego. – Wrocław etc: Ossolineum, 1980.

УДК 004:37.091.3:81'373.423

А. И. Головня (Минск)

#### **КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ ПРИ ИССЛЕДОВАНИИ И ПРЕПОДАВАНИИ ГРАММАТИЧЕСКОЙ ОМОНИМИИ**

*В статье рассматриваются основные компьютерные программные средства для работы с грамматическими омонимами. Установлена предпочтительность начала анализа словосочетаний и предложений с неомонимичных форм с учетом управления глаголов и предлогов. Выявлено свойство грамматического кода, позволяющее с учетом глагольного и предложного управления снимать грамматическую омонимию.*



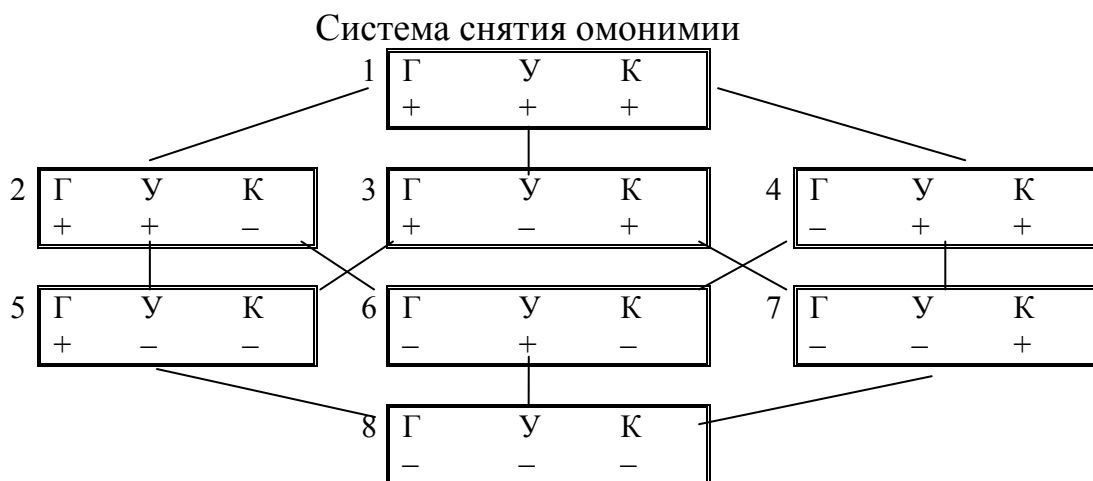
Корпус текстов русского языка, которым располагает Интернет, – это мощный лингвистический ресурс, предназначенный для научных исследований и преподавания русского языка и литературы, риторики, стилистики и других дисциплин. Корпус стал необходимым и обязательным инструментом современного филолога и используется максимально эффективно, если преподаватель ясно представляет принципы его устройства и функциональные возможности [2; 3; 6; 7; 8; 9].

Большинство текстов корпуса русского языка размечены *морфологически* – это означает, что в тегах указана лексема (словарная форма), которой принадлежит данная словоформа, информация о части речи и набор грамматических признаков. Если в корпусе используется *семантическая* разметка, то теги, сопровождающие словоформу, содержат сведения о семантике слова.

Размеченный (морфологически проанализированный) текст позволяет получать ответы на запросы преподавателя и помогать ему составлять задания для студентов.

Омонимия является одной из самых сложных тем, особенно для студентов иностранцев. Пронизанность всей системы языка омонимией на разных уровнях (достаточно указать, что в русском языке нет ни одного существительного, которое бы не имело омонимичных форм внутри собственной парадигмы) позволяет выяснить всевозможные разнородные количественно-качественные разновидности и типы омонимии и описать системные способы ее снятия. Это можно сделать с помощью наполнения трехмерных кубов, которые являются своеобразными экспертными системами.

Исследователи этой проблемы отмечают три основных момента, вытормаживающих омонимию в письменных текстах: контекст, графику, ударение [1, с. 92]. Но три этих способа в комбинаторике дают уже полную систему из 8-ми вариантов снятия омонимии: графика (Г), ударение (У), контекст (К). Для снятия грамматической омонимии нами для наглядности и лучшего усвоения материала используется трехмерный куб В. А. Карпова [4].



Подсистема 8 – обязательная ниша, где: 1) омонимия не снимается; 2) снимается другими способами; 3) находятся неомонимичные слова.

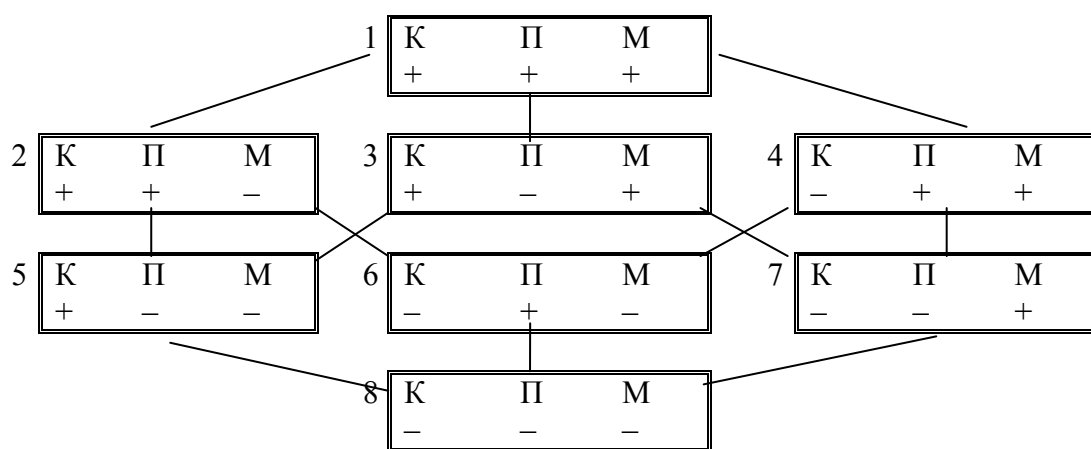
Подсистема 5 представляет омонимы, различающиеся графикой: частично различие имен нарицательных и собственных в середине и в конце предложения или словосочетания (но не в начале).

Подсистема 6 содержит омонимы, различающиеся ударением. Этот активно действующий способ различает: вид глагола (*вЫсыпАть*, *нарЕзАть* и т. п.), падеж и число (*трАвы*, *вОды* и т. п.), семантику (*мУка*, *стОйка* и др.) и другие категориальные признаки. Отметим, что мы имеем в виду ударение в письменных текстах.

Подсистема 7 представляет контекст, вытормаживающий омонимию. Это может быть как собственно окружение слова, так и словарь предметной области, ограничивающий омонимию. Так, омонимы *хлопок* и *хлопок* вряд ли попадают в рамках одной предметной области, а значит, и конкретного текста.

Подсистемы 2, 3, 4 и 1 представляют комбинаторные способы снятия омонимии. Таким образом, мы получили 7 способов снятия омонимии, связанные с семью же способами развития омонимии в системе.

Теперь представим детально, как работает ударение. Построим трехмерный куб по трем исходным компонентам: класс (часть речи) обозначим через «К», позицию через «П» и корневую матрицу через «М».



1. Различаются: *ОтрУб*, *засЫпАть*, *зАвАнский*; не различаются: *полька*, *финка* и подобные;

2. Различаются: *АтлАс*, *УжИн*; не различаются: *однотонный*, *засыпать* (спать) и *засыпать* (сыпать) и т. п.;

3. Различаются: *трАвы:травЫ*, *сЕлите:селИте*; не различаются: *мышь*, *лирика*, *дали*, *ели* и многие другие словоформы;

4. Различаются: *вОлОк*, *мАстерскАя*; не различаются: *запевала*, *часовой*, *кондитерская* и т. п.;

5. Различаются: *нАрЫ*, *чУМА*; не различаются: *кода*, *бука* и т. п.;

6. Различаются: *дОрОги*, *намОлОт*; не различаются: *трону*, *распил*, *пух*, *надел*, *простой* и др.;

7. Различаются: *пОдАть, прОпАсть; пОчЕсть; дОмА*; не различаются: *тепло, зло, добро, отцов, трут* и т. п.

8. Различаются: *сОльЮ, пОлзАла*; не различаются: *при (при и переть), жало (жалить и жать)* и другие.

Как видно из примеров, ударение частично снимает, частично не снимает омонимию во всех 8-ми подсистемах, что говорит о том, что, кроме ударения, должны существовать и другие способы снятия омонимии. Если рассмотреть омонимию с точки зрения порождения и восприятия текста, то можно отметить следующие моменты: а) для говорящего и пишущего омонимии нет, если автор специально не хочет использовать омонимы в каких-то своих целях (двусмысленность, шутка и т. п.); б) для слушающего и читающего омонимии почти нет, так как ударение, контекст снимают практически все омонимичные ситуации; в) при машинной обработке текстов и при машинном переводе омонимия возрастает, так как при двух системных омонимиях возникает еще и межсистемная омонимия.

Использование естественного языка в интеллектуальных системах ставит перед исследователями и разработчиками множество серьезных задач. Одной из таких задач является проблема омонимии от уровня словоформ до уровня синтаксиса. При обработке словоформы на входе ЭВМ это создает большие помехи и ставит задачу выбора словоформ.

Авторы монографии «Языковые средства автоматизированных информационных систем» (считающие, что в машинном словаре основы и окончания должны быть отдельными) отмечают: «Проверка основы и окончания слова на совместимость позволяет в основном преодолеть трудности морфологического анализа, связанные с омонимией основ слов. Однако при этом остается неразрешенной такая омонимия основ слов, которая может привести к совпадению некоторых форм различных слов. Например, у слов **техник** и **техника** совпадают несколько форм единственного и множественного числа, и вне контекста по одной форме слова нельзя определить, о каком слове идет речь. Такого рода омонимия может быть разрешена только средствами синтаксического анализа, а в некоторых случаях потребуются и семантический анализ контекста» [1, с. 153].

Межкорневая, или лексическая, омонимия частично снимается при создании словарей предметной области тем, что определенная часть таких омонимов распределяется по разным словарям (ср.: *проказа* шалость и *проказа* 'болезнь', *каток* 'место для катания' и *каток* 'машина' и т. п.).

При снятии грамматической омонимии на входе единственным на сегодня средством является анализ окружения омонимической словоформы и создание лакмус-контекстов, приводящих к однозначному выбору нужного слова [5].

Первым вопросом при анализе и преподавании грамматической омонимии должно быть выяснение количественной насыщенности текста

разными видами омонимов. Для определения того, насколько часто в художественном тексте отмечаются омонимичные формы и лексические омонимы. Мы со студентами провели эксперимент. В ходе его нами обработаны 10 страниц текста М. Булгакова «Мастер и Маргарита». Оказалось, что на 10-ти страницах текста из общего количества в 2 898 словоупотреблений отмечено 500 омонимичных словоформ, что составляет приблизительно 18 % всего отрывка.

При анализе выяснилось, что, как и ожидалось, наибольшая грамматическая омонимия отмечается у существительных (по Обратному словарю (ОС) их 56 332), затем у прилагательных (24 786) благодаря их объемам в словаре и в тексте и развитым парадигмам. Глаголы (37 319) внутри парадигмы омонимичны не настолько и не проявили межклассовых омонимичных свойств в анализируемом отрывке. Чаще в тексте встречается омонимия именительного и винительного падежей (180 случаев). Это закономерное явление: по данным словаря Э. А. Штейнфельдт именительный (33,6 %) и винительный (19,5 %) в сумме занимают почти 53 процента всех употреблений существительных [4, с. 52]. В тексте эта омонимия снимается, в основном, предлогами (так как им. п. не употребляется с предлогами), глагольным управлением и согласованием прилагательного с существительным, например: *...вернулся на балкон* – омонимия снимается управлением и предлогом *на*; *собирал совещание, утверждает приговор* – здесь омонимия снимается только управлением и *объяснение было странное* (глагол *быть* практически не сочетается с существительными в винительном падеже, за исключением существительных со значением времени или эллиптических конструкций – *Булочку будешь?* (есть).

На втором месте находится омонимия родительного и винительного падежей, чаще всего она связана с одушевленностью. Отмечено 65 случаев, что также укладывается в данные Э. А. Штейнфельдт (родительный 24,6 %) и винительный (19,5 %) в сумме дают свыше 44 % всех употреблений. Снимается эта омонимия: а) предлогами, которые управляют или родительным, или винительным падежами; б) глагольным управлением и в) отрицанием (родительный падеж чаще употребляется с отрицанием, чем винительный).

Следующий тип – это омонимия дательного и предложного падежей (дательный до 5, предложный до 9 %). Комбинаторика остальных падежей менее омонимична.

Второй по количественному составу является омонимия прилагательных (176 случаев). Но здесь омонимия снимается согласованием прилагательного с существительным определенного рода в определенном падеже и числе.

Омонимия местоимений и несклоняемых местоименных прилагательных оказалась редкой (18 случаев) и снимается контекстом предложения.

Остальные омонимичные пересечения классов (частей речи), грамматических позиций и семантик на данном объеме отмечаются в единичных случаях.

Наличие в русском языке хорошо развитых парадигм у существительных, прилагательных, местоимений и глаголов делает проблему межпарадигматической омонимии предельно важной, так как существуют своеобразные «узлы», связывающие парадигмы как одного, так и разных классов: *пол:пола:поле:полоть:полить:Поль: Поля:Пол*. Естественно, что в языках с меньшей развитостью грамматической аффиксации такой тип омонимов будет встречаться реже, уступая первенство лексической омонимии.

Это должно настраивать исследователя омонимии русского языка на достаточно надежный текстовый материал. С этой целью мы расширили объем эксперимента и взяли уже пять разных авторов (по 16 полным произвольно взятым страниц). Этим мы устранили возможность случайности тенденции, описанной выше на тексте М. Булгакова. Сплошная выборка на общем объеме около 23 тысяч словоупотреблений выявила 7 303 омонимичных формы. Подтвердилась общая количественная тенденция: при увеличении объема выборки в 10 раз число отмеченных омонимичных слов возросло почти в 15 раз.

Качественные распределения (части речи и совпадающие грамматические позиции) также сходны с вышеописанными. Чаще всего отмечена омонимия форм прилагательных (родительный, дательный, творительный, предложный – у прилагательных женского рода) и существительных (именительный: винительный и родительный: винительный). Эти совпадения отмечены в количествах, превышающих 100 примеров. Остальные совпадения отмечаются уже меньше, доходя до единичных примеров.

Наиболее надежным способом снятия омонимии в двух языках в плане грамматическом и лексическом является, по нашему мнению, системное кодирование с учетом управления [4; 5].

Так, если в коде глагольного управления глаголов *видеть* и *стоять* будет упакована информация о том, что первый управляет вин. п., а в коде второго такой информации не зашито, то обращение к коду управления позволит в композиции *вижу стол* присвоить слову *стол* код вин. п., а в композиции *стоит стол* – код им. п., что обеспечит на выходе правильное перекодирование и их смысловую обработку.

Аналогичное кодирование прилагательных, предлогов, местоимений, числительных, учитывающее согласование, управление и примыкание как основные типы связей, позволяет надеяться на успешное решение этой проблемы, находящейся на данном этапе в стадии научной проработки.

В рамках работ по использованию языка в интеллектуальных системах В. А. Карповым был разработан такой код на базе симметрии / асимметрии, реализованный И. В. Совпелем при создании словаря грамматических форм с кодами. Этот теггированный словарь позволяет выявить практически все омонимичные словоформы и создать многотысячный

словарь грамматических омонимов. Но проблема снятия омонимии в автоматизированных переводных интеллектуальных системах не снимается даже при наличии такового, так как один словарь грамматических омоформ, не учитывающий валентную сочетаемость глаголов и предложное управление не позволяет снимать омонимию при автоматизированном переводе с одного языка на другой без участия человека.

На наш взгляд, анализ словосочетаний и предложений надо начинать с неомонимичных форм, учитывая управление глаголов и предлогов.

Следовательно, грамматическое кодирование с учетом глагольного и предложного управления при наличии толково написанных программ и при детальной проработке неомонимичных форм, способных снять омонимию с помощью глагольного или предложного управления на входе и выходе ЭВМ, в какой-то мере смогут снимать грамматическую омонимию, хотя на сегодняшний день эта проблема без участия человека до или после автоматизированного перевода не решена.

Надеемся, что размеченные морфологически тексты национального корпуса русского языка позволят более полно и широко исследовать грамматическую омонимию и помогут найти более реальные способы ее снятия в переводных экспертных системах.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Белоногов, Г. Г. Языковые средства автоматизированных информационных систем / Г. Г. Белоногов, В. А. Кузнецов. – М.: Наука, 1983.
- 2 Гришина, Е. А. Мультимедийный русский корпус (МУРКО): проблемы аннотации / Е. А. Гришина // Национальный корпус русского языка: Новые результаты и перспективы. – М.: Нестор-История, 2009. – С. 175–214.
- 3 Добрушина, Н. Р. Корпусные методики обучения русскому языку / Н. Р. Добрушина // Национальный корпус русского языка: Новые результаты и перспективы. – М.: Нестор-История, 2009. – С. 335–351.
- 4 Карпов, В. А. Язык как система / В. А. Карпов. – М., 2003.
- 5 Колесников, Н. П. Синтаксическая омонимия в простом предложении / Н. П. Колесников. – Ростов, 1981.
- 6 Кустова, Г. И. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы / Г. И. Кустова, О. Н. Ляшевская, Е. В. Падучева, Е. В. Рахилина // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. – М: Индрик, 2005. – С. 155–174.
- 7 Ляшевская, О. Н. О морфологическом стандарте Национального корпуса русского языка / О. Н. Ляшевская, В. А. Плунгян, Д. В. Сичинава // Национальный корпус русского языка: Новые результаты и перспективы. – М.: Нестор-История, 2009. – С. 111–135.
- 8 Савчук, С. О. Обучающий корпус русского языка / С. О. Савчук, Д. В. Сичинава // Национальный корпус русского языка: Новые результаты и перспективы. – М.: Нестор-История, 2009. – С. 317–334.
- 9 Национальный корпус русского языка [Электронный ресурс]. Режим доступа: <http://www.ruscorpora.ru>. – Дата доступа: 28.09.2011.

Ю. Д. Голяк (Минск)

## **КОНЦЕПЦИЯ РАЗРАБОТКИ ЭЛЕКТРОННОГО СЛОВАРЯ ДЛЯ СЕМАНТИЧЕСКИХ СЕТЕЙ И ПАУТИН В КОМПЬЮТЕРНЫХ СИСТЕМАХ (НА БАЗЕ РУССКОГО ЯЗЫКА)**

*В докладе освещена проблема разработки специфического электронного семантического словаря, который может стать основой для создания компьютерной семантической сети как модели представления информации в автоматизированных системах. Предлагается алгоритм дистрибуции значений многозначных слов для построения Семантических сетей на основе этого словаря и принцип его аннотации (тэгирования).*

Работа современных электронных переводчиков, справочных, экспертных, поисковых и иных систем тесно связана с процессом анализа текста на естественном языке на всех его уровнях: от морфологического до самого сложного – семантического.

Такие автоматизированные системы нуждаются в как можно более полной словарной базе как основе работы системы с естественным языком. Такой словарной базой может стать электронный словарь, размеченный особым образом – тэгированный – то есть, снабжённый специальными метками (тэгами), определяющими основные характеристики каждого слова или относящими его к определенной группе.

Аннотированные тэгами электронные словари уже не новы в современной компьютерной лингвистике, а обработка текста на грамматическом уровне проводится достаточно точно. Однако говорить о настоящем «понимании» системой текста можно лишь в том случае, когда ей доступны смысл, содержание, а это автоматически ставит в центр проблемы уровень семантического анализа.

Изучение семантики слова является одной из сложнейших задач современной компьютерной лингвистики, ведь лексическое значение – наиболее сложно формализуемый аспект существования слова в языке. Лексическое значение – самая «субъективная» из всех характеристик слова: она основана на условной договоренности понимать под определенным набором символов некоторый смысл или несколько смыслов.

Таким образом электронный словарь в основе высокоуровневой автоматизированной системы должен содержать не только парадигму слова, грамматические сведения о нём, но и семантические данные – значение слова, информацию о его лексической сочетаемости, место в иерархонеиерархических системах внутри лексического фонда. Однако и этого недостаточно для полноценного функционирования слова, если мы будем рассматривать его как составляющую графа Семантической сети.

Семантическая сеть – это информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними. Таким образом, семантическая сеть является одним из способов представления знаний. Сеть в математике представляет собой разновидность графа. В Семантической сети роль вершин графа выполняют понятия базы знаний, а дуги задают отношения между ними.

Для корректного функционирования систем, основанных на Семантических сетях и паутинах необходимо предельно точно определить значение каждого слова в базе и его сочетаемость. И здесь особенно остро встает вопрос дистрибуции значений многозначного слова или проблема полисемии. Провести строгую дистрибуцию значений слова и формализовать эти различия – задача достаточно сложная, поэтому во многом целью данного исследования стал поиск алгоритма, способного развести значения слова, опираясь на формальные показатели – семантические маркеры – внутри предложения.

На примере многозначного глагола *лежать* в русском языке мы постараемся продемонстрировать, что возможно выделить в предложении грамматические и семантические показатели, которые позволят при автоматическом анализе текста с достаточно высокой точностью отнести конкретный случай употребления глагола к тому или иному значению, а значит определить точную позицию слова в компьютерной базе знаний, представленной в виде Семантической сети.

Представим, что существует такая Семантическая сеть для конкретного языка (или нескольких языков), опирающаяся на базу слов языка (тэгированный электронный словарь). Каждое из слов будет промаркировано как можно большим количеством семантических маркеров (тэгов), показывающих отнесенность слова к той или иной семантической группе. Тогда при анализе конкретного слова внутри фразы/запроса/предложения система сможет определиться в его значении исходя из принадлежности тэгов, закрепленных за словами, окружающими искомым словом. Тэг, таким образом, показывает принадлежность слова к конкретной лексико-семантической группе. Тогда каждому из значений слова в базе значений будут соответствовать тэги-маркеры, которые и позволят определить значение слова в конкретном предложении. Помня о том, что в сочетании с семантическими маркерами используются также и грамматические (например, указание на то, каким падежом управляет глагол), вероятность правильного определения значения довольно высока.

Итак, для конкретного примера мы рассмотрели значения глагола *лежать* по толковым словарям русского языка (Таблица 1). Опираясь на собранную базу случаев реального употребления глаголов в текстах, мы выделили еще несколько не указанных в словарях, но функционирующих в языке значений.



Таблица 1 – Значения глагола *лежать* в русском языке по Словарю современного русского литературного языка Института языкознания АН СССР

Значение по словарю	Управление	Пример
1а. Находиться в горизонтальном положении, быть распростёртым всем телом на чём-либо. О человеке, животном.	На, в+П.п.	Я лежал на диване, устремив глаза в потолок.
1б. Будучи больным, находиться в постели	На, в+П.п. С+Тв.п	Лежал он больной три месяца.
2. Находиться, быть помещённым на какой-то поверхности горизонтально, широкой частью. О предметах.	На, в+П.п.	Перед ним на тарелке лежал рябчик.
3. Находиться, храниться где-нибудь	В+П.п.	А золото спокойно в сундуке лежит себе.
4. Иметь место положение, простираться, быть расположенным	На, в+П.п. перед+Тв.п.	Перед ним теперь лежало широкое поле.
5. Проходить, пролегать (о пути, дороге)	На, в+П.п. На+В.п.	И лежит нам путь далёк.
6. Находиться на чьей-либо ответственности, быть обязанностью	На+П.п.	На нём лежала обязанность смотреть за стрелками и за путями.
7. Быть погребённым, покоиться	В+П.п.	Владимир Ленский там лежит, погибший рано смертью смелых.
8. Находиться, имея целью отдых	На, в+П.п.	Он часами лежал на пляже.

Затем необходимо было определить грамматические и семантические маркеры для каждого значения и построить минимальные семантические сети для этого глагола, представленные в виде центр-ориентированного графа, где вершины представляют семантические и грамматические маркеры для каждого значения глагола, а ребра связывают сам глагол и маркеры отношениями значения, номера которых на ребрах соответствуют номерам в таблицах.

Таблица 2. – Корреляция значений глагола и формальных показателей (маркеров)

№	Значение	Формальный показатель
Глагол «лежать» в русском языке		
1а	Находиться в горизонтальном положении, быть распростёртым всем телом на чём-либо. О человеке, животном.	На, в+П.п// слово ЛСГ «мебель», «поверхность», «часть тела»
1б	Будучи больным, находиться в постели	С+ слово ЛСГ «болезнь, симптом», «больница»
2	Находиться, быть помещённым на какой-то поверхности горизонтально, широкой частью. О предметах.	На, в+П.п// слово ЛСГ «мебель», «поверхность», «часть тела»
3	Находиться, храниться где-нибудь	В+П.п//ЛСГ «вместилище»
5	Иметь место, положение, простираться, быть расположенным	ЛСГ «географический объект или рельеф», «покров (в т. ч. в переносном значении)»
5	Проходить, пролегать (о пути, дороге)	ЛСГ «дорога»; На+В.п//ЛСГ «стороны света»
6	Находиться на чьей-либо ответственности, быть обязанностью	На+П.п//ЛСГ «человек», «ответственность»
7	Быть погребённым, покоиться	ЛСП «смерть» (наиб. характерно в роли подлежащего)
8	Находиться, имея целью отдых	на+П.п//ЛСГ «место отдыха»

Приведем пример такого графа:



Рисунок 1 – Семантический центр-ориентированный граф

Соответствия «значения-маркеры» стали основой для создания программы-семантического анализатора, задачей которой и было определение значения глагола *лежать* в конкретном, вводимым пользователем с клавиатуры, предложении, с чем она успешно справляется, хотя существующие на данный момент базы слов с присвоенными им тэгами ограничены (Рисунок 2).

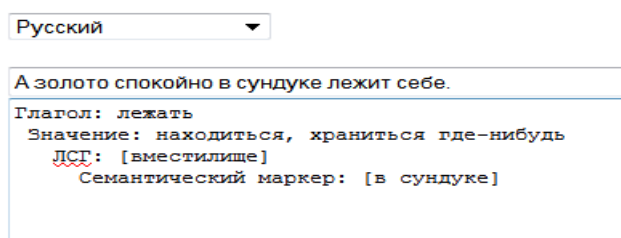


Рисунок 2 – Интерфейс программы-семантического анализатора

Таким образом, можно утверждать, что задача автоматизированного определения значения слова, как и общелингвистический вопрос дистрибуции значений многозначного слова, могут быть решены описанным способом корреляции значения с определенными семантическими и грамматическими маркерами. Присвоение этих тегов-маркеров каждому слову – задача трудоемкая, требующая больших временных затрат и труда экспертов, однако в долгосрочной перспективе работа над таким объемным и подробным анотированным электронным словарем обеспечит естественно-языковую базу для сложнейших интеллектуальных систем. Со временем же можно ожидать, что процесс присвоения тэгов автоматизируется благодаря потенциальной «самообучаемости» такого рода компьютерных систем, а создание подобных словарных баз на разных языках приблизит нас к реализации глобальной идеи о создании единой международной Семантической сети.

## СПИСОК ЛИТЕРАТУРЫ

- 1 Зубов, А. В. Информационные технологии в лингвистике / А. А. Зубов, И. И. Зубова. – М., 2004.
- 2 Шаров, С. А. Средства компьютерного представления лингвистической информации / С. А. Шаров. – М., 1996.

УДК 81'322.4

Т. А. Грибач (Гомель)

### **СИСТЕМЫ МАШИННОГО ПЕРЕВОДА, ОСНОВАННЫЕ НА ПАРАЛЛЕЛЬНЫХ КОРПУСАХ ТЕКСТОВ**

*В статье рассматривается проблема построения систем машинного перевода, основанных на параллельных корпусах текстов. В качестве примера приводится СМП PROMT DeepHybrid, представленная в июне 2012 г. в Монреале в рамках конференции американского отделения Ассоциации компьютерной лингвистики (NAACL 2012) на 7-ом семинаре по статистическому машинному переводу.*

За прошедшие полвека в области компьютерной лингвистики были получены значительные научные и практические результаты: созданы системы машинного перевода текстов с одних естественных языков на другие, системы автоматизированного поиска информации в текстах, системы автоматического анализа и синтеза устной речи и многие другие. Но были и разочарования. Например, проблема машинного перевода текстов с одних языков на другие оказалась значительно сложнее, чем это представляли себе пионеры машинного перевода и их последователи.

Сегодня активно разрабатываются системы машинного перевода, основанные на параллельных корпусах текстов.

Корпус – это информационно-справочная система, основанная на сборке текстов на некотором языке в электронной форме.

Параллельный корпус (Parallel Corpora) – это электронный аналог параллельных переводных текстов, как правило, состоящий из множества блоков "текст-оригинал и один/несколько его переводов". Электронные тексты в корпусе могут представлять собой целое оригинальное словесное произведение или какую-либо его часть.

За последнее десятилетие в мире было создано множество двуязычных/многоязычных корпусов, из них:

EUROPARL – 20.000.000 словоупотреблений, открытый корпус Европарламента на 11 языках (автор Philipp Koehn)  
<http://www.isi.edu/~koehn/publications/europarl/>

CHEMNITZ GERMAN-ENGLISH TRANSLATION CORPUS – 1.000.000 словоупотреблений (тексты – академические, политика, туризм)

<http://www.tu-chemnitz.de/phil/InternetGrammar>

KACENKA (Korpus anglicko-cesky; Czech) – 3.000.000 словоупотреблений <http://www.phil.muni.cz/angl/kacenska/kachna.html>

OPUS (an open source parallel corpus) – это попытка собрать переводы в Интернет, выровнять и разметить корпус, добавить лингвистическую информацию (5 языков). <http://logos.uio.no/opus/>

English-French Canadian Hansard (Brown et al, 1991; Gale and Church, 1991);

Lancaster's ITU, включающий английский, французский и испанский язык (McEnergy et al., 1997);

Англо-норвежский параллельный корпус (Johansson and Hofland, 1993);

Англо-китайский параллельный корпус HKUST, созданный в Гонконге (Wu, 1994);

Восточный многоязычный корпус, включающий болгарский, чешский, эстонский, венгерский, румынский и словенский языки (Erjavec et al., 1998);

Корпус "Agenda 21", включающий датский, английский, французский и немецкий языки (Kraaij, 1997);

Многоязычный корпус, включающий переводы Библии на английском, французском, датском, финском, греческом, латинском, шведском, испанском, вьетнамском языках (Resnik et al., 1999)

Англо-хорватский параллельный корпус (Tadic, 2000).

INTERSECT – англо-немецкий параллельный корпус (Salkie, 1995) [8].

Если в качестве примера взять Национальный корпус русского языка, то в настоящее время включает также параллельные корпуса, в которых можно найти все переводы для определенного слова или словосочетания на русский язык или с русского языка. В настоящее время для поиска доступны англо-русский, русско-английский, немецко-русский, украинско-русский, русско-украинский, белорусско-русский, русско-белорусский и многоязычный параллельные корпуса [7].

Параллельные корпуса текстов-образцов (в виде Базы Данных) особенно полезны в том случае, когда переводчик работает со строго нормированными (конвенциональными) текстами, жанрово-стилистическое и стилистическое оформление таких текстов практически не допускает варьирования, отступления от определенных социокультурных норм. Это тексты деловой переписки, тексты-рецепты, тексты-прогнозы погоды, тексты-контракты и т. д. Тексты различных стилей различаются как словарем лексических единиц, употребляемых в определенных текстах, так и грамматическими и синтаксическими структурами предложений, заключенных в них. Менее нормированные тексты допускают варьирование инвариантной модели. Параллельные корпуса текстов-образцов и их типологические модели-характеристики, составленные на этапе предпереводческого анализа исходного текста, могут служить для переводчика таким же эффективным вспомогательным средством.

Параллельные корпуса нужны при поиске эквивалента для перевода терминов. При быстром появлении терминов во многих научных, технологических или политических областях, создание современных терминологических справочников отстаёт, что вызывает проблемы у переводчиков и в результате – противоречивые и разнообразные переводы. Параллельные корпуса переведенных текстов могут использоваться как ресурс для автоматического извлечения терминов и терминологических словосочетаний. Здесь актуальны параллельные корпуса, основанные на научно-технических, политических, экономических и юридических текстах [19].

Между единицами оригинального и переводного текста (обычно — между предложениями) с помощью специальной процедуры устанавливается соответствие; эта процедура называется выравниванием, а тексты, соответственно, выравненными.

В настоящее время на сайте Национального корпуса русского языка размещены следующие параллельные выравненные корпуса: англо-русский, русско-английский, немецко-русский, украинско-русский, русско-украинский, белорусско-русский, русско-белорусский и многоязычный.

Действующие системы машинного перевода ориентированы на конкретные пары языков и используют, как правило, переводные соответствия либо на поверхностном уровне, либо на некотором промежуточном уровне между входным и выходным языком. Качество машинного перевода зависит от объема словаря, объема информации, приписываемой лексическим единицам, от тщательности составления и проверки работы алгоритмов анализа и синтеза, от эффективности программного обеспечения. Современные аппаратные и программные средства допускают использование словарей большого объема, содержащих подробную грамматическую информацию. Информация может быть представлена как в декларативной (описательной), так и в процедурной (учитывающей потребности алгоритма) форме.

Современный машинный перевод следует отличать от использования компьютеров в помощь человеку-переводчику. В последнем случае имеется в виду автоматический словарь, помогающий человеку быстрее подбирать нужный переводной эквивалент. Хотя и в том, и в другом случае компьютер работает вместе с человеком (переводчиком или редактором), в содержание термина «машинный перевод» входит представление о том, что главную, бóльшую часть работы по переводу и отысканию переводных эквивалентов и переводных соответствий машина берет на себя, оставляя человеку лишь контроль и исправление ошибок, в то время как компьютерный словарь в помощь человеку – это чисто вспомогательное средство для быстрого нахождения переводных соответствий; при этом, однако, в такого рода словарях в ограниченной степени могут быть реализованы и некоторые функции, присущие системам машинного перевода.

В практике переводческой деятельности и в информационной технологии различаются два основных подхода к машинному переводу. С одной стороны, результаты машинного перевода могут быть использованы для

поверхностного ознакомления с содержанием документа на незнакомом языке. В этом случае он может использоваться как сигнальная информация и не требует тщательного редактирования. Другой подход предполагает использование машинного перевода вместо обычного «человеческого». Это предполагает тщательное редактирование и настройку системы перевода на определенную предметную область. Здесь играют роль полнота словаря, ориентированность его на содержание и набор языковых средств переводимых текстов, эффективность способов разрешения лексической многозначности, результативность работы алгоритмов извлечения грамматической информации, нахождения переводных соответствий и алгоритмов синтеза. На практике перевод такого типа становится экономически выгодным, если объем переводимых текстов достаточно велик (не менее нескольких десятков тысяч страниц в год), если тексты достаточно однородны, словари системы полны и допускают дальнейшее расширение, а программное обеспечение удобно для пост-редактирования. Такого рода системы машинного перевода используются в организациях, потребности которых в оперативных и качественных переводах достаточно велики [10]. В основном сегодня говорят о существовании двух подходов к машинному переводу: статистическом и фразеологическом.

Статистический машинный перевод – это метод машинного перевода. Он использует сравнение больших объёмов языковых пар, так же как и машинный перевод основанный на примерах. Статистический машинный перевод обладает свойством «самообучения». Чем больше в распоряжении имеется языковых пар и чем точнее они соответствуют друг другу, тем лучше результат статистического машинного перевода. Статистический машинный перевод основан на поиске наиболее вероятного перевода предложения с использованием данных из двуязычных корпусов текстов. В результате при выполнении перевода компьютер не оперирует лингвистическими алгоритмами, а вычисляет вероятность применения того или иного слова или выражения. Слово или последовательность слов, имеющие оптимальную вероятность, считаются наиболее соответствующими переводу исходного текста и подставляются компьютером в получаемый в результате текст. В статистическом машинном переводе ставится задача не перевода текста, а задача его расшифровки. Мы предполагаем, что статья, написанная на английском языке, на самом деле является статьей написанной на английском, но текст зашифрован (или искажен шумом). При таком подходе становится понятно почему, чем дальше языки, тем лучше работает статистический метод, по сравнению с классическими подходами.

Более обоснованной является концепция фразеологического машинного перевода. Главным тезисом этой концепции является утверждение, что наименования понятий в текстах чаще всего обозначаются не отдельными словами, а словосочетаниями, и при переводе текстов с одного языка на другой в качестве основных единиц смысла следует использовать прежде всего фразеологические сочетания, выражающие понятия, отношения

между понятиями и типовые ситуации. Эта концепция реализована в виде системы RETRANS (системы русско-английского и англо-русского машинного перевода полиметаллических текстов) [6].

Работа статистических, так и фразеологических систем происходит в двух режимах: обучения и эксплуатации. В режиме обучения просматриваются параллельные корпуса текста и вычисляются вероятности переводных соответствий. Строится модель языка перевода. Тут же определяются вероятности каждого n-грамма. В режиме эксплуатации, для фразы из исходного текста ищется фраза переводного текста, так, чтобы максимизировать произведение вероятностей.

Статистические системы машинного перевода активно разрабатывались (и разрабатываются) компанией IBM. Благодаря ее разработкам, были созданы модели перевода IBM Model 1-5. Но наибольшую известность этот метод приобрел благодаря компании Google. Кроме переводчика Google существует еще ряд систем и библиотек, использующих статистический подход: Giza++; Moses; Pharaoh; Rewrite; BLEU scoring tool.

Компании-производители систем статистического машинного перевода: PROMT, Google, SDL Language Weaver, Microsoft, Asia Online, IBM.

Преимущества и недостатки систем машинного перевода, основанных на статистическом и на фразеологическом подходах, во многом совпадают. Однако огромным преимуществом первых является то, что обучение таких систем происходит без участия человека. Для статистического перевода не нужна дополнительная разметка корпусов текста, это значительно упрощает их построение. С другой стороны для качественного обучения и тех и других нужны значительные объемы параллельных текстов. Потому часто переводчики являются дополнительными сервисами поисковых систем (Google, Яндекс). На данный момент, статистические системы являются лидерами по соотношению цена / качество для всех систем машинного перевода [1].

Кроме того, можно выделить ещё несколько технологий машинного перевода, в частности:

Rule – based Machine Translation (RBMT, Машинный перевод, основанный на правилах).

Эта технология основана на словарной информации и анализе грамматических правил конкретных языков. Такие системы строятся на основе лингвистического описания двух естественных языков (двуязычных словарей и других баз данных, содержащих морфологическую, грамматическую и семантическую информацию), формальных грамматик и собственно алгоритмов перевода. Качество перевода зависит от объемов лингвистических баз данных (словарей) и глубины описания естественных языков, т. е., необходим учет максимального количества особенностей грамматической структуры как входного, так и выходного языка.

Существует два типа rule-based систем:

системы по типу Transfer – предполагают морфологический, синтаксический и семантический анализ текста на языке входа; преобразование в структуру выходного языка (Transfer); синтез текста на выходном языке,

системы по типу Interlingua – предполагают анализ входного текста в терминах метаязыка и синтез метаструктуры текста на выходном языке.

Преимущества RBMT-систем:

- синтаксическая и морфологическая точность,
- стабильность и предсказуемость результата,
- возможность настройки на предметную область.

Недостатки RBMT-систем:

- трудоемкость и длительность разработки.
- необходимость поддерживать и актуализировать лингвистические базы данных.

Компании-производители систем машинного перевода на основе технологии rule - based: PROMT, Systran, LINGUATEC

Hybrid Machine Translation (HMT, Гибридный машинный перевод)

В связи с тем, что технологии статистического машинного перевода и системы машинного перевода, основанные на правилах, имеют свои недостатки и сложности, а также достигли определенного предела в своем развитии, разработчики решений по МП рассчитывают на технологический прорыв за счет создания гибридной технологии перевода. Эта технология основана на совмещении этих двух методов. Такой подход позволяет взять сильные стороны обеих технологий.

Компании-производители систем машинного перевода на основе гибридной технологии: PROMT, Systran.

Компания PROMT ведет разработки в области применения статистических методов, результатом которых стала гибридная технология перевода PROMT DeepHybrid.

PROMT DeepHybrid использует RBMT-технологию и статистические методы для:

- автоматического построения словарных баз на основе корпусов параллельных текстов,
- формирования нескольких вариантов перевода – на уровне лексики и структуры предложения,
- постредактирования в автоматическом режиме,
- выбора лучшего (наиболее вероятного) перевода на основе Языковой модели.

Такой подход позволяет сохранить преимущества технологии rule-based (формирование синтаксически связанного и грамматически правильного текста, однородность терминологии), получить преимущества статистического МП (быстрая обучаемость, получение данных из параллельных корпусов в автоматическом режиме, гладкость текста).

Следует также отметить, что если статистический машинный перевод требует колоссальных объемов параллельных текстов, то при гибридной технологии можно обойтись сравнительно небольшими объемами.

Гибридную технологию можно использовать как в корпоративном сегменте, где уже могут быть готовые объемы параллельных текстов, так и



на онлайн-сервисах, предложив интернет-сообществу принять участие в создании параллельных корпусов в режиме онлайн.

Внедрение нового решения PROMT каждый раз реализуется в виде индивидуального проекта. Первые пользователи решения появились в 2011 году на американском рынке.

Преимущества PROMT DeepHybrid:

быстрая автоматическая настройка на основе Translation Memory заказчика;

терминологическая точность и единство стиля;

получение дополнительных полезных данных – глоссариев, двуязычных терминологических словарей [9].

Перспективы развития машинного перевода связаны с дальнейшей разработкой и углублением теории и практики перевода, как машинного, так и «человеческого». Для развития теории важны результаты сопоставительного языкознания, общей теории перевода, теории закономерных соответствий, способов представления знаний, оптимизации и совершенствования лингвистических алгоритмов. Новые и более эффективные словари с необходимой словарной информацией, строгие теории терминологизации лексики, теория и практика работы с подязыками помогут повысить качество перевода лексических единиц. Формальные грамматики, ориентированные на перевод, дадут возможность оптимизировать алгоритмы нахождения переводных соответствий в данной коммуникативной ситуации, которая может быть описана в рамках соответствующих прикладных теорий представления знаний. Наконец, новые возможности программирования и вычислительной техники также будут вносить свой вклад в совершенствование и дальнейшее развитие теории и практики машинного перевода.

#### СПИСОК ЛИТЕРАТУРЫ

1 Белоногов, Г. Г. Каким должен быть машинный перевод в XXI веке / Г. Г. Белоногов и [др.] // Перевод: традиции и современные технологии. – М.: Наука, 2002. – С. 56–59.

2 Всероссийский институт научной и технической информации (ВИНИТИ) [Электронный ресурс]. Режим доступа: <http://www2.viniti.ru/index.php>. Дата доступа: 19.08.2012.

3 Зубов, А. В. Информационные технологии в лингвистике / А. В. Зубов, И. И. Зубова. – М.: Издательский центр «Академия», 2004. – 208 с.

4 Марчук, Ю. Н. Компьютерная лингвистика / Ю. Н. Марчук. – М.: АСТ Восток – Запад, 2007. – 317 с.

5 Международная конференция по компьютерной лингвистике [Электронный ресурс]. Режим доступа: <http://www.dialog-21.ru/dialog2012>. Дата доступа: 15.07.2012.

6 Научно-техническая информация. Серия 2: Информационные процессы и системы. [Электронный ресурс]. Режим доступа: [http://elibrary.ru/title\\_about.asp?id=9678](http://elibrary.ru/title_about.asp?id=9678). Дата доступа: 20.07.2012.

7 Национальный корпус русского языка [Электронный ресурс]. Режим доступа: <http://www.ruscorpora.ru/corpora-other.html>. Дата доступа: 01.08.2012.

8 МИРЭСПЕРАНТО [Электронный ресурс]. Режим доступа: [http://miresperanto.narod.ru/pri\\_angla/mashinnyj\\_perevod.htm](http://miresperanto.narod.ru/pri_angla/mashinnyj_perevod.htm) Дата доступа: 30.07.2012

9 Официальный сайт компания Promt [Электронный ресурс]. Режим доступа: <http://www.promt.ru>. Дата доступа: 01.07.2012.

10 Переводчики для переводчиков [Электронный ресурс]. Режим доступа: <http://www.translationmemory.ru/info/news/1401.php>. Дата доступа: 20.07.2012.

11 Рябцева, Н. К. Информационные процессы и машинный перевод. Лингвистический аспект / Н. К. Рябцева. – М.: «Наука», 1986. – 169 с.

12 Соловьёва А. В. Профессиональный перевод с помощью компьютера / А. В. Соловьёва – М.: Питер, 2008. – 160с.

13 Соснина, Е. П. Параллельные корпуса в обучении языку и переводу / Е. П. Соснина // Гуманитарный факультет Ульяновского государственного технического университета [Электронный ресурс]. Режим доступа: [http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus\\_education\\_translation/](http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_education_translation/) / Дата доступа: 01.07.2012.

14 Biber, D. Corpus Linguistics / D. Biber, S. Conrad, R. Reppen. – Cambridge University Press, 1998. – 300 p.

УДК 004+378+8

О. Е. Елисеева (Минск)

## **ЭЛЕКТРОННЫЕ ВИДЕОКУРСЫ В ПРЕПОДАВАНИИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ СТУДЕНТАМ ФИЛОЛОГИЧЕСКИХ СПЕЦИАЛЬНОСТЕЙ И ПОВЫШЕНИИ КВАЛИФИКАЦИИ ПРЕПОДАВАТЕЛЕЙ**

*В работе рассматривается один из подходов к преподаванию компьютерных дисциплин студентам и преподавателям филологических специальностей, заключающийся в использовании заранее подготовленных видеоуроков (скринкастов). Данный подход как по оценке успеваемости, так и по отзывам учащихся показал свою эффективность.*

Актуальность эффективного использования современных информационных и, в частности, компьютерных, технологий в филологии в настоящее время никто не оспаривает. Как следствие этого – выделение компьютерного направления на филологических специальностях. Тем не менее, за указанной очевидностью скрывается масса вопросов. Перечислим лишь некоторые из них:

- Каким образом повысить эффективность работы филолога с помощью компьютера?
- Как измерить степень этой эффективности?
- В каких направлениях работы использование компьютера наиболее целесообразно?
- Как преподавать учебные курсы, связанные с использованием компьютерных технологий будущим филологам так, чтобы у них формировалось понимание перечисленных выше проблем, а также путей их решения?

– Каким образом повышать квалификацию преподавателей-филологов, основным предметом деятельности которых являются далеко не компьютерные технологии?

В дополнение к перечисленным вопросам достаточно остро стоит проблема эффективного распределения и перераспределения учебной нагрузки как студентов, так и преподавателей с учетом использования либо неиспользования компьютерной техники. Далеко не секрет, что помимо собственно аудиторной нагрузки, во внеаудиторной деятельности филологу приходится тратить достаточно много времени на чтение всевозможной литературы. В то же время освоение компьютерных программ тоже требует немалого времени. Указанная проблема отчасти решается в том случае, если будущий филолог, выпускник средней школы, в достаточной степени овладел компьютерной техникой. Однако практика преподавания курса «Основы информационных технологий» студентам филологического факультета показывает, что это далеко не всегда так. Более того, у студентов гуманитарных специальностей зачастую сформировано негативное отношение к компьютерным дисциплинам, которые, по их мнению, являются пустым времяпрепровождением. С сожалением, следует констатировать, что во многих случаях это действительно так, поскольку практические задания, выполняемые студентами на занятиях по информатике, традиционно практически никак не связаны с филологическими задачами.

Исходя из вышесказанного, в основу преподавания курса «Основы информационных технологий» студентам филологических специальностей необходимо положить прагматичный, деятельностно-ориентированный подход. Опираясь на то, что студенты в стенах школ должны были в достаточной мере овладеть навыками использования основного прикладного программного обеспечения, задания на занятиях по информатике в вузе должны опираться на конкретные филологические задачи, а также другие задачи, с которыми сталкиваются студенты в процессе обучения в вузе. Так, например, практика показала, что большинство студентов не достаточно хорошо понимают, как эффективно использовать инструменты Microsoft Office, чтобы оформить в соответствии со стандартами свои курсовые работы, подготовить презентацию к докладу и пр.

Если говорить о специфике подготовки филологов в белорусских вузах, следует также заметить, что многие из выпускников ориентированы на дальнейшую работу в средних школах. Это обстоятельство помогает осознать, что будущий преподаватель должен знать, как он сможет эффективно использовать компьютер в учебном процессе и организации педагогической деятельности. В связи с этим, на занятиях информатики вполне уместны задачи по созданию демонстрационных презентаций учебного характера, а также компьютерных тестов с использованием тех же инструментов Microsoft Office.

Аналогичные проблемы и задачи были выявлены и в процессе подготовки семинаров-практикумов для преподавательского состава

филологического факультета. Таким образом удалось сформировать содержательную составляющую учебных курсов по информационным технологиям. Так, в частности, для практического освоения инструментов Microsoft Office были выделены следующие разделы:

– Практика подготовки и оформления учебно-методических материалов в текстовом редакторе Microsoft Word.

– Практика эффективного использования электронных таблиц Microsoft Excel в филологических исследованиях.

– Практика создания мультимедиа-презентаций, электронных упражнений и тестов с помощью инструментов Microsoft PowerPoint.

Очевидно, что помимо инструментов MS Office в работе филолога используется и другое программное обеспечение. Так, в работе [1] анализировались возможности использования сети Интернет в филологической подготовке, а в рамках мастер-классов конференции «Компьютерная лингвистика: научное направление и учебная дисциплина» (Гомельский государственный университет им. Ф. Скорины, 2010 г.) старший преподаватель Гродненского государственного университета А. Ю. Станкевич продемонстрировала возможности обучению использованию другого программного обеспечения для языковых исследований.

Еще одним из немаловажных навыков современного филолога является владение гипертекстовой технологией. Это обусловлено двумя причинами. Во-первых, вся сфера информационных технологий неуклонно развивается в сторону расширения сетевых методов взаимодействия, хранения и обработки информации посредством глобальной сети Интернет. В сети Интернет уже сложились свои традиции и стандарты, которые тем или иным образом связаны с понятием гипертекста. Вся информация, размещаемая на веб-ресурсах, соответствующим образом размечается с использованием языков разметки, которые развиваются в сторону семантизации с той целью, чтобы и людям (пользователям Сети), и роботам (программам, функционирующим в Сети) было проще и быстрее найти ту или иную релевантную информацию. Разобравшись в особенностях гипертекстовой разметки текстов, филолог сможет лучше понять идеологию семантической разметки и таким образом стать активным участником процесса всемирной глобализации сетевых знаний.

Во-вторых, с каждым годом все больше и больше информации учебного характера размещается в сети Интернет. Уметь грамотно это делать – одна из задач молодого преподавателя языка и литературы.

Учитывая выше сказанное, в дополнение к приведенному выше перечню разделов учебных курсов по информационным технологиям был сформирован курс «Практика создания гипертекста».

Выше кратко были рассмотрены содержательные аспекты преподавания компьютерных дисциплин на филологических специальностях. Рассмотрим организационно-методологические аспекты. Практика работы с группами студентов и преподавателей филологического факультета показала, что

многим из них не всегда просто быстро сориентироваться в огромном объеме абсолютно новой для них и в терминологическом, и в технологическом плане предметной области, к коей относится информатика. Это, а не отсутствие логического мышления или, еще хуже, недостаток интеллекта, на наш взгляд, является главной причиной затруднений, возникающих у гуманитариев, когда они начинают работать с компьютером. Именно поэтому были опробованы разные способы организации работы, а именно: подготовка подробных описаний заданий в текстовом виде с использованием скриншотов (копий экранов); проведение практических занятий с использованием видеопроектора в режиме «смотри и повторяй за мной».

Однако самым эффективным оказалось использование видеоуроков (скринкастов). Содержательно скринкаст представляет собой практическое занятие, на котором преподаватель информатики показывает и рассказывает, как решить ту или иную задачу с использованием соответствующих прикладных программ. Например, одним из таких занятий является составление частотного словаря словоформ художественного произведения с помощью Microsoft Word и Excel. Все, что делает преподаватель, записывается в видеофайл с экрана компьютера вместе с речевыми объяснениями и комментариями.

Таким образом, была создана совокупность видеоуроков в рамках перечисленных выше курсов. Сначала видеоуроки использовались локально, распространялись непосредственно в учебной аудитории. В дальнейшем на основе созданного учебного материала был создан веб-ресурс, на котором в настоящее время все указанные курсы представлены в свободном доступе. С ними можно ознакомиться по адресу <http://it.lang-study.com>. Некоторые из видеоуроков используются также в специализированной группе «Электронное обучение, прикладная и компьютерная лингвистика» в социальной сети по адресу: <http://vk.com/compling>.

Ниже перечислены дополнительные преимущества предлагаемого подхода:

учащиеся (как студенты, так и преподаватели) получили возможность изучать компьютерные технологии в удобном для них индивидуальном темпе, не боясь допустить ошибку и не зная, как объяснить причины своих затруднений преподавателю информатики;

у учащихся появилась возможность выполнять задания в любое удобное время (что особенно актуально для преподавателей, загруженных своей учебной нагрузкой);

иностранные студенты, имеющие дополнительные трудности, связанные с овладением русского языка, получили возможность прослушивать и просматривать любую часть урока, не требуя от преподавателя повторить объяснение;

наличие ориентированных на студентов-филологов учебных курсов, доступных также и преподавателям-филологам, способствует лучшему пониманию и формированию межпредметных связей филологических специальностей: преподаватели знают, какими технологиями владеют студенты,

и могут ориентироваться на их возможности при формировании практических навыков работы в различных областях языкознания.

#### СПИСОК ЛИТЕРАТУРЫ

1 Елисеева, О. Е. Сеть Интернет как инструмент и средство изучения прикладной и компьютерной лингвистики / О. Е. Елисеева // Компьютерная лингвистика: научное направление и учебная дисциплина [Текст]: сборник научных статей. Вып. 1. – Гомель: ГГУ им. Ф. Скорины, 2010. – С. 196–200.

УДК 004.822:514

А. В. Зубов (Минск)

### **О ВОЗМОЖНОСТЯХ БЕЛОРУССКО-ИНОЯЗЫЧНЫХ ПАРАЛЛЕЛЬНЫХ КОРПУСОВ ТЕКСТОВ**

*Объектом исследования является система правил организации различного рода информации (морфологической, лексической) в текстах, позволяющей извлекать из параллельных тегированных иноязычно-белорусских текстов информацию, необходимую для проведения лингвистических исследований и совершенствования процесса преподавания белорусского, русского, английского и немецкого языков. Детально рассматривается структура белорусско-русского, белорусско-английского и белорусско-немецкого тегированных корпусов текстов.*

С 2006 года кафедра информатики и прикладной лингвистики Минского государственного лингвистического университета занимается созданием параллельных корпусов текстов. В 2010 году закончено создание трех параллельных белорусско-иноязычных корпусов текстов, каждый объемом в 300 000 словоупотреблений.

В белорусско-русский подкорпус текстов был включен 91 текст переводов на русский язык фрагментов произведений Якуба Коласа, Владимира Короткевича, Ивана Шамякина, Рыгора Бородулина, а также переводы на белорусский язык стихотворений Сергея Есенина и фрагментов из «Мастера и Маргариты» Михаила Булгакова.

В белорусско-английский подкорпус включены 72 текста, являющихся переводами на белорусский язык англоязычных публицистических и научных текстов и стихотворений.

Белорусско-немецкий подкорпус текстов содержит 31 текст переводов художественной немецкой литературы на белорусский язык и переводы белорусских публицистических и научных текстов на немецкий язык.

Основной особенностью этих текстов является то, что все словоупотребления этих текстов получили наборы морфологических тегов [1]. Они могут быть с успехом использованы для решения целого ряда

лингвистических и методических задач. Так, с опорой на эти тексты можно использовать [2; 3; 4]:

1. В лексикографии и лексикологии (для составления различных словарей, определения значений многоязычных слов, выявления ассоциативных связей слов в тексте, выявление терминов и терминологических словосочетаний и т. д.).

2. В грамматике (для определения частоты употребления грамматических морфем в текстах различного типа, выявления наиболее употребляемых типов словосочетаний и предложений, определения частоты употребления классов слов и т. д.).

3. В лингвистике текста (для дифференциации типов текста, создания конкордансов, выявления связи между предложениями в абзацах и между абзацами и т. д.).

4. При автоматическом переводе текстов (для поиска контекстов слов, имеющих несколько переводных эквивалентов терминологических и фразеологических словосочетаний в параллельных текстах и т. д.).

5. В учебных целях (для выбора цитат, отдельных произведений и примеров, используемых в процессе создания учебников и учебных пособий и т. д.).

Специалисты, активно использующие корпуса текстов для различных научных исследований, отмечают, что «возможность проводить собственные наблюдения при обеспечении их статистической достоверности не может не сказаться на организации учебного процесса и на роли в нем студента и преподавателя. Опора на корпус ставит также вопрос о балансе между творчеством и использованием готовых фраз в языке и заставляет задуматься о том, как можно использовать такие методы описания для интеграции семантической, прагматической и культурной информации в учебных материалах» [4, с. 66]. Для этого сейчас создаются учебные корпуса текстов (Learner Corpora).

Начиная с 2011 года, кафедра информатики и прикладной лингвистики занимается созданием параллельного тегированного русско-белорусского корпуса учебных текстов.

Среди двуязычных параллельных учебных корпусов текстов можно выделить:

1. Параллельные учебные корпуса переводов.
2. Параллельные учебные корпуса текстов для лингвометодических целей.

Первые из них содержат тексты какого-либо одного языка и их переводы на другой язык, сделанные студентами, получающими квалификацию «Переводчик». Здесь можно отметить, например, созданный в Ульяновском государственном техническом университете русский учебный корпус переводов RuTLC (Russian Translation Learner Corpus) [5]. Целью создания этого корпуса текстов явилось выявление у обучаемых наиболее типичных лексических, стилистических и грамматических ошибок, совершаемых ими в процессе перевода текстов с английского языка на русский. В результате анализа

английских текстов и их переводов на русский язык (общий объем около 1 млн. словоупотреблений) было обнаружено около 4 500 ошибок. При этом лексические ошибки составили 50 % от числа всех ошибок. Из них 33 % таких ошибок связано с выбором неверного значения слова в процессе перевода.

Ко второму виду учебных текстов (двуязычные учебные корпуса) относятся параллельные тексты, созданные на материале учебных текстов двух языков. Именно такой параллельный учебный русско-белорусский корпус учебных текстов и создается в Минском государственном лингвистическом университете. Необходимость его создания связана с тем, что в Беларуси два государственных языка – белорусский и русский. Соответственно, школьные учебники по всем предметам издаются на двух этих языках и являются переводами с русского на белорусский язык.

В процессе обучения белорусскому и русскому языкам, при подготовке учебников и учебных пособий возникает целый ряд лингвистических и методических проблем, связанных со спецификой этих языков. Их решению и будет способствовать создаваемый параллельный русско-белорусский корпус учебных текстов. В исследуемый корпус учебных текстов вошли тексты русских и белорусских учебников по семи школьным дисциплинам: «История Беларуси», «География», «Трудовое обучение», «Информатика», «Биология», «Обществоведение» и «Физика».

Предварительно был проведен анализ отобранных белорусских и русских учебников. В итоге такого анализа выяснилось, что в них зафиксированы различные виды информации («Теоретические темы», «Материал для повторения», «Упражнения» и др.).

Так как предполагается, что в процессе работы будет создаваться тегированный текст, то необходимо было выбрать определенную систему тегирования создаваемых текстов, т. е. разработать систему букв и цифр, которые бы указывали на тот класс слов (существительное, глагол, прилагательное и т. п.), к которому относится конкретное словоупотребление текста, и на морфологические и структурные признаки такого словоупотребления. В качестве такой системы был использован набор тегов CES (Corpus Encoding Standart), который уже применялся ранее в Европе для создания различных корпусов текстов. К тому же, этот код был использован нами при разработке параллельных англо-белорусского и немецко-белорусского корпусов текстов. Дополнительно в эту систему тегов введены структурные признаки словоупотреблений (число слогов в словоупотреблении и место ударного слога в нем). Они могут понадобиться при изучении стихотворных текстов.

В дополнение к вышесказанным возможностям параллельных корпусов текстов, учебные параллельные корпуса текстов позволяют:

1. Создавать белорусско-русские и русско-белорусские словари по различным видам информации учебников и по всем текстам учебников.

2. Отбирать текстовые примеры из художественной литературы на белорусском и русском языках как непосредственно для обучения в классе, так и для создания учебников и учебных пособий.



3. Создавать двуязычные терминологические словари по различным школьным и вузовским дисциплинам.
4. Находить текстовые примеры, верифицирующие значения многозначных лексических единиц.
5. Изучать стилистические особенности различных авторов.
6. Отбирать предложения и словосочетания определенных синтаксических структур.
7. Выявлять наиболее употребительные типы контрольных вопросов на русском и белорусском языках.
8. Выявлять формулировки наиболее употребительных заданий в упражнениях на русском и белорусском языках.
9. Находить и классифицировать ошибки обучаемых.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Зубов, А. В. О принципах тегирования слов корпуса текстов белорусского языка / А. В. Зубов, В. А. Кощенко, Н. В. Зыгмантович // Материалы ежегодной научной конференции преподавателей и аспирантов университета. 24–25 апреля 2007 г.: в 2 ч. – Минск: МГЛУ, 2007. – Ч. 2. – С. 116–119.
- 2 Беляева, Л. Н. Лексикографический потенциал параллельного корпуса текстов / Л. Н. Беляева // Труды Международной конференции «Корпусная лингвистика – 2004». 11–14 октября 2004 г. – Санкт–Петербург: СПбГУ, 2004. – С. 55–64.
- 3 Сичинава, Д. В. Комплексное исследование одноязычных параллельных корпусов в грамматических исследованиях / Д. В. Сичинава // Труды Международной конференции «Корпусная лингвистика – 2011». 27–29 июня 2011 г. – Санкт–Петербург, 2011. – С. 316–321.
- 4 Медведева, И. Л. «Лингвистика большого корпуса» в теоретическом и прикладном аспекте / И. Л. Медведева // Вопросы лингвистической семантики. – Тула, 1976. – С. 58–66.
- 5 Соснина, Е. П. О разработке и использовании российского учебного корпуса переводов / Е. П. Соснина [Электронный ресурс]. – 2010. – Режим доступа: <http://ling.ulstu.ru/linguistics/chair/ltcturers/sosnina/development/>. – Дата доступа: 26.09.2011.

УДК 81‘373:398.9:004.77

В. И. Коваль

### **РОЛЬ ИНТЕРНЕТ-РЕСУРСОВ В ИССЛЕДОВАНИИ ФРАЗЕОЛОГИИ (НА МАТЕРИАЛЕ УСТОЙЧИВОГО СЛОВСОЧЕТАНИЯ *ПЕРЕМЫВАТЬ КОСТОЧКИ*)**

*В статье рассматриваются особенности анализа немотивированных фразеологизмов, соотносящихся с мотивированными словосочетаниями. Отмечаются основные закономерности функционирования фразеологической единицы перемывать косточки в интернет-текстах, включая ресурсы НКРЯ. Указывается возможность синхронного и диахронного анализа фразеологизмов в сетевом пространстве.*

Сегодня вряд ли можно встретить лингвиста, кто бы мог «вчистую» отрицать важность обращения к интернет-источникам в процессе исследования единиц языка, в том числе – фразеологизмов. Естественно, что при этом необходимо учитывать научный уровень и «статус» таких источников, к которым не могут быть отнесены низкопробные и непрофессиональные интернет-публикации, авторы которых претендуют на собственное «слово в науке».

Впервые к Интернету как источнику фактического материала мы обратились в 2008 году в процессе описания идеологизированного концепта «славянское единство». Различные (нередко взаимоисключающие) подходы и точки зрения по данному вопросу (Славянское единство – миф или реальность?; Каковы основы единства славян в настоящее время? Имеются ли исторические перспективы единства славян? и др.) дали возможность выявить несколько подходов в понимании этого концепта, проиллюстрировав каждый из них соответствующими материалами [1]. Впоследствии оказалось, что рассмотрение «в зеркале Интернета» некоторых устойчивых словосочетаний (*настоящий мужчина, настоящая женщина* [2]; *женская логика* [3]), а также собственно фразеологизмов (*публичная женщина, публичный мужчина* [4]; *съесть собаку* [5]; *бесплодная смоковница* [6]) позволяет не только представить разные точки зрения относительно их внутренней формы, но (что особенно важно) выявить и исследовать реальную динамику фразеологического состава, проявляющуюся в трансформациях их компонентного состава и в изменении узуальной семантики. Интернет (в первую очередь популярный ресурс «Национальный корпус русского языка» – [www.ruscorpora.ru](http://www.ruscorpora.ru)) оказывается весьма надежным (хотя и существующим виртуально) источником, дающим исчерпывающее представление о фразеологических единицах русского языка в их статике и динамике.

Особенно интересные результаты могут быть получены через «пропускание» через Интернет немотивированных фразеологизмов, потенциально (а иногда и реально) соотносящихся с нефразеологическими, переменными словосочетаниями (*вознать в краску, съесть собаку, наставить рога, съездить в Могилев, выносить сор из избы, вбить осиноый кол* и др.). К числу таких оборотов относится и фразеологизм *перемывать косточки* злословить, сплетничать, судачить о ком-либо<sup>4</sup>. Каковы могут быть результаты его «интернетного измерения»?

Во-первых, помещение этого устойчивого словосочетания в поисковую систему (например, в Google.com) позволяет быстро получить исчерпывающую информацию о его происхождении, которая благодаря авторитету академика В. В. Виноградова «перекочевала» в школьные и вузовские учебники, многочисленные научные работы. Еще в 1954 году в статье «Из истории русской лексики и фразеологии» известный ученый отнес фразеологизм *перемывать косточки* к языковым фактам, реконструкции которых способствуют «данные материальной и духовной культуры в связи с теми семантическими намеками, которые заключены в смысловой структуре

выражения». Развивая данную идею, он отмечает необходимость связи данного выражения «с тем значением слова *кости* (*косточки*), которое относилось к останкам умерших». Это, в свою очередь, дало основание для соотнесения происхождения фразеологизма с «обычаем перемывания костей умершего». Далее В. В. Виноградов указывает не только на содержащееся в работе академика Е. Е. Голубинского «История русской церкви» описание обряда «вторичного захоронения» останков покойника у греков, при котором извлеченные из земли кости обмывались водой и вином (свидетельство XI века), но и на известное южным славянам – сербам и хорватам – «двоструко сахрањивање», заключавшееся в откапывании (спустя три года или семь лет после похорон), перемывании сохранившихся костей водой и вином и повторном погребении [7].

Во-вторых, в Интернете можно обнаружить примеры языковой игры, в которых отражена «двойная жизнь» словосочетания *перемывать косточки*, воспринимаемого и как свободное, и как устойчивое. Сравн.: *Больше всех любят перемывать косточки археологи; Археологи перемывают косточки Александру Македонскому.*

Кроме того, представляют определенный интерес найденные в Интернете картинки, наглядно воплощающие эту «двойную жизнь» рассматриваемого оборота: раскопки захоронений и перемывание костей археологами (1), перемывание человеческих останков в лаборатории при одновременном заинтересованном общении женщин (2), обсуждение подругами сплетен, слухов (3, 4). Подобный «видеоряд» с успехом может быть использован в процессе проведения занятий по развитию речи с иностранными студентами:



(1)



(2)



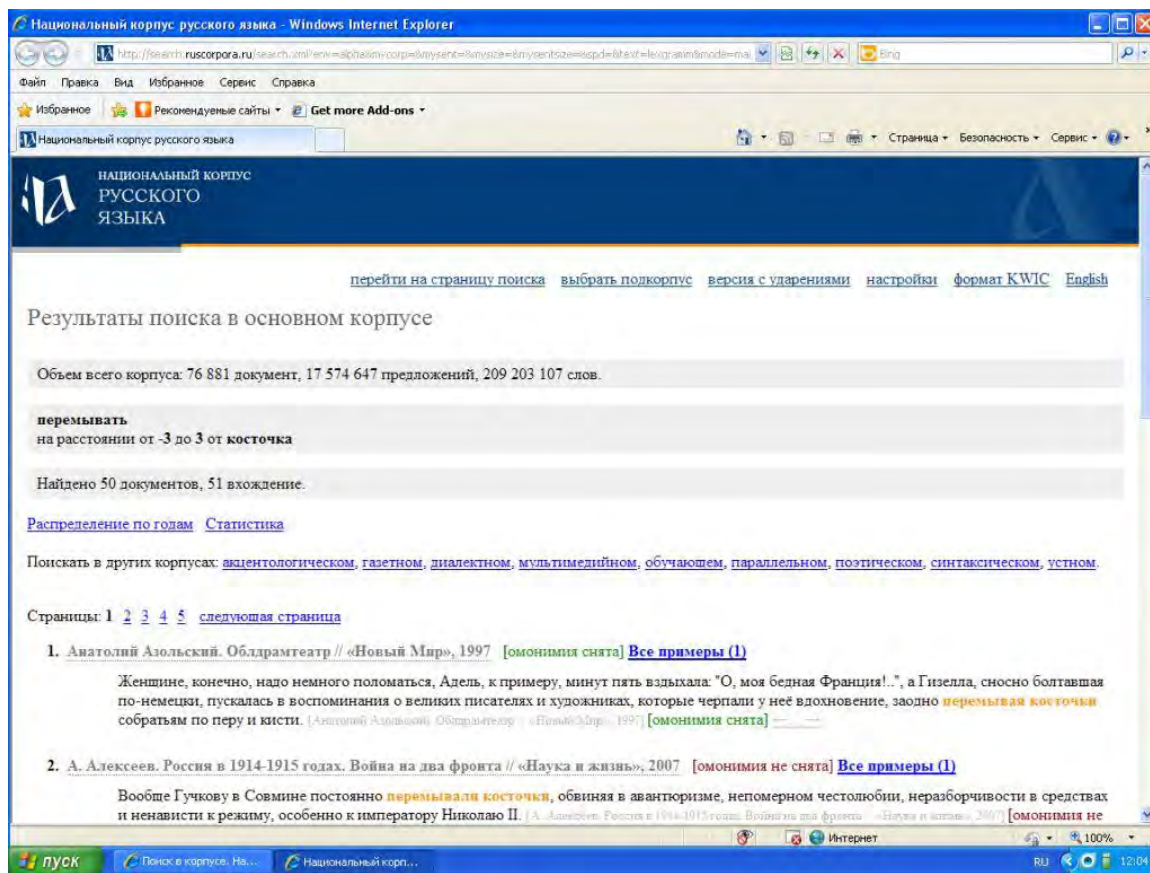
(3)



(4)

В-третьих, богатейший материал для изучения особенностей употребления фразеологизма в художественных и публицистических текстах дают ресурсы Национального корпуса русского языка (НКРЯ). Так, только на

запрос *перемывать* и *косточка*, помещенный в раздел «Лексико-грамматический поиск» в Основном корпусе, мы получили 50 документов:



Всего из Основного и Публицистического корпусов НКРЯ нами было получено 132 результата (контексты, иллюстрирующие употребление фразеологизма *перемывать косточки* в разных вариантах), что дало возможность выявить ряд интересных закономерностей.

1. Прежде всего, обнаруживается выразительная «дефразеологизация» этого фразеологизма. Так, в одном из материалов приводится описание ритуала, с которым связано происхождение данного оборота: *Западные славяне в прямом смысле слова перемывали кости – тело расчленили, выварили в святой воде, затем в новой порции святой воды перемывали кости, припоминая заодно злые и добрые поступки покойного* (Комсомольская правда, 2003.08.18). В другом контексте отражено использование этого словосочетания применительно не к ритуальной, а к бытовой ситуации: *Если бы не эта запруда, кто знает, может быть, мальчика давно уже не было бы в живых. И, как говорила бабка, река давно бы уже перемыла его кости и вынесла бы их прямо в Иссык-Куль, и разглядывали бы их там рыбы и всякая водяная тварь* (Ч. Айтматов. Белый пароход). В следующем случае в результате «столкновения» в пределах одного контекста рассматриваемого фразеологизма и свободного словосочетания (*перемывать*) посуду возникает юмористический эффект: *После ухода гостей хозяйка перемывала им косточки быстрее, чем посуду.* (Н. Богословский. Заметки на полях шляпы).

2. В ряде случаев в НКРЯ выявляются контексты, иллюстрирующие определенное «обнажение» внутренней формы фразеологизма в результате конкретизации значения его именного компонента, развивающего контекстуальную семантику деталь, частность, отдельная черта характера: *Переполоскали, перемыли Марусины косточки – и не о чем больше* (Е. И. Замятин. На куличках); *После того, как пришелец был рассмотрен, изучен и перемыт по косточкам, разговоры за столами вернулись в обычное русло* (М. Дяченко, С. Дяченко. Привратник); *Что должны ощущать почтеннейшие ее редакторы, читая, как «Помои» перемывают ее косточки и в каждой косточке прозревают измену* (М. Е. Салтыков-Щедрин. Письма к тетеньке); *Коллеги на «Ленфильме» недоумевали, перемывая мои военные косточки* (М. Кураев. Записки беглого кинематографиста).

3. Поскольку свойственное рассматриваемому фразеологизму значение – злословить, сплетничать, судачить о ком-либо – содержит в себе эксплицитную эмоциональную сему интенсивности, этот оборот употребляется в сочетании с соответствующими наречиями: *Пресняковым радовостно перемывают косточки глянцевого журналы и театральные критики* (Комсомольская правда, 2004.07.14); *Тут насытившаяся Рита начала энергично перемывать кости Соне, а Лиза поддакивала ей и про себя радовалась избитой теме* (Д. Симонова. Шанкр); *Пожилые дамы расположились на диванах в гостиных и неутомимо перемывали косточки своим ближним* (Н. А. Островский. Рожденные бурей); *Тетя Глаша все еще продолжала уборку и заодно тщательно перемывала косточки школьным шалунам* (Е. Чеповецкий. Непоседа, Мякиш и Нетак); *С Малаховым общаться тоже сложно: это в «Большой стирке» он резво перемывал кости окружающим* (Комсомольская правда, 2004.09.15).

4. Представленные в НКРЯ материалы дают возможность выявить и другие синтагматические характеристики фразеологизма *перемывать косточки*, в частности – особенности его сочетаемости с близкими по значению фразеологизмами, свободными словосочетаниями и отдельными словами: *Перемыли косточки всем знакомым, обсудили все их слабости и глупости, воздали всем сестрам по серьгам* («Семейный доктор», 2002.09.15); *Никто бы не лез ни к кому в душу, никто бы не перемывал друг другу косточки... »* (Т. Шмыга. Счастье мне улыбалось...); *Можно копаться в грязном белье и перемывать косточки, а можно и «ток», и «шоу» о добром и вечном, о смешном и грустном* (Труд-7, 2002.04.30); *Они вели язвительные разговоры, перемывали косточки всем знакомым. Госпожа Дулебова рассказывала о перчатках Потерина* (Ф. К. Сологуб. Капли крови); *Весь театр рассказывал мне, какая она нехорошая, все перемывали ей кости и ругали на чем свет стоит* (Труд-7, 2006.05.05).

5. Приведенные в НКРЯ контексты, включающие оборот *перемывать косточки*, могут быть рассмотрены не только в лингвистическом аспекте, но и с точки зрения отражения в них некоторых психологических характеристик сферы «Злословье, сплетни»:

5.1. Объектом обсуждения, злословья могут быть не только обычные люди (знакомые, сослуживцы, родственники, «публичные» персоны), но и спортсмены, начальники, руководители (в том числе – и руководители государства). В таком случае в данном фразеологизме в большей мере актуализируется значение критиковать, осуждать<sup>1</sup>: *Британские телеэксперты перемыли кости всем игрокам «Арсенала», в том числе и Андрею Аршавину* (Советский спорт, 2010.02.08); *Видя такое отношение к себе, начинают перемывать кости боссам* (Труд-7, 2010.05.23); *Обсудили новости, перемыли кости начальству* (М. Ходаренок. Зенитные ракетные страсти); *Честно говоря, скучно было обсуждать футбольные матчи и перемывать кости политикам из Кремля* (А. Житков. Кафедра).

5.2. Из контекстов следует, что обычно злословят и сплетничают при обсуждении характера, внешнего вида или особенностей поведения: *После длинного разговора о любви одна из девиц встает и уходит. Оставшиеся начинают перемывать косточки ушедшей. Все находят, что она глупа, несносна, безобразна, что у нее лопатка не на месте* (А. П. Чехов. Из записок вспыльчивого человека); *Зато уж если наши провинциальные премьерши примутся кому-нибудь перемывать косточки, то делают это с неподражаемым совершенством, тем более что тем для такого занятия бывает много, ибо город изобилует легкими и приятными нравами* (А. И. Куприн. Впотьмах); *Стали перемывать друг другу косточки. Оказалось, у многих крали заведены были* (В. Я. Шишков. Емельян Пугачев).

5.3. В ряде контекстов имеются сведения не только о субъектах, но и о месте и времени «перемывания костей»: *Такие люди преспокойно забывают козла, распивают «на троих» или целыми вечерами где-нибудь на лавочке перемывают кости соседям* («Комсомольская правда», 1971.06.10); *А потом с горя начала собирать по вечерам на посиделки кумушек и с ними перемывать кости мачехиным дочерям* (Е. Датнова. Возвращение на кухню); *Сейчас две старухи закроются в баньке, поднимут по стопке за Преполовенье и тебе же перемоют кости* [В. Личутин. Любостай).

Таким образом, имеющиеся в свободном доступе богатые информационные интернет-ресурсы (включая возможности НКРЯ) в значительной степени способствуют всестороннему исследованию фразеологического состава русского языка: устойчивые словосочетания «в зеркале Интернета» могут быть исследованы как в диахронном, так и в синхронном аспектах, а также с точки зрения их парадигматических и синтагматических характеристик.

#### СПИСОК ЛИТЕРАТУРЫ

1 Коваль, В. И. Концепт «славянское единство» по данным Интернет-источников / В. И. Коваль // «Восточные славяне: историческая и духовная общность», международная научно-практическая конференция (2008; Гомель). «Восточные славяне: историческая и духовная общность» 15–16 апреля 2008 г. [Текст] : [посвящ. 15-летию Общества Кирилла Туровского : материалы] / редкол. : В. И. Коваль (отв. ред.) [и др.]. – Гомель : ГГУ им. Ф. Скорины, 2008. – С. 19–24.

2 Коваль, В. И. *Настоящий мужчина и настоящая женщина* в зеркале Интернета / В. И. Коваль // Славянская фразеология и паремиология в XXI веке: сборник научных статей / под ред. В. М. Мокиенко, Е. Е. Иванова. – Могилёв: МГУ им. А. А. Кулешова, 2009. – С. 148–155.

3 Коваль, В. И. Фразеологизм *женская логика*: семантика, концептуализация, употребление / В. И. Коваль // Фразеология и языковая динамика: сборник научных трудов, посвященных 55-летию профессора Харри Вальтера, почетного доктора Санкт-Петербургского университета. Материалы XXXIX Международной филологической конференции (15–19 марта 2010 года). Вып. 22 / под ред. проф. Мокиенко В. М. и доц. Савченко А. В. – СПб.: СПбГУ, 2011. – С. 259–263.

4 Коваль, В. И. Фразеологизмы *публичная женщина* и *публичный мужчина*: история и современность / В. И. Коваль // Фразеология, познание и культура: сб. докл. 2-1 Междунар. науч. конф. (Белгород, 7–9 сентября 2010 года): в 2 т. / отв. ред. проф. Н. Ф. Алефиренко. – Белгород: Издво БелГУ, 2010. – Т. 1. Фразеология и познание. – С. 181–185.

5 Коваль, В. И. Фразеологизм *съесть собаку* в пространстве Интернета / В. И. Коваль // Фразеология во времени и пространстве: сб. научных трудов / Науч. редакторы: Х. Вальтер, В. М. Мокиенко, А. В. Савченко. – Greifswald – Sankt-Peterburg, 2012. – С. 167–171.

6 Коваль, В. И. К происхождению фразеологизма *бесплодная смоковница* / В. И. Коваль // Фразеология и Библия: Материалы XLI Международной филологической конференции (г. Санкт-Петербург, 26–31 марта 2012 года) (В печати).

7 Виноградов, В. В. История слов. Перебивать косточки / В. В. [Электронный ресурс]. – Режим доступа: Виноградов <http://www.wordhist.ru/peremivat.html>.

УДК 81'276:004.738.5

М. А. Корсак (Гомель)

### **СТРУКТУРНЫЙ АНАЛИЗ ЗООДЕМОТИВАТОРОВ (НА ПРИМЕРЕ ИНТЕРНЕТ-МЕМА *COVA ART-СТУДЕНТ*, ИЛИ *\*OVO\* ART STUDENT OWL*)**

*В статье предпринимается попытка охарактеризовать зоодемотиваторы (рабочее название), которые представлены в социальной сети <http://vk.com>. Устанавливаются особенности функционирования мема как креолизованного текста, который может быть исследован с филологической точки зрения.*

*Зоодемотиватор* – это разновидность демотиватора, в структуре которого в качестве графического изображения (картинки, фото, комикса и т. д.) выступает зооним. Зооним в таком демотиваторе персонифицирован.

На сайте Вконтакте выложены все группы зоодемотиваторов, функционирующие на данный момент в русскоязычном интернет-пространстве и постоянно пополняющиеся анонимными пользователями: *Агрессивно-социальный пингвин, Жаба холостяк, Классический кот, Пингвин-социофоб,*

*Смелый волк, Собака советчица, Сова Арт-студент \*O v O\* Art Student Owl, Типичная собака, Философрантор.* В свою очередь, в каждой из выше названных групп зоодемотиваторы объединены в альбомы. Анализ групп зоодемотиваторов позволил выделить наиболее представленную в количественном отношении группу – *Сова Арт-студент \*O v O\* Art Student Owl* [1]. На сайте Вконтакте находится шесть альбомов, которые включают данный зоодемотиватор. Все альбомы (в альбом может входить 501 дем) различны в количественном отношении (#1 – 260 демов; #2 – 488; #3 – 341; #4 – 419; #5 – 488; #6 – 341), иногда включают одинаковые демы.

Зоодемотиватор *Сова Арт-студент \*O v O\* Art Student Owl* по способу образования является собственно интернет-мемом, заимствованным в русскоязычную среду Интернета из англоязычной. В оригинале название данного дема звучит как *Art Student Owl*.

Данный дем посвящён всем, кого постигла участь обучаться в художественной школе/художественном училище/вузе. Дем интерпретируется следующим образом: студент-художник настолько связан с искусством, что все остальное становится неважным, ему едва хватает денег, чтобы поесть, он смотрит свысока на тех, кто живет в «реальном мире», а не в мире искусства. Весь мир для студентов как искусство. Дем включает в свою композицию фотографию довольно скучно смотрящей на всё совы (кто знал, что сова может выглядеть скучно?) с сигаретой, свисающей из клюва. Называется дем так не потому, что студент полностью поглощён и увлечён искусством, а потому, что студенты-художники, сосредоточенные на своем ремесле, полностью рассеяны. Текст вокруг дема, как правило, имеет какое-то отношение к искусству, художественной школе, художникам, или тому, что связано каким-то образом с искусством. Употребление дема уместно в тех случаях, когда затрагиваются темы, связанные с искусством.

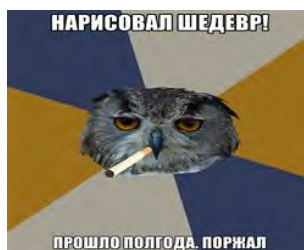
В «русском» Интернете данный демотиватор первоначально распространялся с текстом, который являлся переводом с английского языка.

Зоодемотиватор как разновидность демотиватора, который представляет поликодовое образование, имеет четкую композицию и содержит три основных элемента: изображение, расположенное на определённом фоне; слоган, набранный крупным шрифтом; таглайн – пояснительную надпись к слогану, комментирующую или в редких случаях дублирующую слоган. Однако в отличие от классического демотиватора, в структуре которого изображение заключено в чёрную рамку, фоном для фото совы в зоодемотиваторе *Сова Арт-студент \*O v O\* Art Student Owl* служит поле, разбитое на шесть трёхцветных сектора (оливковый, серый, синий). Слоган, всегда набранный крупным шрифтом, в анализируемом зоодемотиваторе сопровождается таглайном, который отличается от слогана более мелким шрифтом (демотиватор 2 и демотиватор 3) или имеет такой же размер шрифта, как и слоган (демотиватор 1). На наш взгляд, одинаковый размер шрифта, который выбирается для пояснительной и основной надписи, затрудняет различение слогана и таглайна (демотиватор 1).

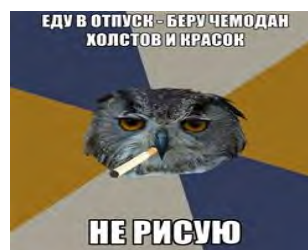




Демотиватор 1



Демотиватор 2

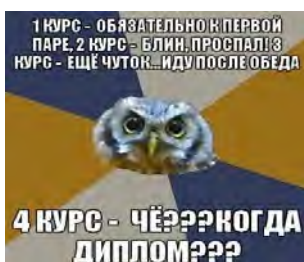


Демотиватор 3

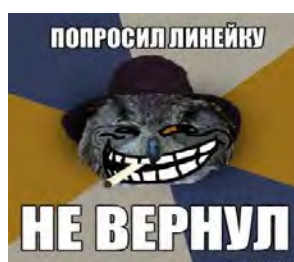
Подписи, окружающие демотиватор, не имеют постоянного расположения: слоган может находиться вверху, а таглайн – внизу (демотиватор 1 и демотиватор 2), слоган пишется внизу, таглайн – вверху (демотиватор 3).

После того как зоодемотиватор *Сова Арт-студент \*O v O\* Art Student Owl* получил широкую известность в русскоязычной среде Интернета и адаптировался к ней, он претерпел определённые качественные и количественные изменения. Надо отметить, что зоодемотиватор настолько «обрусел», что послужил базой для возникновения новых демов с изображением Сова, которые составляют отдельные серии и на сайте Вконтакте представлены следующими альбомами: *Сова-профессор* (390 единиц), *Другие Арт-Совы* (219), *Совёнок-новичок* (77).

В композиции демов, производных от *Сова Арт-студент \*O v O\* Art Student Owl*, варьируются практически все формообразующие элементы. Однако наиболее часто в новых сериях модифицируется структура изображительной части. Во-первых, происходит трансформация самого изображения совы. Это можно наблюдать в серии *Совёнок-новичок* (демотиватор 4), а также в демах, где сова представлена как *trollface* (демотиватор 5). В альбоме *Совёнок-новичок* собраны демы, в центре которых располагается фото совёнка. *Сова-trollface* может заменять, например, изображение *Сова Арт-студент*.



Демотиватор 4



Демотиватор 5

Во-вторых, изображение совы активно персонифицируется. Персонификация достигается за счёт дополнительных предметов (дем 9), помещения совы в пространство, в котором живёт человек (дем 8), стилизации под лицо человека, полностью или частично (демотиваторы 6 и 7) и т. д. Однако чаще всего персонификация происходит на основе нескольких признаков, например, стилизация самого изображения и дополнительные

предметы (демотиватор 6). Особенно наглядно это демонстрируют демы, собранные в альбоме *Другие Арт-Совы*.



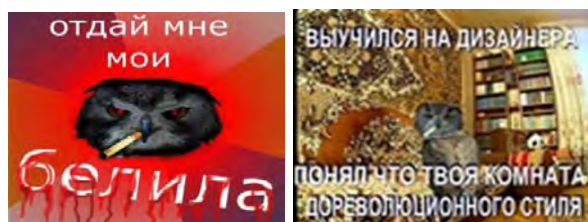
Демотиватор 6    Демотиватор 7    Демотиватор 8    Демотиватор 9

На Демотиваторе 8 видно редкое изображение совы в «полный рост».

Стилизация изображения совы может ориентироваться на различные культуры. В альбоме *Сова-профессор* встречается изображение Совы-профессора с атрибутами самурая.



В-третьих, иногда изменяется сам фон и цвет зоодемотиваторов (демотиватор 8 и 10). Это также можно наблюдать на примере демов, в альбоме *Другие Арт-Совы*.



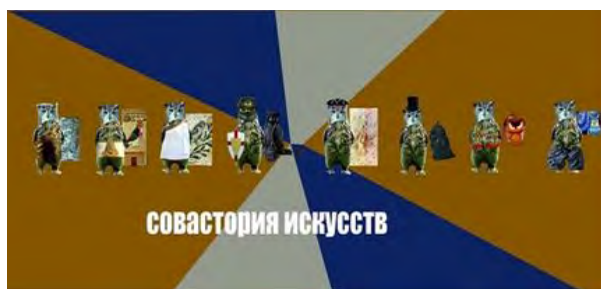
Демотиватор 10    Демотиватор 8

В-четвёртых, сам зоодемотиватор как креолизованный текст дальше усложняется другими знаковыми системами, которые располагаются в различных местах дема (демотиватор 11 и 12).



Демотиватор 11    Демотиватор 12

В-пятых, зоодемотиваторы становятся многокомпонентными: компоненты могут располагаться линейно (демотиватор 13) и вертикально (демотиватор 14).



Демотиватор 13



Демотиватор 14

В-шестых, демотиваторы различных типов могут контаминироваться. Это положение можно продемонстрировать на примере зоодемотиватора 14, который представляет собой особую единицу: демотиватор в демотиваторе. Это многокомпонентный классический дем, в котором изображение располагается на чёрном фоне. Однако фон всех компонентов зоодемотиватора соответствует фону дема *Сова Арт-студент*. Изображение совы в каждом компоненте (кроме первого) данного зоодемотиватора динамично, так как стилизовано под студента и представлено в «полный рост».

В-седьмых, изображение совы практически на всех мемах статично, так как представляет собой голову совы. Оно становится динамичным в нескольких случаях: 1) происходит стилизация головы совы под человеческое лицо, тогда можно наблюдать изменение «головных» соматизмов (демотиватор 6; 7; 12); 2) сова стилизуется под человека и изображается в «полный рост», в таком случае активными становятся «конечности» (демотиватор 13 и 14).

Таким образом, зоодемотиваторы могут служить предметом филологического исследования, так как представляют собой креолизованный текст, построенный по особым правилам. Возникнув в иностранной среде, зоодемотиваторы быстро осваиваются в русскоязычном пространстве Интернета, что приводит к значительным структурным изменениям данного типа текста.

#### СПИСОК ЛИТЕРАТУРЫ

1 Вконтакте [Электронный ресурс] / Сова Арт-студент \*O v O\* Art Student Owl. - URL: [http://vk.com/art\\_student\\_owl](http://vk.com/art_student_owl) (дата обращения 17. 10.2012)

А. Ю. Крохмальник (Гомель)

## **СВЯЗЬ КЛАССИЧЕСКОЙ И АКАДЕМИЧЕСКОЙ «ТОШНОТЫ» И «ВОДЯНИСТОСТИ» ТЕКСТА С ЕГО ФУНКЦИОНАЛЬНЫМ РЕЧЕВЫМ СТИЛЕМ**

*Объектом исследования данной статьи являются характеристики «водянистости», классической и академической «тошноты» текстов, относящихся к различным функциональным стилям. Определены показатели академической «тошноты» текстов, т. е. их перенасыщения повторяющимися словами. Отмечена вариативность показателей «водянистости» и «тошнотности» текстов, т.е. характеристик несущественной для понимания текста информации и повторяющихся лексических единиц.*

«Тошнота текста – это насыщенность текста ключевыми словами» [1]. Показатель «тошноты» определяет «натуральность» текста либо его прямую ориентацию под поисковые запросы. Высокие показатели «тошноты» текста негативно воспринимаются поисковыми роботами. Это является знаком того, что текст написан для поисковиков, а не для людей. Определить процентное содержание элементов тошноты можно при помощи инструмента «Семантический анализ текста онлайн» биржи текстов *Адвего*.

Различают классическую и академическую тошноту документа (текста). «Классическая тошнота – это коэффициент, показывающий зашамленность текста самым повторяющимся словом. Определяется как квадратный корень из количества повторений самого тошнотного слова. Самое тошнотное слово может снижать релевантность по остальным словам. Оптимальное значение классической тошноты равно 7. Большая тошнота может помешать продвижению сайта» [1]. Правда, на практике рекомендуется коэффициентом классической тошноты считать 4. «Академическая тошнота характеризует натуральность документа. Если в документе много различных повторяющихся слов, то показатель академической тошноты будет большим. Академическая тошнота измеряется в процентах и зависит от частоты повторения слов и объема текста. Вычисляется как отношение числа повторов самых частых слов к общему числу слов в тексте» [1]. На практике рекомендуется не 7-процентный показатель академической тошноты текста.

Нами было проведено исследование, в ходе которого определялись показатели классической и академической тошноты текстов, относящихся к различным функциональным речевым стилям. Как известно, «речевые нормы устанавливают закономерности употребления языковых средств в том или ином функциональном стиле и его разновидностях» [2, с. 27]. Вместе с тем данные закономерности тесно связаны с повтором определенных

лексических средств в текстах конкретного функционального стиля и даже относительным показателем водянистости текста, который характеризует степень употребления в документе тех или иных слов и выражений, не участвующих в раскрытии основной мысли текста.

Исследование было проведено на практическом занятии по одноименной теме при участии академической группы РФК-43. Для анализа были отобраны приблизительно одинаковые по объему отрывки текстов, принадлежащих различным функциональным речевым стилям.

Результаты исследования были оформлены в виде компьютерных презентаций и прокомментированы. Так, например, студенткой Очеретяной Ольгой в качестве текста научного стиля была взята статья из энциклопедии, посвященная описанию божьей коровки. Показатель воды составил 63 % (при норме 65 процентов). Классическая тошнота документа оказалась равной 2.00. Само словосочетание «божья коровка» встречается в тексте 2 раза. Академическая тошнота текста оказалась равной 9,2 %. Это означает, что в тексте встречаются различные повторяющиеся слова, которые могут быть отнесены к терминам, отсюда и название – «академическая» тошнота документа.

В качестве текста, относящегося к художественному стилю, был взят отрывок с описанием дуба из романа Л. Н. Толстого «Война и мир». Проанализировав текст, программа «Семантический анализ текста» определила, что показатель воды данного текста составляет 75 %. Такая величина обусловлена тем, что в тексте присутствуют описания природы и рассуждения главного героя. Стоп-слова в количестве 63 единиц. Как правило, это частицы, предлоги. Классическая тошнота документа не превышает норму – 2.00. Академическая тошнота текста составляет 10 %.

В качестве текста разговорного стиля был взят отрывок из произведения В. Ларина «Нейронный шок». Показатель воды в данном случае составляет 70,8 %. Устная речь часто «засоряется» ненужными оборотами при описании того или иного явления. Классическая тошнота документа равна 1,41 %, а академическая тошнота – 6,1 %. Это значит, что в данном случае качество разговорной речи достаточно приемлемое, она не переполнена повторами, как это часто происходит.

Текстом публицистического стиля явился текст-воспоминание советской актрисы и режиссера Софьи Гиацинтовой о Василии Качалове, в котором героиня рассуждает о воспитанности. Классическая тошнота документа составляет 1,73 %. Академическая тошнота документа в 8,0 % считается несколько выше нормы, поскольку разного вида повторы могут выполнять воздействующую (суггестивную) функцию. Показатель воды в тексте высокий – 77,4 %.

Текстом официально-делового стиля стало одно из постановлений 2011 года, принятое на территории Российской Федерации. Классическая тошнота документа составляет 1,41 %. В тексте 57,1 % воды, академическая тошнота документа не превышает норму – 6,1 %.

Студентка Макарова Людмила провела сравнительный анализ текстов различных функциональных стилей и свои выводы оформила в конце своего исследования. По ее наблюдениям, самым большим количеством «воды» обладает пример, представляющий собой разговорный стиль речи (78,8 %), а также пример, представляющий собой научный стиль речи (71,9 %). Художественный стиль речи имеет 70,7 % «воды», публицистический – 70,0 %. Наименее «водянистым» оказался официально-деловой стиль речи. Значительный процент «водянистости» указывает на то, что в тексте употреблено много слов, которые не раскрывают тему, например цифры, служебные части речи: предлоги, союзы, частицы.

При анализе стилей речи было выявлено, что в текстах встречаются «классическая тошнота документа» и «академическая тошнота документа»: повторы, тавтология, употребление синонимов, словоформ, а также слов, понимание которых затруднено.

Наиболее существенной *классической тошнотой* обладают следующие стили речи: научный (1,73 %), художественный (1,73 %), официально-деловой (1,73 %), разговорный (1,73 %). Публицистический стиль речи имеет наименьший показатель *классической тошнотности* – 1,41 %.

Наибольшая *академическая тошнотность документа* у текстов официально-делового стиля речи (13,7 %) и научного стиля речи (13,6 %). Далее следуют тексты разговорного стиля речи – 10,9 %; художественного стиля – 8,1 % и публицистического стиля – 5,1 %.

Студентка Анастасия Лашкевич составила сводную таблицу показателей водянистости, а также классической и академической тошнотности текстов:

	«Водянистость»	«Классическая» тошнота	«Академическая» тошнота
Разговорный стиль	71,7 %,	2,45	4,1 %,
Публицистический стиль	68,1 %	4,36	6,8 %
Официально-деловой стиль	51,4 %	4,80	10,7 %
Художественный стиль	67,9 %	3,87	7,5 %
Научный стиль	54,8 %	3,32	7,8 %

Студентка Наталья Прокопенко сделала следующие выводы по показателям тошнотности текстов:

*Классическая* тошнота текстов

- Официально-деловой стиль 1.41
- Научный стиль 1.41
- Публицистический стиль 2.00
- Художественный стиль 2.45
- Разговорный стиль 1.00

*Академическая* тошнота текстов:

- Официально-деловой стиль 8.0 %
- Научный стиль 12.9 %

- Публицистический стиль 7.2 %
- Художественный стиль 16.0 %
- Разговорный стиль 0.0 %

Таким образом, получены различные результаты классической и академической тошнотности текстов, относящихся к различным функциональным речевым стилям. Так, например, академическая тошнота текста научного стиля оказывается наиболее высокой либо одной из наиболее высоких, а классическая тошнота может оказываться на низком уровне или не превышает средний показатель. По всем остальным текстам разных стилей наблюдается варьирование соответствующих значений. Данный факт может быть связан как со структурными особенностями самих текстов (тематикой, характером изложения информации), так и с тем фактом, что повтор лексических единиц в принципе характерен для текста любого функционального стиля. Наибольшим процентом *водянистости* отмечены тексты разговорного стиля, поскольку разговорная речь обычно отличается либо политематичностью, либо большим количеством второстепенных лексических единиц. Вместе с тем делать выводы о *водянистости* либо *тошнотности* текстов какого-либо функционального стиля видится нам преждевременным ввиду недостаточного количества исследованного материала.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Тошнота текста / Адвего. Документация [Электронный ресурс]. – Режим доступа: [http://wiki.advego.ru/index.php/%D0%A2%D0%BE%D1%88%D0%BD%D0%BE%D1%82%D0%B0\\_%D1%82%D0%B5%D0%BA%D1%81%D1%82%D0%B0](http://wiki.advego.ru/index.php/%D0%A2%D0%BE%D1%88%D0%BD%D0%BE%D1%82%D0%B0_%D1%82%D0%B5%D0%BA%D1%81%D1%82%D0%B0). – Дата доступа: 2012.01.10.
- 2 Плещенко, Т. П. Стилистика и культуры речи: учеб. пособие / Т. П. Плещенко, Н. В. Федотова, Р. Г. Чечет; Под ред. П. П. Шубы. – Мн.: «ТетраСистемс», 2001. – 544 с.

УДК 81-139:81'371:81'373.612.2

С. Б. Кураш (Мозырь)

#### **ФРАГМЕНТ КОРПУСНОГО ИССЛЕДОВАНИЯ КОНЦЕПТА «МЕТАФОРА»: МЕТАФОРА И ЕЁ АНТИПОДЫ**

*В статье представлен фрагмент корпусного исследования концепта «метафора» в аспекте выявления того, каким феноменам и сущностям («антиметафорам») данный концепт противопоставляется в современных дискурсивных практиках.*

**Введение.** В работах [1–5], выполненных с применением методов корпусной лингвистики главным образом на материале различных интернет-источников, нами было показано, что слово «метафора» на данном временном срезе вышло далеко за пределы поэтико-риторической терминологии и активно концептуализируется в различных дискурсивных практиках.

В данной работе будет представлен обзор контекстов, в которых концептуализируются сущности, мыслимые как антиподы метафоры.

**Методика исследования.** Поисковой системе Национального корпуса русского языка (НКРЯ) (<http://www.ruscorpora.ru>) нами был задан запрос на поиск всевозможных контекстов, содержащих смысловую формулу «*X – это не метафора*» в разных подкорпусах. Пример одного из вариантов запроса с наиболее релевантным результатом поиска (более 90 % аутентичных контекстов) приводится в нижеследующем скриншоте.

Всего в НКРЯ нами было выявлено около 100 релевантных контекстов, подавляющее большинство из них (более 90 %) – в основном и газетном подкорпусах.

**Результаты исследования.** Выявлены следующие способы концептуализации «антиметафоры» в современных дискурсивных практиках.

1) **Метафора vs. реальность как таковая, возможное в принципе, факт:** – *Первый шаг – взглянуть правде в глаза и понять, что вымирание русских – вовсе не метафора, это наша реальность, и идет этот процесс намного быстрее, чем кажется. [Невидимая война (2003) // «Советская Россия», 2003.02.15]; А для религиозного человека – это, конечно, не затрепанная метафора, а реальность. [Поэт и толпа (2003) // «Театральная жизнь», 2003.07.28].*

2) **Метафора vs. реальность объекта:** *Эти звезды, Алекс, это ведь не просто метафора, Алекс, не правда ли? [Василий Аксенов. Новый сладостный стиль (2005)]; Это была не метафора, если под копейкой понимать тогдашний круглый бронзовый французский франк. [Н. Н. Берберова. Курсив мой (1960–1966)].*

3) **Метафора vs. реальность конкретного события.** Именно этот тип концептуализации антиметафоры преобладает в газетном подкорпусе



НКРЯ: *Из рваной воронки в асфальте бил гейзер. Настоящий фонтан горячей воды – не метафора. Это происходило на левосторонней набережной имени И. В. Сталина, в двух шагах от Дворца бракосочетаний, которому Салам, будь на то его воля, дал бы имя своего дядюшки Лео, так часто тот вступал в различные браки и выступал из них.* [Теймураз Мамаладзе. «Здравствуй, осел!» (1999) // «Дружба народов», 1999.07.15]; *И это вовсе не метафора: зной пришел в Москву из Средней Азии, раскаленный воздух закачивает в центр России антициклон, который висит над южным Уралом.* [Юлия СМИРНОВА. До конца недели в Москве 33-градусное пекло // Комсомольская правда, 2010.06.22].

4) **Метафора vs. реальность ощущения:** *это не метафора, не образ, – где его «душа», где тело? – но опять физическое, телесное ощущение холода, – ощущение смерти.* [З. Н. Гиппиус. *Задумчивый странник (о Розанове) (1923)*]; *Игроки Анчелотти зубами вцепились в штрафную римлян, а Амброзини наглядно продемонстрировал присутствующим, что умение сверкать глазами – это не литературная метафора, а обычное для разъяренного человека состояние.* [Филипп Бахтин. *Две крепости. Пьер Луиджи Коллина разбудил первородное зло (2003)* // «Известия», 2003.02.17].

5) **Метафора vs. наука, учение, умозаключение:** *Это не метафора, а утверждение более или менее основательное.* [Андрей Балдин. *Московские праздные дни (1997)*]; *Выражение "любовная лихорадка", таким образом, не метафора, а точный медицинский "диагноз"* [Лескова Наталия. *ФОРМУЛА ЛЮБВИ* // Труд-7, 2006.03.09]; *Слово «равновесие» – не метафора. Я вкладываю в него вполне конкретный научный смысл, имея в виду природное явление, известное в науке под названием «изостазия».* [И. Усейнова. *Кто художник географической карты?* // «Знание – сила», 1987].

6) **Метафора vs. реальность смысла (денотата выражения), то, что нужно понимать буквально:** *Расхожее выражение – найти свою половину – это ведь не просто метафора, она имеет вполне реальный смысл.* [Илона Броневицкая: «У меня всегда было обостренное чувство собственного дома» (2004) // «Мир & Дом. Residence», 2004.04.15]; *В данном случае упоминание о небесах – не расхожая метафора, ее надо понимать буквально.* [Павлючик Леонид. *ПОЛЕТЫ ВО СНЕ И НАЯВУ* // Труд-7, 2005.09.10]; *И, к сожалению, выражение «до безумия» – в данном случае не метафора.* [Борис Васильев. *Оглянись на середине* // «Октябрь», 2003]; *И потому евангельское изречение о том, чтобы не заботиться о завтрашнем дне, а жить, как птицы небесные, – не есть метафора, а утверждение существующего закона всякой животной общественной жизни.* [Л. Н. Толстой. *Христианское учение (1896)*].

7) **Метафора vs. иное поэтологическое понятие:** *Не метафора, а метаморфоза, как любил говорить Блок, думая о теургии вместо стихов.* [Борис Парамонов. *Трава родины, или Сталь и шлак* // «Октябрь»,

2003]; *Элегические розы и слезы – не метафора, но предметное значение вытеснено в них стилистической окраской.* [А. П. Чудаков. *Разговариваю с Гинзбург* // «Новое литературное обозрение», 2001]; *Важно понять, что здесь «море» не отвлеченная метафора, а вполне конкретный образ.* [М. М. Морозов. *Язык и стиль Шекспира (1941)*].

Для многих контекстов искомого характера характерно присутствие особых прагматических усилителей отрицания (а по сути – нанесения урона «авторитету» метафоры), ср.: *не расхожая метафора, не броская метафора, не затасканная, это вам не метафора* и т. п.

Характерны и такие контексты, когда метафора ставится в один синонимический ряд с такими понятиями, как *воображение, выдумка* и пр., ср.: *То есть стихи... да, это не метафора и не воображение* [Людмила Улицкая. *Медея и ее дети (1996)*]; *Тогда лишь поймете вы, что выражения – «волосы становятся дыбом», «руки опускаются» – не метафоры, а истинная правда, – то что бывает в жизни, что рассказы о том, как в одну ночь седеют люди – не выдумка.* [Л. И. Шестов. *Шекспир и его критик Брандес (1898)*].

Ряд контекстов демонстрирует допустимость того факта, что высказывание может быть и метафоричным, и неметафоричным (буквальным) одновременно, ср.: *Кстати сказать, железный занавес – не только метафора.* [Владислав Быков, Ольга Деркач. *Книга века (2000)*]; *Но когда хоккеистов сравнивают с артистами или солистами – это не всегда метафора.* [Лысенков П.. *...А лучшего новичка отправили в Японию. И еще 18 наград лауреатам сезона, которые вручила КХЛ* // *Советский спорт, 2010.05.26*]; *Старый стишок о Гленгайлах – Копит, копит смолоду Наш помещик золото – не только метафора, но и правда.* [Гилберт К. Честертон. *Неведение отца Брауна*].

Лишь немногие высказывания, апеллирующие к анализируемому концепту, оценивают метафору не как антипод реальности, а как единственно возможное средство концептуализации того или иного феномена действительной или вымышленной реальности, подчас даже находящееся в выигрышном положении в сравнении с неметафорическим способом мироосмысления, ср.: *А почему это, в конце концов, приятно? Что такое вечность, как не метафора? Ведь о неметафорической вечности мы ничего не знаем.* [Игорь Сухих. *Остается только метафора...* // «Звезда», 2002]; *Все же, видимо, это была не гениальная метафора, а какое-то медицинское пособие.* [Виктор Пелевин. *Чапаев и пустота (1996)*].

Ещё более примечательны в этом плане контексты, в которых говорящий, пытаясь уйти, как бы отречься от метафоры, тем не менее выражает свою мысль не иначе, как метафорически, иносказательно, фигурально, тем самым наталкивая на мысль о том, что «слабость» метафоры есть не что иное, как обратная сторона её «силы» как инструмента мыслеформирования, ср.: *Картофелина крикнула: «АЙ!» И это не метафора. Конструкторы на протяжении десятилетий проектировали картофелеуборочные*

машины вслепую, не имея никакого представления о том, как картофельные клубни относятся к сконструированным ими устройствам [обобщенный. Заметки (Вокруг земного шара) // «Техника – молодежи», 1975]; **Об этой деревне Мануилово вспоминается, как о рае земном. Слово: «рай» – здесь вовсе не метафора.** Благодаря раннему поступлению в училищную жизнь – в 8 лет, благодаря умственному богатству семьи моих родителей, в 11 лет я был развит не менее, чем мои товарищи 14-ти и 15-ти лет, а в этом возрасте уже много в душе зарождается не только впечатлений, но и мыслей серьезных. [В. П. Мецкерский. Мои воспоминания (1897)].

Таким образом, несмотря на то, что в абсолютном большинстве случаев вне сферы специальных лингвометафорологических исследований метафора концептуализируется со знаком «минус», тем не менее сам факт частотности апелляции к данному понятию в метаязыковых целях, а также случаи наподобие мнимого отказа от метафоры в её же пользу, не говоря уже принципиальной метафоричности осмысления многих феноменов реальности, характерной для современной языковой картины мира, заставляют ещё раз убедиться в том, что метафора – не просто термин, не просто языковое средство, – это, безусловно, концепт, характерный для языкового (и метаязыкового) сознания образованного носителя русского языка, в чём убеждает применение методов корпусной лингвистики к анализу контекстно-семантических наполнений данного слова.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Кураш, С. Б. Нетерминологические употребления слова *метафора* в современном русском языке (по данным Интернет-источников) / С. Б. Кураш // Русский язык: система и функционирование: сб. матер. IV междунар. науч. конф. В. 2 ч. Ч. 1. – Минск: РИВШ, 2009. – С. 184–187.
- 2 Кураш, С. Б. От метафоры корпуса – к корпусам метафор / С. Б. Кураш, О. И. Ревуцкий // Компьютерная лингвистика: научное направление и учебная дисциплина: сборник научных статей. Вып. 1. – Гомель: ГГУ им. Ф. Скорины, 2010. – С. 120–126.
- 3 Кураш, С. Б. «Метафора» vs. «не метафора»: метаязыковой аспект метафорогенеза / С. Б. Кураш // Язык и межкультурные коммуникации: матер. III междунар. науч. конф. / БГПУ им. М. Танка; редколл.: В. Д. Старичёнок и др. – Минск: БГПУ, 2011. – С. 132–134.
- 4 Кураш, С. Б. Метафора и смыслопорождение: pro & contra / С. Б. Кураш // Язык – текст – дискурс: проблемы интерпретации высказывания в разных коммуникативных сферах: матер. междунар. науч. конф., Самара, 12–14 мая 2011 г.; науч. ред. Н. А. Илюхина. – Самара: Изд-во «Универс групп», 2011. – С. 144–148.
- 5 Кураш, С. Б. Функционирование концепта *метафора* в различных дискурсивных практиках (по данным Национального корпуса русского языка) / С. Б. Кураш // Современная филология в контексте взаимодействия языков и культур: сб. м-лов междунар. науч.-практ. конф. / Отв. ред. А. Л. Фатыхова. – Стерлитамак: СГПА им. З. Бишевой, 2011. – С. 174–180.

А. С. Лашкевич (Гомель)

## О СУБСТИТУЦИИ КОМПОНЕНТОВ В СТАТУСНЫХ ЕДИНИЦАХ

*В статье анализируются способы трансформации так называемых статусов – языковых единиц, используемых в процессе интернет-коммуникации. Обращается внимание на использование статусов для емкой характеристики различных эмоциональных состояний и доведения до участников коммуникации последней и наиболее важной информации.*

Статусные единицы в последнее время очень активно используются участниками интернет-общения в социальных сетях. В компьютерном толковании под термином *статус* понимается поле и текст в нём, которые видят пользователи при посещении страниц социальных сетей'.

Каждая статусная единица имеет особенности оформления, именно это вызывает интерес со стороны специалистов и побуждает к анализу данного интернет-явления.

В последнее время одним из самых распространённых способов создания статусных единиц стал способ трансформации известных слоганов, в результате чего избитые и известные большинству русскоговорящих людей фразы стали нести совершенно иную информацию. *Слоган* – это рекламный девиз, который в сжатом виде передает рекламное сообщение, часть долговременной коммуникационной платформы бренда' [1].

Существует даже структурная классификация слоганов. Так, выделяют:

- 1) связанные – включают названия продукта («*L'Oreal – ты этого достойна!*»);
- 2) прямые – содержат личное обращение к потенциальному потребителю («*Будущее зависит от тебя!*»);
- 3) привязанные – соотносятся с названием ритмически и фонетически («*Каждый день только твой, если Always с тобой!*»);
- 4) свободные – самодостаточные и независимые слоганы («*Мечты сбываются!*»).

Интерес авторов статусных единиц к форме слоганов неслучаен. Если слоганы, которые создаются в коммерческой сфере, «употребляется во всех видах рекламной коммуникации для привлечения внимания целевой аудитории, повышения её лояльности бренду, стимулирования продаж» [1], то статусные единицы, воссоздающие их форму, так же призваны привлечь внимание аудитории, стимулировать интерес, вызвать на «словесную дуэль», «поединок». Такого рода статусные единицы выполняют, помимо развлекательной функции, функцию привлечения внимания или самопрезентации, именно в этом и заключается основная цель их создания.

Так, знакомство со статусной единицей *«Куплю себе сковородку Tefal. Пусть хоть она обо мне думает!»* способствует припоминанию у пользователей некогда очень навязчивой рекламы, которая предлагала к покупке продукцию известной международной марки, специализирующейся на производстве посуды с антипригарным покрытием. Слоган этой компании звучал следующим образом: *«Тэфаль. Ты всегда думаешь о нас»* (*«Tefal. You always think of us»*). Таким образом, автор обыгрывает в своей статусной единице известный слоган, причём ему удаётся обратить внимание пользователей на собственное одиночество весьма оригинальным способом. Частица *хоть* усиливает впечатление от статусной единицы, в данном случае выступает в качестве эквивалента частицы *хотя бы* в значении *по крайней мере*. Автор акцентирует внимание на том, что среди его окружения нет людей, которые заботились бы о нём и выражали заинтересованность по отношению к его жизни, поэтому он находит крайне специфический выход из данного положения – купить сковородку Tefal, которая по заверениям производителей, думает о каждом клиенте. Несмотря на то, что статусная единица создавалась с целью заявить о собственном одиночестве, её можно назвать юмористической.

Следующая статусная единица содержит в себе целый ряд реминисценций на слоганы известных фирм: *«МегаФон сказал, что будущее зависит от меня. Кока-кола обещала, что всё будет хорошо. Газпром утверждает, что мечты сбываются! Tefal думает обо мне, Maybelline говорит, что все в восторге от меня, с Always этот день мой и Loreal – я этого достойна. Так что всё зашибись!»*. Автор статусной единицы практически во всех случаях дословно воссоздаёт девизы компаний, небольшие изменения коснулись лишь трёх слоганов. Так, трансформации подвергся слоган *«Всё будет Coca-Cola»*, однако необходимо заметить, что в данном случае сочетание *coca-cola* – это своеобразный синоним словам *замечательно, хорошо, отлично, превосходно*, или к английской аббревиатуре *OK*. В контексте слогана это сочетание воспринимается носителями русского языка именно так. Можно сравнить со слоганом: *«Всё будет охрустенно!»* или *«Живи охрустенно!»*, где неологизм *охрустенно* тоже выступает в качестве синонима к слову *замечательно*. Слоган фирмы *«Always»* был усечён, в оригинале он звучит следующим образом: *«Каждый день только твой, если Always с тобой»*. То есть автором статусной единицы была отброшена та часть слогана, которая, по его мнению, не несёт в себе нужной для данной статусной единицы информации.

Итак, рассматриваемая статусная единица строится по определённой модели: первая её часть, наиболее объёмная, содержит в себе элементы популярных слоганов, которые рисуют очень приятные и оптимистичные для пользователей картины, вторая часть – это вывод, который можно сделать, исходя из содержания использованных слоганов.

Статусная единица *«Долги бывшего СССР? Сначала – йогурт «Даниссимо»*. *И пусть мир подождёт!»* содержит в себе слоган компании,

специализирующей на производстве йогуртов и молочных продуктов: «*Danissimo. И пусть весь мир подождет!*». Если обратиться к представленной выше классификации, то можно сказать, что это связанный слоган, так как он содержит в себе название фирмы. Цель создания данной статусной единицы – развлекательная. Вне контекста в слогане нет ничего смешного, однако присоединение фразы «долги бывшего СССР» привносит в статусную единицу комический элемент. Всё дело в семантике слова *долг* – *взятое взаймы, то, что необходимо отдать или возместить*<sup>4</sup>. Таким образом, долг по-разному воспринимается людьми. Должники часто стремятся отсрочить выплату долгов, а заёмщики заинтересованы в получении представленной в долг суммы в кратчайшие сроки. Поэтому для СССР, выступающего в качестве должника, который занимал у всего мира, йогурт Даниссимо – позитивное явление, ведь это возможность отсрочить выплату долгов, в чём заверяет слоган компании.

В статусных единицах довольно часто реализуются реминисценции на популярные в народе рекламы, фильмы и т. п. Обязательное условие – использование тех фраз, которые на слуху у большинства пользователей сети. Фильмы из сферы арт-хаус или рекламы с иностранных каналов вряд ли будут пользоваться интересом при создании статусов.

Статусная единица «*Раньше мои волосы были сухими и безжизненными, но я попробовала Dove, теперь они мокрые и шевелятся*» построена на антитезе. В её структуре нет полноценного слогана компании, однако присутствует отсылка к рекламе шампуня Dove, которая начинается словами: «Сухие и безжизненные волосы?...» и далее содержит в себе перечисление плюсов продаваемого товара. В статусной единице два ряда антитез: *сухие – мокрые, безжизненные – шевелятся*. Причём в первом случае противопоставляются два прилагательных, во втором – прилагательное и глагол, которые становятся антонимами лишь в контексте, уже содержащем в себе антитезы. Связь между этими словами прослеживается при рассмотрении их семантики. Так, в прилагательном *безжизненный* – *лишённый признаков жизни, мёртвый*<sup>5</sup> главным признаком жизни избирается наличие двигательной активности у предмета (переносное значение слова *безжизненный* – *невыразительный, вялый*<sup>6</sup>). Это делает возможным употребление слова *шевелиются* в качестве антитезы к прилагательному *безжизненные*. Волосы, которые шевелятся, то есть могут проявлять двигательную активность, приобретают признак жизни. Стоит отметить, что в статусной единице имеет место реминисценция на известный древнегреческий миф о Медузе Горгоне, которая вместо волос имела много шипящих и извивающихся змей. Таким образом, в статусной единице присутствует *поэтика безобразного*, которая «определяется как проявляемое внешне нарушение определенной внутренней меры бытия» [2]. Шевелящиеся волосы – это нарушение нормы, на что и делает акцент автор статусной единицы.

Интересна, например, статусная единица «*Не верю!*» – *сказала Скалли, стоя на борту НЛО*», где обыгрывается известная фраза

К. С. Станиславского, которая достаточно часто употребляется в речи людей. Известный театральный режиссёр, преподаватель и актёр К. С. Станиславский использовал эту фразу, желая подчеркнуть ненатуральность, фальшь в игре актёра. Она существует и в следующем виде: «Станиславский сказал бы: «Не верю!». Кроме того, в статусной единице содержится отсылка к популярному в недавнем времени фантастическому сериалу «Секретные материалы». Употребление фразы Станиславского в речи главной героини сериала «Секретные материалы» Скалли придаёт комичность статусной единице, поскольку сюжет сериала заключался в расследовании дел, в которых зачастую были задействованы сверхъестественные силы, но чаще всего – иноземный разум. Упомянутый неопознанный летающий объект, на борту которого находилась Скалли при произнесении фразы «не верю», усиливает комический эффект от статусной единицы.

Таким образом, мы пришли к выводу, что использование при создании статусных единиц слоганов и иных элементов из известных продуктов современного искусства привносит в статусные единицы комизм. Вместе с тем, подобные статусные единицы способны нести в себе скрытый философский подтекст и прочитываться двояко.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Рекламный слоган [Электронный ресурс]. – 2006. – Режим доступа: [http://ru.wikipedia.org/wiki/Рекламный\\_слоган](http://ru.wikipedia.org/wiki/Рекламный_слоган). – Дата доступа: 21.04.2012.
- 2 Безобразное [Электронный ресурс]. – 2001. – Режим доступа: <http://ru.wikipedia.org/wiki/Безобразное>. – Дата доступа: 21.04.2012.

УДК 811.512.161‘367.3:81‘322.4

С. А. Лашукевич (Минск)

### **КРИТЕРИИ И ПРИНЦИПЫ СЕГМЕНТАЦИИ ТУРЕЦКОГО ПРЕДЛОЖЕНИЯ В СИСТЕМЕ МАШИННОГО ПЕРЕВОДА ДЕЛОВОЙ ДОКУМЕНТАЦИИ**

*В статье представлена разработка формальных правил деления на минимальные сегменты турецкого предложения в деловых письмах. Деловая переписка осуществляется в рамках официально-делового стиля речи, который характеризуется четко структурированными, нормированными предложениями с отчетливой тенденцией к однозначности их компонентов. Результатом исследования стала база правил, позволяющих сегментировать предложение и создать словарь фразеологических единств с их русскими эквивалентами.*

Построение электронного корпуса текстов и словаря является неотъемлемой задачей при создании алгоритмов для программ машинного перевода текстов. Отличительная особенность словарей программ машинного перевода заключается в содержании и способе оформления словарной статьи. Словарная статья такого словаря может иметь вид некоего фразового единства, в котором элементы имеют прямую взаимосвязь и семантическую завершенность, что позволяет подобрать к ним соответствующий русский эквивалент. Однако построение такого словаря требует обработки определенного количества текстов общей тематической и стилистической направленности. Обработка таких текстов, в первую очередь, сводится к сегментации предложений, которая основывается на принципах, вырабатываемых исследователем с опорой на синтаксическую, морфологическую и лексико-семантическую специфику исследуемого языка.

После деления исходного текста на сегменты он проверяется на наличие повторяющихся сегментов. Найденные дублированные сегменты надо перевести только один раз, а в следующих их появлениях программа обработки подставит в текст уже имеющийся перевод.

Сегментация производится делением на такие смысловые блоки, как предложения и их части, ориентируясь на обозначенные границы фрагментов [1, с. 197].

Формальное деление на сегменты турецкого предложения представляет собой задачу в рамках создания базы данных для системы фразеологического перевода турецких текстов на русский язык. Поскольку материалом для проводимого исследования служат материалы деловой переписки, оформленные по правилам официально-делового стиля в соответствии с литературной нормой языка [2, с. 14], то, на наш взгляд, представляется возможным осуществить сегментацию предложений таких писем. Предлагаемая сегментация подразумевает движение от более крупных к максимально мелким элементам. Сегментация турецкого предложения опирается как на синтаксические, так и на морфологические сигнализаторы границ элемента.

При поступлении турецкого предложения для сегментации на фразеологические элементы процесс сегментации состоит из нескольких этапов. Рассмотрим эти уровни на примере делового письма-извинения на турецком языке.

*Первый этап* сегментации турецкого предложения основывается на делении предложения с опорой на знаки препинания. Запятые, двоеточия и другие знаки препинания будут сигнализировать о границе сегмента.

*Второй этап* осуществляется с опорой на лексические сигнализаторы границ сегментов. Этими сигнализаторами являются союзы, частицы и послелоги, которые, наряду, со словарем вводных устойчивых конструкций, являются частью базы данных для сегментации и перевода.

Далее, необходимо выделить сегменты, которые присущи официально-деловой переписке, а именно конструкты со значением времени и пространственных отношений, а также конструкции со словом *taraf*



(сторона) в значении «стороны договора или переговоров» и различные местоимения, которые также представлены в базе данных.

Таким образом, сегментация турецкого предложения осуществлялась с опорой на пунктуацию и служебные слова. Дальнейшая сегментация будет проводиться с опорой на морфологические показатели слов.

*Третий этап* сегментации предполагает выделение глаголов, имен, а также их производных, по принципу их идентификации по конечным элементам или их комбинаций (аффиксов). Благодаря тому, что в морфологии турецкого языка представлены две единственные парадигмы имен и глаголов со своими стандартными показателями, формальное отнесение словоформы к классу имен или глаголов является выполнимой задачей, при условии создания базы данных с вариантами и комбинациями возможных аффиксов. Продемонстрируем процесс сегментации на примере предложения:

*Sizin kozmetik alanında önde gelen ithalatçılardan biri olduğunuzu öğrendiğimiz için size kendimizi tanıtmak istiyoruz.* „Мы хотим представить себя вам, поскольку узнали, что вы являетесь одним из лидирующих импортеров в области косметики“.

В турецком предложении, кроме точки, отсутствуют знаки препинания и вводные конструкции, поэтому сегментация начинается с выделения служебных слов – союзов, частиц и послелогов.

*Sizin kozmetik alanında önde gelen ithalatçılardan biri olduğunuzu öğrendiğimiz için* - послелог

*size kendimizi tanıtmak istiyoruz.*

Теперь в полученных сегментах выделяются местоимения.

*Sizin* – личное местоимение

*kozmetik alanında önde gelen ithalatçılardan biri olduğunuzu öğrendiğimiz için* – послелог

*size* – личное местоимение

*kendimizi* – возвратное местоимение

*tanıtmak istiyoruz.*

В состав полученных сегментов не входят конструкты с временным, пространственным значением, поэтому дальнейшее сегментирование проходит на основании морфологических признаков элементов предложения, а именно на основании выделения конечных аффиксов и их комбинаций в словах. Таким образом, выделяются грамматически маркированные имена.

*Sizin* – личное местоимение

*kozmetik alanında* – имя существительное в единственном числе и в местном падеже

*önde* – имя существительное в единственном числе и местном падеже

*gelen ithalatçılardan* – имя существительное во множественном числе и исходном падеже

*biri* – имя числительное в винительном падеже

*olduğunuz öğrendiğimiz için* – послелог

*size* – личное местоимение

*kendimizi* – возвратное местоимение  
*tanıtmak istiyoruz*.

Далее отрезки сегментируются с опорой на выделенные глагольные формы тем же способом, что и имена, а именно путем определения их конечных аффиксов или их комбинаций.

*Sizin* – личное местоимение

*kozmetik alanında* – имя существительное в единственном числе и в местном падеже

*önde* – имя существительное в единственном числе и местном падеже

*gelen* – глагол в форме деепричастия настоящего или прошедшего времени

*ithalatçılardan* – имя существительное во множественном числе и исходном падеже

*biri* – имя числительное в винительном падеже

*olduğunuzu* – глагольное имя (масдар)

*öğrendiğimiz* – глагольное имя (масдар)

*için* – послелог

*size* – личное местоимение

*kendimizi* – возвратное местоимение

*tanıtmak* – глагольное имя (масдар) в форме инфинитива

*istiyoruz* – глагол в финитной форме настоящего продолженного времени множественного числа первого лица.

После такой детальной сегментации целесообразным будет объединить некоторые сегменты по формальному признаку. В этой связи послелог необходимо «привязать» к предшествующему ему слову: *öğrendiğimiz + için = öğrendiğimiz için*. Элементы *önde* и *gelen* необходимо также объединить, так они являются устойчивым выражением с русским эквивалентом лидирующий. В результате получены следующие элементы предложения по формальному признаку: *Sizin / kozmetik alanında / önde gelen / ithalatçılardan / biri / olduğunuzu / öğrendiğimiz için / size / kendimizi / tanıtmak / istiyoruz*.

В процессе проведения сегментации турецкого предложения по названным принципам движение идет в описанном направлении, и в результате остаются минимальные сегменты, которые вместе со своими русскими эквивалентами и словарными формами входят в базу данных. После определения конечных сегментов предложения целесообразным будет установить их синтаксическую принадлежность и соотнести со структурой предложения на русском языке.

#### СПИСОК ЛИТЕРАТУРЫ

1 Кит, М. Об использовании повторяемости сегментов текста для повышения эффективности процесса перевода / М. Кит // Вестник РГГУ. – № 11 (73). Научный журнал. – М., 2011. – 300 с.

2 Коробейникова, Л. С. Документационное обеспечение делового общения: учеб. пособие / под ред. проф. Д. А. Ендовицкого, Л. С. Коробейниковой, О. М. Купрюшина. – М.: Магистр, 2009. – 302 с.

## **ФРЕЙМОВОЕ МОДЕЛИРОВАНИЕ КАК СПОСОБ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ (НА МАТЕРИАЛЕ ТУРИСТИЧЕСКОГО ДИСКУРСА)**

*В статье анализируется теория фреймов как способ репрезентации знаний. На материале туристического дискурса изучается специфика фрейма как концептуальной модели. В аспекте фреймового моделирования рассматриваются возможности экспертных систем для создания текстов определенного типа дискурса.*

На сегодняшний день внимание научного сообщества все чаще обращено на изучение механизмов восприятия, хранения и переработки информации в сознании человека. Исследования в этой области носят междисциплинарный характер. Выявление сущности данных процессов интересно как с точки зрения психологии, так и лингвистики. Работы в рамках этих областей знания вместе с разработками в сфере нейробиологии, математической логики, философии и некоторых других наук образуют общую межпредметную исследовательскую область когнитивистики, где предпринимаются попытки моделирования человеческого знания.

Работы по созданию некоторой модели знания отмечаются многообразием применяемых теорий и способов. Одним из наиболее интересных является фреймовый подход репрезентации знаний.

В соответствии с этим подходом предполагаемой основой для протекания процессов обработки информации, поступающей из окружающей среды, служат особые когнитивные структуры или концептуальные схемы. В сознании человека имеется целый ряд таких структур, именуемых фреймами. Они служат для представления стереотипных ситуаций [1, с. 2; 2, с. 16]. При этом ситуация понимается более широко: как любое явление, объект или процесс окружающей действительности. В качестве основной структурной единицы в системе репрезентации знания теория фреймов использует понятие или его аналог, вокруг которого строится представление. Фрейм, таким образом, является моделью обобщенного до определенного уровня представления о каком-либо предмете, процессе, явлении окружающей действительности.

У каждого взрослого человека имеется большой набор когнитивных структур для репрезентации как простых представлений о предметах, к примеру, распознавание мяча или стула, так и целых классов ситуаций, например сценарий знакомства, совершения покупки в магазине, прощания, модель построения текста аргументации либо описания некоторого объекта и т. д. Часто структура ситуаций вышеупомянутой категории более сложного уровня

схожа у представителей одного культурного сообщества и, предположительно, имеет конвенциональную природу [2, с. 17]. Значит, характеристика соответствующих фреймов позволяет сформировать понятие, как об этом сообществе в целом, так и об отдельном его представителе.

Фреймовый подход может применяться и в исследованиях различных типов дискурса (под дискурсом понимается определенная совокупность тематически соотнесенных текстов [3, с. 16]). С точки зрения теории фреймов в сознании человека существует определенная когнитивная модель, в соответствии с которой происходит восприятие, понимание и продуцирование текста данного типа. Так, в рамках кулинарного дискурса в сознании человека имеется фрейм кулинарного рецепта, а для туристического дискурса, например, фрейм текста описания туристического объекта. Семантически близкие или связанные фреймы объединяются в системы фреймов, в совокупности репрезентируя весь тип дискурса.

Фрейм можно представить в виде графа с узлами и связями между ними. Связи между узлами или терминалами представлены условиями, предъявляемыми к заданиям этих терминалов. Простейшие из условий именуется маркерами [1, с. 2]. Это, к примеру, требование, чтобы в качестве задания терминала был предмет определенной формы.

К преимуществам теории фреймов относят учет ожиданий и предположений о ситуации. Такая информация размещается в «заданиях отсутствия», заранее готовых значениях терминалов. «Задания отсутствия» представляют собой некоторого рода предположение о ситуации, которое может подтвердиться, и тогда терминал оказывается уже заполненным. Если же ожидание оказывается ошибочным, задания легко заменяются более подходящими [1, с. 2]. Таким образом, структура фрейма является достаточно гибкой, и в определенной ситуации некоторые элементы могут быть эксплицированы, а некоторые терминалы или целые подструктуры, где потенциально может размещаться информация, несущественная в данной ситуации, остаются незаполненными и присутствуют в вероятностной форме. К примеру, фрейм текста описания замка потенциально содержит подструктуры, терминалы которых могут быть заполнены данными о местонахождении этого туристического объекта, его истории, особенностях конструкции и внешнего вида, историко-культурной значимости данного замка и некоторыми другими. При этом в различных текстах описания замка не обязательно представлена информация по всему перечню возможных параметров. Некоторые терминалы или даже целые подструктуры фрейма так и остаются в вероятностной имплицитной форме, например, не приводятся какие-либо характеристики интерьера замка, что с точки зрения создателя текста не представляет особой важности в данной ситуации: для этого замка.

С каждым фреймом ассоциирована информация нескольких типов. Во-первых, это указание, каким образом следует применять эту структуру данных. Во-вторых, что возможно последует за выполнением фрейма. Наконец, что необходимо предпринять в случае несоответствия ситуации

выбранной модели [1, с. 2]. Воспринимая новую информацию, человек выбирает определенную концептуальную схему интерпретации из ряда имеющихся. Набор таких структур в сознании человека определяется его принадлежностью к определенному национально-культурному сообществу, а также его опытом, уже полученными знаниями о мире и некоторыми другим личностными характеристиками. Когда выбранный фрейм каким-либо образом не соответствует ситуации либо задания терминалов не могут выполнить предъявляемых условий, по сети поиска информации подбирается более подходящая ситуационная модель [1, с. 3]. Избирая наиболее оптимальную когнитивную структуру, человек заполняет ее имеющимися данными и, изменяя некоторые элементы, приспособливает данный фрейм или систему фреймов к конкретной ситуации, тем самым усваивает новое знание.

Таким образом, репрезентация знания и интерпретация механизмов когнитивной обработки информации в терминах фреймов позволяет построить модели некоторого текста или всего типа дискурса. Следует отметить, что такие структуры данных могут быть представлены в компьютерной системе, а значит и работа по обработке и созданию соответствующих текстов может быть до известной степени автоматизирована. Система сама подбирает необходимую структуру и вероятное ее заполнение в соответствии с типом дискурса, выбранным пользователем. Специалисту необходимо заполнить предлагаемую системой модель имеющимися данными. Использование экспертной системы такого рода облегчит и ускорит процедуру создания текстов определенного типа. Кроме того, построенный в соответствии с выделенным для данного культурного сообщества фреймом текст будет успешен с точки зрения его восприятия адресатом.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Минский, М. Фреймы для представления знаний / М. Минский. – М.: Мир, 1979. – 77 с.
- 2 Ван Дейк, Т. А. Язык. Познание. Коммуникация: пер. с англ. / Т. А. ван Дейк // Сост. В. В. Петрова; под. ред. В. И. Герасимова. – М.: Прогресс, 1989. – 132 с.
- 3 Чернявская В. Е. Дискурс как объект лингвистических исследований / В. Е. Чернявская // Текст и дискурс. Проблемы экономического дискурса: сб. науч. тр. / С.-Петербург. гос. ун-т экономики и финансов. – СПб., 2001. – С. 16.

УДК 801.316.4

М. В. Половец (Гродно)

#### **СТРУКТУРА ГЛОССАРИЯ ТЕРМИНОВ КОМПЬЮТЕРНОЙ ЛИНГВОДИДАКТИКИ**

*Статья посвящена особенностям составления глоссария компьютерной лингводидактики. Несформированность и открытость терминосистемы компьютерной лингводидактики обуславливает необходимость*

*отражения ее динамики в глоссарии терминов. Описывается модель структуры такого глоссария.*

Актуальность разработки глоссария терминов той или иной области связана с необходимостью исследования и систематизации терминологии. В традиционном понимании глоссарий – словарь узкоспециализированных терминов, использованных в определенном произведении. В переводоведении глоссарии служат для унификации терминологии в переводе при работе над одним текстом несколькими специалистами. Необходимость систематизации терминов возникает в процессе развития области знаний, появления новых задач, роста, развития и совершенствования ее терминосистемы.

Компьютерная лингводидактика как область знания зародилась в 80-х годах XX века. Хотя компьютерная лингводидактика развивается очень динамично, на данный момент она находится в стадии становления, а терминология этой области представляет собой открытую, интенсивно развивающуюся, формирующуюся и неустойчивую систему. Несмотря на то, что компьютерная лингводидактика использует терминологию нескольких научных направлений (лингвистики, лингводидактики, информационных технологий), на основе которых она сформировалась, термины приобретают иное значение при переходе в новую формирующуюся область компьютерной лингводидактики. Междисциплинарный характер направления обусловлен общими интеграционными процессами во всех сферах человеческой деятельности [1, с. 7]. Несформированность и открытость терминосистемы компьютерной лингводидактики обуславливает специфику глоссария терминов данной области. Традиционно глоссарий терминов отражает статическое состояние сформировавшейся терминосистемы. Но при составлении глоссария терминов компьютерной лингводидактики необходимо учитывать динамический характер терминосистемы как ее основной характеристики. Рассмотрение терминологии в динамическом аспекте позволит выявить как динамику развития терминологии в целом, так и изменения в объеме отдельных понятий, что, в свою очередь, поможет спрогнозировать тенденции в развитии целевой терминологии. Поэтому терминология компьютерной лингводидактики представляет особый интерес для исследователей, а именно процесс ее формирования, «исследования живых изменчивых и сложных фактов становления языка науки и обоснование терминопорождения реальными запросами самой науки» [2, с. 3]. На данный момент не существует лексикографических продуктов, отражающих предметную область компьютерной лингводидактики. В настоящее время такой глоссарий разрабатывается на материале специализированного электронного журнала «Language Learning and Technology», который находится в свободном доступе в Интернете [3]. Данный англоязычный журнал посвящен компьютерной лингводидактике, он освещает наиболее актуальные проблемы направления в прикладном и теоретическом аспектах. Несмотря на то, что «Language Learning and

Technology» спонсируется Гавайским (the University of Hawai‘i) и Мичиганским (Michigan University) университетами, представители которых и составляют редколлегию, в материалах журнала соблюдаются нормы международного английского. К публикации принимаются материалы от авторов со всего мира: Китай, Япония, Австралия, Сингапур, страны Европы и Африки. Сравнительно небольшое количество материалов, написанных авторами из русскоговорящих стран. Поэтому целесообразным представляется построение систематизированного глоссария англоязычных терминов с описаниями объема закрепляемых ими понятий на русском языке. Это позволит не только показать уровень развития информационных технологий в мире, дать стимул для развития области знаний в русскоязычной традиции, но и в последующем гармонизировать русскоязычную и даже белорусскоязычную терминологию. Описание объема понятий, закрепляемых терминами, основывается на исследовании всех контекстов встречаемости соответствующего термина в исследуемом языковом материале. Текст – среда функционирования терминов, обращение к тексту абсолютно необходимо для выявления объема закрепляемых терминами понятий в случаях неустойчивых терминосистем.

Динамика терминологии проявляется в изменении количества единиц в терминосистеме (в большую либо меньшую сторону), а также в изменении значения терминов. Следовательно, глоссарий будет содержать лексические лакуны и различные варианты толкований значения термина в зависимости от номера журнала, в котором он встречается. Для отражения специфики терминологии компьютерной лингводидактики предполагается включение дополнительных полей в структуру глоссария. Учитывая все отмеченные выше особенности, модель словарной статьи можно представить следующим образом:

- термин, лексический вход словарной статьи (начальная форма на английском языке);
- описание понятия (на русском языке), основанное на анализе контекстов в максимальной степени отражающих признаки понятия и специфику сочетаемости термина;
- источник (номер журнала, названия статей или рубрик, где термин встречается);
- примеры контекстов;
- частота встречаемости в исследуемом материале (количество раз).

В случае изменения значения термина он будет иметь несколько вхождений в глоссарии в соответствии с количеством понятий, им обозначаемым. Таким образом можно отразить изменения в плане содержания терминологической единицы.

С целью отражения системных связей необходимо группировать термины по тематическому принципу. Для этого первоначально выделяются ключевые (гнездообразующие) термины. Затем определяются термины, образованные от ключевых терминов и входящие в данное тематическое гнездо.

Глоссарий терминов, отражающий развитие терминосистемы, может послужить инструментом стандартизации терминологии компьютерной лингводидактики.

#### СПИСОК ЛИТЕРАТУРЫ

1 Головкин, Е. А. Социальный заказ и становление компьютерной лингводидактики как области научного знания / Е. А. Головкин, Т. Н. Ломтева // Вестник Ставропольского государственного университета. №1 (60). – Ставрополь: Изд-во СГУ, 2009. – 182 с. – С. 5–9.

2 Ивина, Л. В. Лингво-когнитивные основы анализа отраслевых терминосистем (на примере англоязычной терминологии венчурного финансирования) / Л. В. Ивина // Учебно-методическое пособие. – М. : Академический проект, 2003. – 304 с.

3 Language Learning and Technology // Mode of access <http://llt.msu.edu/>. – Date of access: 27.09.2012.

УДК 37.013:811.161.3'373.2:004

А. Ф. Рогалев, В. А. Рогалев (Гомель)

### **МАТЕРИАЛЫ САЙТА «БЕЛОРУССКАЯ ОНОМАСТИКА И ЭЗОТЕРИЧЕСКАЯ ЛИНГВИСТИКА» (BEL-ONOMASTIKA.UCOZ.RU) В УЧЕБНОЙ РАБОТЕ**

*Рассматривается возможность использования интернет-ресурсов в учебной работе по лингвистическим дисциплинам. Описывается структура сайта и его тематическое многообразие.*

Среди существующих интернетовских сайтов не так уж и много полифункциональных. Одним из них является представляемый нами сайт *bel-onomastika.ucoz.ru*, выполняющий три основных функции – научно-познавательную, просветительную, учебную.

**1. Общие сведения о сайте.** Сайт *bel-onomastika.ucoz.ru* непосредственно связан с компьютерной лингвистикой – направлением в современном языкознании, ориентированным на использование компьютерных программ и технологий для обработки и систематизации данных, для моделирования функционирования языка в тех или иных условиях, ситуациях и жизненных сферах, в том числе в сфере учебной деятельности. Это направление активно разрабатывается и пропагандируется в последние годы на кафедре русского, общего и славянского языкознания Гомельского государственного университета имени Франциска Скорины.

Тематика материалов сайта – топонимия, антропонимия Беларуси, топонимическое краеведение, этнолингвистика, лингвокультурология, лингвопалеонтология, философия языка, теория языкознания, когнитивная лингвистика.

Назначение сайта – поиск скрытого смысла в именах, названиях, словах, устойчивых выражениях; расширение ключевых концептов не только



белорусской (русской) или восточнославянской, но и общечеловеческой концептосферы.

Задачи сайта – расширение и углубление знаний о языке как знаковой системе, органически связанной с индивидуальным и коллективным (этническим) сознанием.

Определение *эзотерическая* (лингвистика) в названии сайта следует понимать с учётом этимологического значения греческого по происхождению слова *эзотеризм* (*эзотерика*) – «направленный внутрь». В связи с этим область эзотерической лингвистики может быть определена следующим образом: познание языковых фактов как «ключей» к глубинным информационным пластам человеческого сознания и всей ноосферы Земли.

Объединение в рамках сайта проблем ономастики и эзотерической лингвистики вполне оправдано, поскольку имена собственные всех видов обладают очень высоким информационным потенциалом, причём в значительной степени скрытым. Об этом с высот философии языка основательно рассуждал в своё время в работе «Философия имени» (1927 год) русский философ и филолог Алексей Фёдорович Лосев (1893–1988).

Выявление скрытого смысла в словах, именах и названиях ведётся нами на фоне текстов – мифов, легенд, преданий, сказок и целых художественных произведений, например, сочинений Н. В. Гоголя – выдающегося мастера литературной эзотерики.

Источники материалов для сайта – преимущественно личные разработки, исследования и наблюдения администратора сайта профессора А. Ф. Рогалева. Подход к материалам иных авторов – аналитический: при сохранении авторского права и системы ссылок используются пересказ, изложение основных идей, положений и выводов. Нередко размещается не весь текст соответствующего сочинения, а только основные и наиболее интересные его фрагменты с комментариями.

Принцип отбора материалов – строгий, критический подход с целью установления достоверности и степени научности тех или иных интерпретаций.

**2. Учебные дисциплины**, с которыми непосредственно связаны материалы сайта: «Введение в языкознание», «Общее языкознание», «Философия языка», спецкурсы «Введение в ономастику», «Введение в антропонику», «Введение в топонимику», «Имя и образ».

Материалы сайта активно используются студентами при подготовке докладов, рефератов, курсовых и дипломных работ.

### **3. Основные разделы и страницы сайта:**

а) **«Библиотека»** (включает большие по объёму материалы – монографии, словари, научные статьи, в частности, о происхождении географических названий и фамилий);

б) **«Исследования»** (средние по объёму файлы, связанные с общей тематикой сайта, например, «Названия стран и народов» в нескольких частях, «Гидроним *Нарочь*», «Ойконим *Туров*», «Символ розы», «Концепт

*Бог*», «Концепт *любовь*», «Скрытый смысл астронима *Марс*», «Чернобыльская легенда и название *Чернобыль*», «Этнонимы», «Сущность микротопонима», «Концепт *нос* и символ носа в художественной литературе», «Имя и образ в драме А. Н. Островского *Гроза*», «Символ тамплиеров», «Философские законы и их проявление в языке», «К вопросу о сакральной сущности азбуки», «Нейролингвистическое программирование» и многие другие);

в) **тематические страницы информационно-аналитического плана**, в частности, «Топонимические предания», «Мифологические символы», «Исторические символы», «Космические символы», «Цветовая символика», «Эзотерические мысли», «Природа языка», «Язык воды», «Мыслеформы», «Язык идеофонов», «Птичий язык», «Язык пейзажа», «Язык цветов», «Язык и действительность», «Язык и познание», «Язык и общество», «Язык и этнос», «Язык и структура психики», «Инобытие и язык» и другие.

Определённое место на сайте отведено заданиям к семинарам, библиографии, тестам. Одна из страниц сайта предназначена для диалога и носит название «Вопрос – ответ». Диалог активно ведётся и на странице «Гостевая книга».

#### **4. Некоторые ключевые идеи сайта.**

- Язык – не сугубо социальный, а био-социальный феномен, имеющий звуко-кинетическую природу.

- Помимо звуко-кинетического языка, люди активно используют мысленно-энергетический (психоэнергетический) язык, однако не акцентируют на этом внимания.

- Звуко-кинетический и мысленно-энергетический языки есть «среда бытия» человека и общества.

- Сознание человека способно декодировать не только знаки основного для людей звуко-кинетического языка, но и язык чисел и язык космических символов, которые являются кодом для открытия и уяснения тайн мироздания.

- Числовой ряд формулы именованного лица, личное имя, отчество и фамилия человека непосредственно связаны с его жизненной программой.

- Личное имя обычно аллегорически указывает на жизненную задачу личности. Отчество очерчивает базу (основу), позволяющую человеку выполнять жизненную задачу. Фамилия опять-таки иносказательно определяет характер и специфику жизнедеятельности, сопутствующие обстоятельства, способы достижения цели.

- Каждый произнесённый звук и каждая написанная буква, а соответственно, каждое слово и весь текст в совокупности составляющих его слов «заряжены» энергетическим спектром, в котором сочетаются ментальная, эмоциональная (астральная) и психическая (духовная) энергии соответствующего качества, воздействующие в том или ином плане на адресата (собеседника, слушателя, читателя).

- В созданном тексте, даже в одном произнесённом слове закодирована сущность личности.

Материалом для анализа на сайте служат не только словесные тексты, но и тексты иного рода, в частности, фотографии природных объектов, неба, воды, растений, цветов (эзотерические фото Владислава Рогалева) и фотокопии эзотерических картин Н. К. Рериха. Считывание информации из таких текстов развивает как элементарную наблюдательность, так и способность воспринимать проявления инобытия.

Материалы сайта «Белорусская ономастика и эзотерическая лингвистика» служат дополнением к темам названных выше основных учебных дисциплин и спецкурсов, пробуждают у студентов интерес и желание исследовать, помимо явного (открытого) языкового уровня с хорошо известными и во многом уже познанными фонетической, грамматической и лексической подсистемами, также второй (скрытый) уровень языка, позволяющий выявлять в известном неизвестное. Такой подход к языковым фактам соотносится с одной из современных концепций языка, которая формулируется так: «Язык есть двойной код».

### **5. Иллюстрация с сайта (фрагмент анализа языкового факта).**

Концепт «язычник». *«...Есть люди и есть боги. Следовательно, существуют разные уровни сознания. Поэтому мир многомерен, или многоуровнев. Каждый уровень бытия характеризуется определённой энергетической частотой. Улавливание частот этих уровней означает общение с данным уровнем бытия.*

*Человек благодаря семи его основным энергоцентрам может переключаться на частоты разных уровней многомерного мира и общаться с минералами, растениями, животными, ангелами, архангелами, богами – с астральным миром, в котором, как и в человеческом мире, также есть свои градации – положительные и отрицательные.*

*Восприятие информации в виде энергоимпульсов с разных уровней бытия равнозначно знанию языков этих миров. Того, кто владеет разными языками, то есть умеет улавливать разнообразные частоты многоуровневого мира, считывает и передаёт информацию с этих уровней и на эти уровни, издавна и называли язычником, а ещё – ведуном, ведаманом, то есть «обладателем запредельных знаний – Вед.*

*Все люди во все времена делились на ведунов (меньшая часть) и невежд (большинство). Последние не могут воспринимать многоуровневый мир, их сознание функционирует только на частотах плотного физического мира Земли».*

### **6. Пример творческого задания студентам при изучении темы «Язык символов» в курсе «Философия языка».**

*«На берегу синего моря стоит золотая пирамида, позади вершины которой на фоне голубого неба находится красный солнечный шар. На основании пирамиды крупными буквами написано “ДА!” По условию задания требуется истолковать символический рисунок и представить своё толкование в виде комментария на сайте.*

**7. Фрагмент теста** по теме «Язык и действительность» в курсе «Общего языкознания». Студенты находят правильные суждения из представленного перечня и кратко аргументируют свой выбор.

*«Внешний мир шире, разнообразнее и глубже, чем внутренний мир».*

*«Внутренний мир шире, разнообразнее и глубже, чем внешний мир».*

*«Внешний и понятийный мир тождественны».*

*«Понятийный мир человека не соответствует внешнему миру».*

*«Реально только нечто вещественное».*

*«Внутренний мир человека реален».*

*«Всё реальное всегда объективно»...*

**8. Пример факультативного задания повышенной сложности.**

*«Свяжите понятия “мыслеобраз” и “имя”, иллюстрируя эту связь примерами из жизни или художественной литературы».* Помещённые на сайте ответы обсуждаются затем на семинарском занятии на тему «Имя и личность» в курсе «Философии языка».

**9. Образцы получаемых через сайт контрольных тест-вопросов по темам курсов «Философия языка» и «Общее языкознание»:**

а) Раскрыть подтекст выражений *«светлые мысли»*, *«тёмные мысли»*, *«творческая атмосфера»*, *«невыносимая атмосфера»*.

б) Каково соотношение понятий «этнос» и «язык»?

в) Можно ли словом менять реальность?

г) Что такое «волшебные слова»? Как их найти?

УДК 81'32

Л. В. Рычкова (Гродно)

## **ЛИНГВИСТИЧЕСКИЕ КОРПУСЫ В СОЦИОКУЛЬТУРНОМ И ЭКОЛИНГВИСТИЧЕСКОМ АСПЕКТАХ**

*Развитие корпусной лингвистики и наращивание объема и разнообразия компьютерных языковых ресурсов актуализировало социокультурные аспекты создания и использования лингвистических корпусов. Назрела также необходимость учета эколингвистических факторов при создании новых корпусов. В статье рассматриваются основные социокультурные, собственно социальные и эколингвистические факторы, учет которых необходим в современной корпусной лингвистике.*

В последнее время наблюдается резкое возрастания интереса к корпусам текстов и корпусной лингвистике как таковой со стороны неспециалистов в этой междисциплинарной области знания. Возрастает также количество попыток создания корпусов непрофессионалами. С другой стороны, отмечается лавинообразный рост корпусов, создаваемых в полном

соответствии с принятыми профессиональным сообществом стандартами; возрастает также их типологическое разнообразие. Таким образом, настало время для рассмотрения корпусов текстов не только как специфического вида компьютерных языковых ресурсов и основы для реализации наиболее современной, доказательно-экспериментальной парадигмы языкознания [1], но и как фактора культуры, отражающего определенный период цивилизационного развития человечества [2]. Открытость корпусов (например, Национального корпуса русского языка [3]) широкому кругу пользователей за пределами лингвистического сообщества существенно влияет на изменение социокультурной значимости не только собственно корпусной лингвистики, но и языкознания в целом. Наконец, возможности, предоставляемые динамичными корпусами (о динамичных корпусах см., например, [4]) для проведения многоаспектного лингвистического мониторинга и контент-анализа, весьма актуальны для эколингвистики. Очевидно, что предпосылки возникновения таких двух междисциплинарных направлений современного языкознания, как корпусная лингвистика и эколингвистика, были разными. Очевидно, однако, и то, что без учета эколингвистических факторов корпус теряет одно из своих необходимых, сущностных свойств – репрезентативность, а эколингвистика – способность к прогностичности, без которой она вынужденно становится констатирующей наукой и сужает поле исследований, дублируя до определенной степени культуру речи.

Нами уже отмечалось ранее [2], что языковой материал, включенный в корпус, представляет собой часть культурного наследия народа. Корпусы, как наиболее репрезентативный тип компьютерных языковых ресурсов, могут служить задаче сохранения малочисленных языков мира (*language preservation*), находящихся под угрозой исчезновения. Представляется целесообразным рассматривать репрезентативные полнотекстовые корпусы как специфическую электронную среду (*digital environment*), релевантную не только для собственно лингвистических исследований (ограничены принятой для каждого конкретного корпуса концепцией языковых объектов, подлежащих разметке), но и для решения широкого круга междисциплинарных гуманитарных проблем, в том числе в сопоставительном и контрастивном аспектах. По-видимому, именно национальные (генеральные) лингвистические корпусы должны и могут стать ядром «национальной киберинфраструктуры» (*national cyberinfrastructure*), поскольку призваны представлять все состояния общенационального языка, а не только его литературную форму. Независимо от принятой системы разметки, выбор типов языковых объектов и процедуры лингвистического аннотирования отражает как сложившиеся в рамках описательной лингвистики стереотипы, так и национальные лингвистические традиции. К социокультурным аспектам лингвистических корпусов следует отнести и тот факт, что они неизбежно отражают узус, или существующие узуальные практики, коммуникативные стереотипы и шаблоны. Развитие корпусных

технологий позволяет создавать корпуса устной (в том числе разговорной) речи, мультимедийные корпуса, размеченные с учетом невербальных компонентов коммуникации. Это, в свою очередь, превращает корпуса в незаменимый источник материала для исследования межкультурных отличий и социокультурных представлений. Практически не исследована «справочная» функция лингвистических корпусов. Сегодня она пока еще недостаточно, но все же реализуется лишь при обучении иностранным языкам. По-прежнему основными источниками справочной информации для самого широкого круга пользователей остаются словари и грамматики. И если англоязычные словари и грамматики все чаще базируются на корпусных данных, то для русского и, тем более, для белорусского языков – это все еще дело будущего. В определенной степени восполнить эту лакуну для русского языка предполагается с помощью образовательного портала Национального корпуса русского языка [5].

К социокультурным аспектам лингвистических корпусов можно отнести также такие собственно «социальные» характеристики, как оповещение пользователей посредством электронной почты либо мобильной телефонии о пополнении / изменении языковых ресурсов; сервисы по консультированию начинающих пользователей либо по выбору наиболее адекватного конкретной исследовательской задаче языкового ресурса; возможности поделиться с коллегами и просто с заинтересованными лицами результатами использования конкретного корпуса; доступность для оценки различных компонентов и видов обеспечения того или иного языкового ресурса. Очевидно, что многие из социальных аспектов языковых ресурсов аналогичны тем из них, которые являются релевантными для цифровых библиотек (*digital libraries*) [6] и электронных исследовательских сред. Речь идет о создании социально-адекватной инфраструктуры, важную роль в которой должны сыграть мультилингвальные сервисы (*multi-lingual services*).

Проблема инфраструктуры остается пока за пределами корпусной лингвистики. Так, каждый языковой ресурс либо не доступен массовому пользователю, либо доступен как самостоятельная поисковая система. Не существует пока взаимосвязанных, «распределенных», корпусов либо лингвистических баз данных. В связи с наращиванием объема свободно доступных лингвистических, в том числе корпусных, языковых ресурсов возникает насущная потребность в инструментах, позволяющих осуществлять одновременный поиск и обработку данных, полученных из разных источников. Каждая из гуманитарных наук, «работающих» с текстом, нуждается в разных типах данных, тем не менее, существующие системы разметки корпусов этого не учитывают.

Эколингвистические аспекты корпусной лингвистики до сих пор также не исследовались и не учитывались при разработке лингвистических корпусов. Очевидно, однако, что значимость эколингвистического аспекта корпусов особенно велика в следующих случаях: 1) «документирования» малочисленных языков мира, находящихся под угрозой исчезновения;

2) изучения национальных / региональных и иных особенностей различных идиомов – коммуникативных вариантов языков (нациолектов, региолектов, урбанолектов и т. д.); 3) отражения естественного многоязычия языкового социума в условиях «полиглотизма» (Б. А. Ларин). До сих пор не создано ни одного «смешанного» корпуса, отражающего соответствующий языковой материал. В силу ограниченных возможностей существующих корпус-менеджеров, все корпуса «привязаны» к конкретному языку / языкам, поскольку разметка текстов вручную слишком дорогостояща.

Подводя итог, необходимо затронуть еще одну проблему, связанную с аутентичностью языкового материала, положенного в основу любого лингвистического корпуса, и степенью достоверности получаемых в результате поиска данных. Лингвоцентрическая, по сути, традиция создания корпусов не позволяет должным образом решать проблемы полноты, репрезентативности корпусов, а также аутентичности и достоверности корпусных данных. У неподготовленного пользователя при работе с корпусами, как правило, теряется ощущение текста, т. е. самого языкового материала, а критерии поиска, по которым компьютер отбирает искомые объекты, представляются ему не очевидными. Принятая в современной корпусной лингвистике инфраструктура допускает превращение корпусных данных в виртуальные, которые легко можно подвергнуть обработке, как до этого был подвергнут обработке первичный языковой материал в процессе осуществления мета- и собственно лингвистической разметки текстов. Таким образом, назрела потребность в разработке такой инфраструктуры, которая способна обеспечить аутентичность, достоверность и целостность корпусных данных. Особенно это важно, когда исходный языковой материал представлен в рукописной форме и в основу корпуса кладется электронный аналог.

Лингвоцентричность создания корпусов проявляется также в том, что их разновидности обусловлены лишь дифференциацией языкового материала. Ни один из существующих корпусов не создается по тематическому принципу либо специально для решения интердисциплинарной исследовательской задачи.

Итак, учет социокультурных и эколлингвистических факторов важен для развития корпусной лингвистики, а повышение ее социальной значимости сопряжено с решением задачи разработки адекватной инфраструктуры.

#### СПИСОК ЛИТЕРАТУРЫ

1 Пиотровский, Р. Г. Доказательно-экспериментальная парадигма современного языкознания / Р. Г. Пиотровский // Структурная и прикладная лингвистика. – 2008. – Вып. 7. – С. 5–13.

2 Рычкова, Л. В. Социокультурная значимость национальных корпусов / Л. В. Рычкова, С. Н. Киеня // Компьютерная лингвистика: научное направление и учебная дисциплина: сборник научных статей. – Гомель: ГГУ им. Ф. Скорины, 2010. – Вып.1. – С.132–135. – Режим доступа: <http://www.lib.grsu.by/library/data/resources/catalog/161646-347269.pdf>.

3 Национальный корпус русского языка / электронный ресурс. – Режим доступа: <http://www.ruscorgo.ru>. – Дата доступа: 01.10.2012.

4 Баранов, А. Н. Введение в прикладную лингвистику / А.Н. Баранов. – М., 2001. – С. 116.

5 Образовательный портал Национального корпуса русского языка / электронный ресурс. – Режим доступа: <http://studiorum.ruscorgo.ru>. – Дата доступа: 01.10.2012.

6 Many More than a Million: Building the digital environment for the age of abundance. – Washington, DC: Council on Library and Information Resources, 2008. – 32 p.

УДК 81'32

Л. В. Рычкова, А. Ю. Станкевич, Ю. А. Бодак (Гродно)

## ПРИНЦИПЫ СОЗДАНИЯ ГАЗЕТНОГО КОРПУСА СМИ ГРОДНЕНЩИНЫ

*Описаны принципы отбора первоисточников для формирования базы газетного корпуса Гродненщины. Выявлено, что архитектура корпуса задается системой метаразметки, аналогичной той, что применяется в Национальном корпусе русского языка. Отмечено, что в качестве обязательного вида лингвистического аннотирования избрана морфологическая разметка. Определены основные направления использования создаваемого корпуса.*

### **1. Отбор первоисточников.**

Формирование базы корпуса – первый и чрезвычайно важный этап в технологии создания любого лингвистического корпуса. От этого этапа во многом зависят и выбор адекватного формата разметки, и возможности использования конечного продукта. Отбор источников для корпуса, таким образом, должен быть прагматически обусловленным. Нами было предложено 6 принципов (изложены ниже), в опоре на которые в качестве источников для корпуса СМИ Гродненщины были избраны следующие газеты: «Гродзенская праўда», «Вечерний Гродно», «Милицейский вестник», «Перспектива». Назовем и охарактеризуем принципы отбора именно этих СМИ в качестве источников создаваемого корпуса.

#### **Принцип территориальности.**

Отобранные газеты издаются в Гродно, распространяются по подписке и в розницу в Гродно и в Гродненской области.

#### **Принцип широты тематического охвата.**

Газета «Милицейский вестник» освещает деятельность органов внутренних дел, ведет профилактику преступлений и правонарушений, пропагандирует законопослушный образ жизни. Нередкими в газете являются новости культуры и образования. Газета «Перспектива» освещает новости сельского хозяйства, отражает различные происшествия, уделяет внимание



духовной жизни человека (в каждом выпуске есть статья либо очерк, посвященные раздумьям о духовных и моральных ценностях). Газета «*Вечерний Гродно*» оперативно откликается на все события, происходящие на Гродненщине, и, соответственно, отражает самый широкий круг проблем, касающихся жизни города. Газета «*Гродзенская праўда*» отражает новости науки, культуры, политической и экономической жизни города, дает информацию о событиях городской жизни, происшествиях, печатает официальную информацию.

### **Принцип официальности – неофициальности.**

Вид организации-учредителя определяет языковые особенности газет, их тематику, наличие в них рекламных объявлений и степень официальности. Так, газета «*Милицейский вестник*» – ведомственная. Ее учредителем является УВД Гродненского облисполкома. Эта газета издается с 1996 года, выходит 1 раз в неделю, ее объем составляет от 2 до 4 печатных листов, тираж – 8 470 экземпляров. Постоянные рубрики: «Задержания», «Имидж», «Криминал», «Кримобоз», «Новости», «Наши ветераны», «Мир искусства». Учредитель газеты «*Перспектива*» – Гродненский районный совет депутатов. Газета издается с 1939 года, выходит 2 раза в неделю, тираж составляет 10 777 экземпляров, объем – 6 печатных листов. Постоянные рубрики: «Бегущей строкой», «Земля и люди», «История с фотографией», «Официально», «На пульсе жизни», «Память», «Происшествия», «События, факты», «Социум», «Человек и закон». Учредитель газеты «*Вечерний Гродно*» – ЗАО «Редакция газеты «Вечерний Гродно»». Издается с 2000 года, выходит 1 раз в неделю, тираж составляет 25 170 экземпляров. Данная газета – частная, неофициальная, поэтому отражает несколько иную тематику, чем вышеназванные газеты. Постоянные рубрики: «Площадь новостей», «Репортер», «Социум», «Отдыхай» и др. Учредителями газеты «*Гродзенская праўда*» являются областной и городской исполнительные комитеты, областной и городской Совет депутатов, ОУИРП «Редакция газеты «Гродзенская праўда»». Газета издается с 1939 года, выходит 3 раза в неделю, общий тираж недели составляет 75 240 экземпляров, объем – 8 печатных листов. Постоянные рубрики: «Гродно. Городские новости», «Ход в историю», «ЧП», «Суд да дело», «Жизнь, как она есть», «Что волнует?» и др. В газете регулярно печатаются тексты официальных документов. Таким образом, по степени официальности отражаемой информации издания могут быть распределены на 3 группы: 1) «*Гродзенская праўда*»; 2) «*Перспектива*», «*Милицейский вестник*»; 3) «*Вечерний Гродно*».

### **Принцип нацеленности на аудиторию.**

Для определения целевой аудитории отобранных газет были изучены их сайты в Интернете, что позволило выяснить, что целевой аудиторией газеты «*Вечерний Гродно*» является социально активная часть населения города, а газета «*Гродзенская праўда*» ориентирована на самый широкий круг читателей. Несмотря на то, что подобная информация на сайтах газет «*Перспектива*» и «*Милицейский вестник*» отсутствует, по тематике

данных периодических изданий можно предположить, что они «нацелены» на людей среднего и старшего возраста, а также на представителей определенных профессий (востребованы, прежде всего, в правоохранной и аграрной сферах).

#### **Принцип временного охвата.**

В корпус войдут тексты газет, начиная с 2011 г. Данный год выбран как точка отсчета не случайно: двадцатилетнее существование Республики Беларусь как независимого государства позволило сформировать свою систему ценностей, отличную от современной российской культуры, а это неизбежно находит сегодня отражение в текстах СМИ. Региональный принцип также важен, так как за то же время сформировалась новая, не советская, идентичность, что, не в последнюю очередь, выразилось в повышении интереса населения к историческим событиям и знаковым фигурам каждого региона, что также находит отражение в текстах СМИ.

#### **Принцип языкового разнообразия.**

Газеты «Гродзенская праўда», «Перспектива», «Вечерний Гродно» издаются на русском и белорусском языках, в них преобладают публикации на русском языке; газета «Милицейский вестник» полностью издается на русском языке. Следует отметить, что в газете «Гродзенская праўда» публикации на белорусском языке встречаются гораздо чаще, нежели в других двуязычных газетах. Таким образом, создаваемый корпус будет состоять из письменных текстов различных газетных жанров и будет преимущественно русскоязычным.

## **2. Архитектура корпуса как отражение системности представления языкового материала.**

Метаразметка, аналогичная метаразметке газетных текстов в Национальном корпусе русского языка (НКРЯ), позволит работать во всем корпусе и выделять, при необходимости, подкорпусы для работы, например, с текстами только одной газеты либо для формирования тематически выровненных параллельных корпусов. В этом случае предлагается «полексемное» выравнивание, при котором задавалась бы связь между вариантами слова (например, «Гродно – Гродна – Гародня»).

В газетном корпусе НКРЯ тематическая разметка не поддерживается. Однако в основном корпусе НКРЯ для *Нехудожественных текстов* можно, указав значение параметра *сфера функционирования: публицистика*, задать группу параметров «Тематика текста» (при этом группа «Тематика текста» доступна и для других, не только публицистических, нехудожественных текстов; множества имен газет заданного таким способом подкорпуса и Газетного корпуса не равны). Группа «Тематика текста» содержит всего 23 общих параметра неименованной подгруппы (*адміністрацыя і ўправленне; армія і ўзброеныя канфлікты; астралогія, парапсіхалогія, эзотэрыка; бізнес, каммерцыя, эканоміка, фінансы; дом; здароўе і медыцына; досуг, зрэлішча і развлечення; мастацтва і культура; крымінал; адукацыя; палітыка і грамадзянская жывасць; права; прырода; прадукцыя*);

*путешествия; религия; сельское хозяйство; спорт; строительство и архитектура; техника; философия; частная жизнь; наука и технологии*) и 31 параметр подгрупп «Гуманитарные науки», «Естественные науки», «Прикладные науки» (по 12, 10 и 9 параметров соответственно) для определения темы по разделу и отрасли науки.

Предполагается, что система метаразметки создаваемого корпуса СМИ Гродненщины будет включать следующие группы параметров: «Издание», «Тип издания», «Автор текста», «Псевдоним», «Год публикации», «Структурный элемент», «Жанр текста», «Тема текста» (последний параметр будет применяться только для разметки тематического корпуса).

Использование описанной выше системы метаразметки определяет архитектуру создаваемого корпуса СМИ Гродненщины, который будет включать: 1) основной корпус (с возможностью формирования подкорпусов по любой комбинации параметров); 2) тематически размеченный подкорпус; 3) тематически выровненный параллельный корпус.

### **3. Виды лингвистической разметки.**

Обязательным видом лингвистического аннотирования, применяемого в создаваемом корпусе, является **морфологическая разметка**, аналогичная системе морфологической разметки газетного корпуса НКРЯ. Известно, что при разработке этого вида лингвистического аннотирования было использовано описание русской морфологии, воплощенное в «Грамматическом словаре русского языка» А. А. Зализняка.

**Семантическая разметка** будет использована фрагментарно (иногда как дополнение к тематической метаразметке, а также, при необходимости, при полексемном выравнивании). **Синтаксическая разметка** на первом этапе создания корпуса не предусмотрена.

### **4. Направления использования корпуса.**

Предполагается, что предложенный подход к отбору первоисточников позволит создать корпус, с помощью которого можно будет решить следующие задачи:

- 1) выявить особенности узуса русского языка Гродненщины;
- 2) изучить тематический и жанровый охват входящих в корпус периодических изданий;
- 3) исследовать идиостили конкретных авторов.

Данный корпус, отражая особенности узуса русской речи, может использоваться в лингводидактике, особенно в преподавании русского языка иностранцам; при изучении стилистики и культуры речи; в исследовательской деятельности специалистов в области лингвистики, социолингвистики, специалистов по сравнительному изучению русского и белорусского языков.

Особую значимость создаваемому корпусу придает возможность формирования белорусскоязычного регионального подкорпуса. Полагаем, что это позволит выявить те темы, которые преимущественно отражаются на страницах СМИ в текстах на белорусском языке, а также получить

естественные контексты функционирования специальных единиц и лингвистически релевантную информацию об отражаемых в СМИ подъязыках.

В перспективе возможно создание по предлагаемому образцу других региональных корпусов СМИ.

УДК 37.091.3:004.77:81.161.1'243

Н. А. Сивакова (Гомель)

## **ИСПОЛЬЗОВАНИЕ ИНТЕРНЕТ-РЕСУРСОВ ПРИ ОБУЧЕНИИ РУССКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ**

*В статье представлено описание обучающих интернет-ресурсов, позволяющих преподавателю русского языка как иностранного совершенствовать процесс обучения, сделать его более мобильным, дифференцированным, индивидуальным и интерактивным.*

Основная цель изучения русского языка как иностранного (РКИ) – формирование коммуникативной компетенции, все остальные цели (образовательная, воспитательная, развивающая) реализуются в процессе осуществления этой главной цели. Коммуникативный подход подразумевает обучение общению и формирование способности к межкультурному взаимодействию. Помочь преподавателю в решении этой непростой задачи может сочетание традиционных методов обучения и современных информационных технологий, в том числе и компьютерных с использованием ресурсов Интернет. Использование интернет-ресурсов на занятиях по РКИ позволяет сделать процесс обучения мобильным, дифференцированным, индивидуальным и интерактивным.

Можно выделить три аспекта использования интернет-ресурсов на занятиях по РКИ. Во-первых, Интернет как тема занятия, как предмет обсуждения; во-вторых, Интернет как средство урока, обеспечивающие новые мультимедийные возможности; в-третьих, Интернет как вспомогательное средство для подготовки занятия, как огромная база данных. В Интернете содержится множество разнообразных ресурсов, содержащих учебно-методическую и научную информацию, что позволяет организовывать оперативную консультационную помощь, моделировать научно-исследовательскую деятельность, проводить виртуальные учебные занятия (семинары, лекции) в режиме реального времени. Наиболее используемыми ресурсами в организации учебно-методической деятельности являются электронные учебники, чаты, он-лайн словари, блоги и виртуальные обучающие среды.

В Интернете можно найти большое количество сайтов, разработанных специально для тех, кто изучает русский язык, а также сайты, содержащие объемные базы ссылок на разнообразные интернет-ресурсы по РКИ [1].

Представленные интернет-ресурсы можно использовать как на уроке, так и для самостоятельной работы дома. Одной из таких разработок является курс «Время говорить по-русски!» [2], рассчитанный на усвоение элементарного уровня владения русским языком, что в классификации уровней владения языками, разработанной на кафедре теории и методики преподавания русского языка как иностранного факультета международных отношений Белорусского государственного университета, соответствует базовому уровню – уровню минимальной коммуникативной достаточности (УМКД), который предполагает пользование русским языком как средством коммуникации на минимально допустимом коммуникативном уровне в сфере повседневного общения в пределах определенного программой обучения круга ситуаций [3, с. 4].

Проект «Время говорить по-русски!» осуществлен Центром Международного Образования МГУ и, по мнению разработчиков курса А. Петановой и Ю. Коваленко, является доступным, интересным, динамичным, веселым, не перегруженным научной терминологией и грамматикой, что на начальном этапе овладения языком является весьма важным.

Основной курс состоит из отдельных уроков, построенных по модульному принципу. Основу обучающей программы составляет многосерийный анимационный фильм, герои которого пластилиновые куклы. Разработчики подчеркивают, что именно такая форма передает условность происходящего и, вместе с тем, имитирует реальность в виде игры, помогая обучающимся преодолеть психологический барьер: «"Пластилиновая реальность" позволяет относиться к ее героям не совсем серьезно – над ними можно посмеяться и поиронизировать» [2].

Каждый урок содержит определенный блок лексико-грамматических заданий: представлена тема урока, связанная с конкретной темой грамматики и соответствующие лексико-синтаксические конструкции. Специально подобранные упражнения после каждого урока позволяют усвоить, повторить и закрепить материал. Уроки расположены по степени усложнения грамматики и расширения объема лексики. Однако свободная навигация дает возможность выбора: последовательный переход от урока к уроку или свободное перемещение по темам (разделам грамматики).

Каждый урок состоит из следующих частей: сцена (или сцены); словарь к уроку; тест к уроку; страноведческая информация. На странице предъявления урока можно получить конкретную информацию о том, каковы темы отдельных сцен урока, какие грамматические аспекты и активные лексические конструкции рассматриваются в данном уроке и с какой страноведческой информацией (связанной с темой данного урока) можно познакомиться. Например, тема первого урока «В аэропорту» разбита на отдельные сцены: «Паспортный контроль», «Профессии», «Багаж», состоящие из нескольких диалогов, которые можно смотреть в разных режимах: без субтитров, с субтитрами на русском языке, на родном языке и с отдельным текстом диалога.

Каждая сцена состоит из следующих элементов: предпросмотровые задания, знакомство с диалогом, задания на понимание диалога, грамматический комментарий, словарь к данной сцене, система упражнений по фонетике, грамматике и лексике, письменные и речевые упражнения.

Предпросмотровые задания включают в себя лексико-грамматические задания до просмотра сцены фильма: предъявление новой лексики, запоминание и проверка запоминания лексических единиц, знакомство с ключевыми фразами. Благодаря наглядности и аудиозаписи (вся новая лексика снабжена иллюстрациями, которые сопровождает голос диктора), становится возможным качественное усвоение материала.

Задания, направленные на закрепление изученной лексики, позволяют самостоятельно производить проверку материала как на занятии, так и дома. В отличие от традиционных печатных учебников интерактивный курс «Время говорить по-русски!» позволяет учащимся слышать звучащую речь и контролировать свое произношение. По мере продвижения по курсу вперед задания усложняются, большинство из них представлены в форме интерактивных игр.

Помимо основного курса, в общую систему проекта также входят разделы «Грамматический справочник», «Тестирование», «Библиотека», «Страноведение», «Медiateка», «Словари», связанные между собой гипертекстовыми ссылками, объединяющими весь курс в единое виртуальное пространство.

Данный интернет-проект удобен не только для самостоятельной работы, но и для использования в качестве дополнительного средства обучения на занятиях по РКИ. При этом он не заменяет преподавателя, а только дополняет его, играя роль инструмента, который при грамотном использовании значительно повышает эффективность педагогического процесса.

Таким образом, использование интернет-ресурсов при обучении РКИ может значительно обогатить учебный материал звуковой и зрительной наглядностью, сделать его более понятным; индивидуализировать процесс обучения без дополнительных затрат аудиторного времени; программировать и контролировать дидактические ситуации; рационализировать учебный процесс в целом; сделать сам процесс обучения интересным и комфортным для всех обучающихся; создать условия для качественного формирования речевых навыков во внеаудиторное время в процессе самостоятельной работы студентов.

#### СПИСОК ЛИТЕРАТУРЫ

1 Русский язык для жителей СНГ и Балтии [Электронный ресурс]. – Режим доступа : <http://www.russianforall.ru/conf/links.php>. – Дата доступа : 12.09.2012.

2 Время говорить по-русски [Электронный ресурс]. – Режим доступа : <http://speak-russian.cie.ru>. – Дата доступа : 26.04. 2010.

3 Лебединский, С. И. Русский язык как иностранный. Типовая учебная программа для иностранных студентов I–IV курсов нефилологических специальностей / С. И. Лебединский, Г. Г. Гончар. – Мн.: БГУ, 2002. – 170 с.

Д. В. Сичинава (Москва)

**РУССКАЯ КОНСТРУКЦИЯ С *БЫЛО*  
И БЕЛОРУССКИЙ ПЛЮСКВАМПЕРФЕКТ:  
СОПОСТАВИТЕЛЬНЫЙ АНАЛИЗ  
НА МАТЕРИАЛЕ БЕЛОРУССКО-РУССКОГО  
И РУССКО-БЕЛОРУССКОГО ПАРАЛЛЕЛЬНЫХ КОРПУСОВ<sup>1</sup>**

*В статье рассматривается на корпусном материале русская конструкция типа пошёл было в сопоставлении с белорусским плюсквамперфектом (форма типа пайшоў быў). Выявлены некоторые особенности менее изученной белорусской формы – прежде всего, формальные и семантические архаизмы по сравнению с русским было, а также особенности семантической сочетаемости. Особо рассмотрены условия появления той или иной формы, связанные с «вольностью» в переводе.*

**1. Введение. Грамматические исследования на параллельном корпусе.**

С 2011 года в ходе совместного проекта ИРЯ РАН (команда Национального корпуса русского языка) и Института языка и литературы (ныне в составе Центра исследований белорусской культуры, языка и литературы НАНБ) разрабатываются параллельные белорусско-русский и русско-белорусский корпуса. Напомним, что параллельный корпус включает в себя оригинальные тексты и сопоставленные им (выровненные предложение в предложении) переводы этих текстов. Он имеет многообразные применения, в частности – исследовательские, когда ведётся конкретное сопоставление грамматических и лексических структур языков на материале текстов (см. подробное рассмотрение вопроса в [16]).

При сопоставительных (контрастивных) исследованиях на материале корпуса исследованию подвергается как весь массив текстов (оригинальных и переводных), так и конкретные соответствия между оригиналом и переводом. В частности, значение имеют *модели перевода* – то есть наиболее частотные способы перевода той или иной лексемы или конструкции оригинала – и *стимулы перевода*, то есть, напротив, те выражения в оригинале, на основании которых переводчики избирают те или иные выражения в новом тексте.

Ранее мы проводили исследование на материале англо-русского и немецко-русского параллельного корпусов [11], где выделили ряд типичных стимулов, вызывающих в переводе русскую конструкцию с *было* (типа *пошёл было*). Оказывается, что она возникает, в частности, в случаях, когда

---

<sup>1</sup> Статья написана при поддержке Программы Президиума РАН «Корпусная лингвистика» № 36-П и совместного гранта РГНФ-БРФФИ № 11-24-01004a/Bel.

в оригинале имеются ограничивающие небольшой промежуток времени обстоятельства вроде англ. *for a moment* или нем. *einen Augenblick lang* «на мгновение», а также в соответствии с английским прогрессивом, английским и немецким плюсквамперфектом и рядом других глагольных времён. Эта информация позволяет существенно дополнить картину семантики русской конструкции с *было*, вырисовывающуюся на материале одного только одноязычного корпуса (например, Национального корпуса русского языка), а также и сведения о семантике соответствующих глагольных форм в германских языках.

## 2. Формальное соотношение конструкций.

Если мы обратимся к исследованию этой же конструкции (*пошёл было*) в белорусско-русском и русско-белорусском параллельных корпусах (текущий объём – 1,6 млн словоупотреблений; дополнительно привлекались данные также из других источников), то увидим, что наша задача в чем-то упрощается, но одновременно в чем-то усложняется. Дело в том, что в белорусском языке (в отличие от западноевропейских) имеется конструкция, распределение которой весьма близко к русской конструкции с *было*. Это её непосредственное этимологическое соответствие, а именно плюсквамперфект, традиционно известный в грамматиках как «давнепрошедшее время» (даўномінулы час) или «сложная форма прошедшего времени» (складаная форма прошлага часу): *пайшоў быў*.

С формальной точки зрения белорусская конструкция заметно архаичнее русской по целому ряду параметров. Как и в древнерусском и украинском плюсквамперфекте, а также в северорусских говорах, в ней сохранено согласование вспомогательного глагола (в великорусских центральных диалектах оно утрачено в конце XVI в.). Форма эта хорошо известна по старобелорусским памятникам [5, с. 285–287], [1]. Правда, в белорусском, как и в украинском (частица *було*), в качестве варианта существует и несогласуемая частица *было*, однако она сравнительно редка: в корпусе встретилось только 6 примеров, из них 5 – в оригинальных текстах, причём интересно, что только один передан в переводе русской конструкцией с *было*. И. Я. Лепешев считает конструкцию с несогласованным *было* ошибочной и возникшей под влиянием русского языка [7, с. 177–178], хотя его ссылаясь на то, что согласуемая форма «паслядоўна выкарыстоўваецца аўтарытэтнымі аўтарамі», не подтверждается корпусом – среди использующих оба варианта есть такие классики, как Владислав Короткевич и Василь Быков. В одном из случаев несогласованное *было* встретилось в корпусе в контексте причастия: *Загрымелі замоўкшыя было партызанскія кулямёты, сціх на хвіліну нямецкі агонь* [Міхась Лынькоў. Міколка-паравоз (1936)]; подобное употребление (почему-то только для косвенных падежей) И. Я. Лепешев признаёт нормативным, приводя пример из Янки Купалы. Три раза параллельный корпус в качестве соответствия для русского причастия с частицей *было* даёт более характерные для белорусского языка придаточные определительные со стандартной согласуемой конструкцией



(*Валя задержала за рукав собравшегося было бежать Климку.* [Н. А. Островский. Как закалялась сталь (ч. 1) (1930–1934)] // *Валя затрымала за рукаў Клімку, які ўжо збіраўся быў бегчы* [Мікалай Астроўскі. Як гартавалася сталь (ч. 1). Пераклад Л. Ганкіна]). В русском языке частица *было* в составе нефинитных конструкций появляется только в 1820-е–1830-е годы (см. подробнее [10], [12]).

С точки зрения линейного порядка слов (препозиции или постпозиции вспомогательного элемента) белорусская конструкция также архаичнее, хотя и не столь радикально. Как известно, в древности элемент *был-* в составе плюсквамперфекта соблюдал закон Вакернагеля и в соответствующих условиях уходил в препозицию [4]. В белорусских текстах препозиция встретилась в 25 % примеров (в точности как в русском языке XVIII века, [12]), что несколько выше показателей препозиции в разных фрагментах одноязычного НКРЯ (от 17 % до 24 %), причём в 3 примерах белорусской препозиции соответствует русская постпозиция, а еще в 6 случаях в соответствующем русском тексте конструкция с *было* отсутствует:

*Корзун быў памкнуўся на дапамогу, але Савацееў утрымаў яго і шапнуў: – Глядзі!. [Уладзімір Шыцік. Скачок у нішто (1967)] // Корзун рванулся было на помощь, но Саватеев удержал его и шепнул: – Смотри!.* [Владимир Шитик. Скачок в ничто. Перевод Б. Мескина и В. Шитика]

### **3. Семантическая сочетаемость.**

Семантически эти конструкции в целом довольно близки – особенно по сравнению со степенью формальных различий. Как и русская форма с *было*, белорусский плюсквамперфект преимущественно употребляется в контекстах нарушения нормального хода ситуации (применительно к русскому языку см., прежде всего работы А. Барентсена и Ю. П. Князева [15], [6]). Толкование, близкое к тому, какое обычно предлагается для русской частицы *было*, даёт И. Я. Лепешевым [7, с. 177–178]: «дзеянне, якое пачалося ці магло пачацца, але перапынілася іншым дзеяннем». Правда, многие традиционные описания (как и для большинства славянских плюсквамперфектов) дают здесь привычные ярлыки «давнопрошедшее действие» и «предшествование в прошедшем» (что отражает традицию описания западноевропейских языков с таксисом). Ср.: «для перадачы даўномінулага часу дзеяння, якое паперэднічала іншаму завершанаму дзеянню» [14, с. 144]; аналогичную альтернативу своему толкованию предлагает и Лепешев. Обращение к корпусу показывает, однако, существенно большую близость белорусской конструкции к русской – в большинстве случаев одна просто переводится при помощи другой, и это делает картину (но только на первый взгляд) гораздо более простой, чем в случае перевода с немецкого или английского. Сложности же заключаются в том, что отклонения от такого прямолинейного сценария, о которых речь уже шла, всё же встречаются не менее чем в трети случаев и достаточно разнообразны по своим причинам – они зависят как от реальных различий сочетаемости конструкций общего происхождения в близкородственных языках, так

и от стратегий конкретных текстов, авторов и переводчиков. В русском языке употребительность данной конструкции сильно варьирует в зависимости от автора (по данным первопроходческой работы А. Барентсена [15]); то же верно, насколько можно судить, и применительно к белорусским писателям, а также к переводчикам с одного языка на другой.

В работе [10] мы рассмотрели лексико-семантическую сочетаемость русской конструкции с *было* на материале текстов одноязычного основного подкорпуса в составе Национального корпуса русского языка после 1950 г. Объём этих текстов многократно превосходит русские тексты в составе параллельных корпусов. Были получены данные о том, в каком проценте случаев конструкция с *было* сочетается с тем или иным характерным для неё семантическим классом глаголов, а именно: глаголами желания (типа *хотеть, вознамериться*), глаголами начала действия или инициирования ситуации (типа *начать, взлететь, обратиться [с речью], раскрыть рот*), глаголами достижения состояния (типа *построить, умереть*), глаголами попытки (*пытаться, постараться*) и глаголами ментальной деятельности или эмоций (*решить, обрадоваться*). Исследовались также некоторые другие классы, отчасти пересекающиеся с данными, на которых мы в этой статье не останавливаемся.

Рассмотрим отдельно лексико-семантическую сочетаемость конструкции с *было* в оригинальных русских текстах, включённых в параллельный корпус, и в русских текстах, переведённых с белорусского. Затем проведём аналогичные подсчёты для плюсквамперфекта (включая редкую конструкцию с несогласованным *было*) для белорусских текстов – оригинальных и переводных. В отдельной колонке приведём цифры по большому одноязычному русскому корпусу из [10]:

Тип глаголов	НКРЯ	рус. ориг.	рус. перев.	бел. ориг.	бел. перев.
Желания	23 %	24,3 %	32,3 %	17,2 %	32,1 %
Начала действия / инициирования ситуации	43 %	48,6 %	32,3 %	41,4 %	28,6 %
Ментальной деятельности или эмоций	10 %	13,5 %	4,8 %	8,6 %	17,9 %
Попытки	8 %	8,1 %	16,1 %	5,0 %	7,1 %
Достижения результирующего состояния	14 %	5,4 %	17,7 %	24 %	14,3 %

Вероятно, данные этой таблицы удастся сделать более надёжными по мере увеличения объёма корпусов, но общие тенденции заметны и в наличном материале. Прежде всего, показатели русских оригинальных текстов, включённых в параллельные корпуса НКРЯ, несмотря на сравнительно малый объём, за одним исключением весьма близки к вычисленным на большом корпусе. Единственный пункт, в котором эта репрезентативность нарушается – глаголы результирующего состояния; их процент в попавших в параллельный корпус текстах сильно занижен. Причины

этого не вполне ясны (возможно, это связано с индивидуальной манерой авторов). На этом фоне рельефнее выступает одно отличие белорусской конструкции от русской, которое, впрочем, было бы хорошо заметно и при русских данных, более близких к данным большого НКРЯ. Пока у нас нет большого репрезентативного корпуса белорусского языка (одноязычный корпус, доступный онлайн – Corpus Albaruthenicum В. А. Кощенко – составлен на материале научных текстов, для которых интересующая нас конструкция не характерна), сложно судить, насколько цифры, полученные для белорусских оригинальных текстов, отражают сочетаемость конструкции в языке в целом; однако данные этих текстов, связанные с глаголами результирующего состояния, судя по другим цифрам, достаточно надёжны. Процент этих глаголов в составе нашей конструкции очень высок (почти 24 %) не только в оригинальных, но и в переводных белорусских текстах, хотя в последних он предсказуемо снижается под влиянием интерференции со стороны русского. Отметим тот факт, что среди привнесённых переводчиком на русский язык (а не продиктованных оригиналом) конструкций с *было* глаголов достижения состояния не встретилось вовсе.

Эти данные указывают на большую семантическую (а не только формальную) архаичность белорусской конструкции по сравнению с русской. Действительно, широкая сочетаемость с глаголами достижения состояния была обычна в древнерусском, а в XVIII в. еще были распространены контексты чистой ликвидации результата типа *построил было, но разрушил*. Глаголы достижения состояния занимали в русском языке XVIII в. 30 % от лексического наполнения данной конструкции [12]. Распределение семантических классов в белорусской конструкции несколько напоминает русский язык XVIII в. и в других отношениях, например, пониженная частотность глаголов волеизъявления (в русском языке XVIII в. – 16 %) и глаголов попытки (2 %). Сравни следующий пример, где в русский перевод добавлен глагол попытки (вместо \*я было отказался); подробнее о такой стратегии перевода см. в разделе 4:

*Падышоў Антош з пірагамі. Я быў адмовіўся.* [Уладзімір Караткевіч. Дзікае паляванне караля Стаха (1964)]

*Подошел Антош с пирогами. Я было попытался отказаться.* [Владимир Короткевич. Дикая охота короля Стаха. Перевод В. Щедриной]

В остальном по своей грамматической сочетаемости белорусская конструкция похожа на современную русскую; распределение прочих показателей (особенно глаголов начала деятельности и инициирования ситуации), в общем, близко к данным НКРЯ. Другие отличия, в частности, значения, не выражаемые русской конструкцией, можно выделить только при анализе конкретных примеров. И здесь заметно, что эти отклонения во многом связаны с языком конкретных авторов; вероятно, при расширении корпуса процент таких случаев может измениться и в ту, и в другую сторону, но сейчас нам важно само наличие случаев, не характерных для русского языка (по крайней мере, современного литературного). Начнём с

примера, содержащего редкое несогласованное *было*: *Баец падняўся на ногі і, спатыкаючыся, набрыў было да прыступак*. [Васіль Быкаў. Адна ноч (1961)], в переводе просто *побрёл к ступенькам*; своей цели боец достигает (находит автомат), и никакого нарушения хода этой ситуации не происходит. Возможно, выбор конструкции связан с неуверенностью начала действия, или, что вероятнее, с более широкой «макроситуацией» по Ю. П. Князеву [6] – заваленный в помещении, герой долгое время не может из него выбраться. Четыре плюсквамперфекта, отличные от русского «стандарта», встречаются в произведениях Янки Брыля. Все эти примеры задействуют различные глаголы достижения состояния; как мы помним, именно этот класс глаголов отражает наиболее заметное различие между сочетаемостью русской и белорусских конструкций. При этом, как ни парадоксально, все четыре брылевских примера всё же переведены на русский при помощи частицы *было*, что, на наш взгляд, отражает интерференцию в языке переводчиков под влиянием белорусского. Приведём два из этих примеров:

*Праўда, сыры аднойчы былі не ўдаліся, і, каб зацерці сляды, рагатун Рыўка з яшчэ адным такім майстрам, не з нашай вёскі, уз'ехалі з цяжкім возам на стары склеп, і ён праваліўся*. [Янка Брыль. Ніжнія Байдуны (1974–1975)] // *Правда, сыры однажды было не удались, и, чтобы затереть следы, хохотун Ривка еще с одним таким мастером, не из нашей деревни, въехали с груженым возом на старый погреб, и он провалился* [Янка Брыль. Нижние Байдуны. Перевод Д. Ковалёва];

*Перад ваенным пажарам ён пагарэў быў яшчэ чысцей, нават і пагрэбнік тады не засталася*. [Янка Брыль. Глядзіце на траву (1966)] // *До этого пожара он погорел было еще почище, даже и погребка тогда не осталось*. [Янка Брыль. Глядите на траву. Перевод А. Г. Островского]

В первом случае конструкция имеет экспериенциальное значение (*однажды было так, что*). В русских говорах и литературном языке XVIII века ряд таких примеров отмечен, что отражает сближение между частицами *было* и *бывало*. С этой точки зрения белорусские примеры интересны тем, что в них не утрачивается согласование вспомогательного элемента плюсквамперфекта (при том, что частица *бывала* тоже существует). Белорусские конструкции с *было* и *бывала* сближает и относит к «давнопрошедшему» Е. Ф. Карский [5: 376] Во втором случае речь идёт об аннулированном результате в чистом виде, причем даже дважды аннулированном: погоревший дом был отстроен (а потом, как видим, снова погорел). Соответствие для этого примера опять же находится в русском языке XVIII века: *Пожирающий пламень истребил было красоту сего священнаго здания: но промысл Божий неисповедимый, по временам нас наказывающий, се лучшим украшением изобильнейше утешил нас* [Архиепископ Платон (Левшин). Слово при освящении храма (1779)].

Нехарактерное для русского литературного языка употребление этой конструкции в условной контрфактической конструкции вместе с

сослагательным наклонением (показатель наклонения в составе союза *каб*) видим в «Мужицкой правде» Калиновского: *Для таго то і набілі Маскалі Касцюшку, бо каб мужыкі ўсе разам былі збунтаваліся і ўханілі за сякеры, нажы і косы, так бы Маскаль мусіў бы прапасці без паўстання, і мы на век вякоў ужэ былі бы вольныя.* [Канстанцін Каліноўскі. Мужыцкая праўда. № 2 (1862)] // *Оттого-то и побили москали Костюшку, потому что если бы мужики все вместе взбунтовались и взялись за топоры, ножи и косы, то москаль должен был бы пропасть бесповоротно, и мы на веки веков были бы свободными.* [Константин Калиновский. Мужыцкая правда. № 2. По переводу издания 1990 г.]. Такое употребление плюсквамперфекта сослагательного наклонения может быть как продиктовано польским влиянием (ср. аналогичный старобелорусский пример с параллельным польским текстом в [5: 287]), так и отражать диалектные особенности (диалект Калиновского, как и использующего иные нестандартные значения Брыля – гродненский). Употребление *было* в функции *бы* известно в русском языке XVIII в. и несколько раз встретилось также в русском корпусе разговорной речи. Есть основания считать, что оно не чуждо и русской речи Беларуси, во всяком случае *было* парадоксальным образом (если это не опечатка) возникает в переводе на месте тривиального условного наклонения оригинала: *І мне хацелася нешта сачыніць. Гэтае нешта пачыналася б словамі: <...>* [Вітаўт Чаропка. На кругуі свая (1988)] // *Хотелось что-то сочинить. Это что-то начиналось было словами <...>* [Витовт Чаропка. На кругуі свая. Перевод Т. Зарицкой]<sup>2</sup>.

Что касается переводных текстов – как русских, так и белорусских – то параметры исследуемой конструкции в них вообще достаточно сильно отходят от характерных для оригинальных текстов. Данное явление надо связывать с интерференцией и влиянием переводимого текста (а для переводчиков-белорусов, переводящих на русский язык – возможно, и с региональным варьированием русской нормы под влиянием белорусского). Но это не единственная причина наблюдаемых соотношений. Если показатели языка перевода «тянутся» к показателям языка оригинала (достигая некоторых промежуточных значений), такое объяснение естественно, однако в случаях, когда некоторые семантические классы употребляются гораздо чаще, чем даже в языке оригинала, можно искать истоки такого изменения в семантике конструкции. Например, и «переводной русский», и «переводной белорусский» дают резко увеличенные (примерно до того же значения) показатели для глаголов желания. Именно на глаголы желания приходится 60 % «вольных» (привнесённых переводчиком) конструкций с *было*

<sup>2</sup> У Е. Ф. Карского отмечено, по-видимому, еще одно архаичное значение плюсквамперфекта в белорусских говорах — «сдвига начальной точки» [9], или обозначение первого действия в нарративе, отнесённом к неактуальному прошедшему (помимо застывшей формулы *жыла-была*, это такая начальная фраза рассказа, как *раз, у нядзелю, была яна пасціла свінні на полі* [5: 376; Сенно на Витебщине]. Такое употребление представляет большой интерес (известное по берестяным грамотам и тому же русскому *жили-были*, в современных северных говорах оно, судя по всему, сохранилось; см., в частности, [8: 790], [3: 205—206])

в переводах на русский язык. Для русского языка сильно повышена частотность глаголов попытки, для белорусского – глаголов ментальной деятельности или эмоций. Вероятно, в ряде случаев обсуждаемая конструкция привносится в перевод с целью усилить и яснее передать семантику неосуществлённого желания или неудачной попытки, передаваемую в оригинале только глаголом в прошедшем времени; другие семантические группы такого дополнительного усиления не требуют. Это может свидетельствовать о тенденции к модализации семантики нашей конструкции (точка зрения, высказывавшаяся, начиная с А. А. Шахматова, целым рядом учёных, см. подробнее об этой концепции [15]).

*Азірнуцца захацелася яму, ён ужо і намерыўся, але не мог.* [Міхась Стральцоў. Смаленне вепрука (1973)] // Оглянутыся захотелось ему, он уже и намерился было, но не смог. [Михась Стрельцов. Смаление вепря. Перевод В. Тараса]

*В голасе ее ужо не было насмеішкі, было в нем что-то дружеское, приміраючае, і Павка, сбравіішыся нагрубіць гэтай невесты адкуда взяўшейся «барышне», был обезоружен.* [Н. А. Островский. Как закалялась сталь (ч. 1) (1930–1934)] // *У голасе яе ўжо не было насмеішкі, было ў ім нешта сяброўскае, супакойваючае, і Паўка, які ўжо збіраўся быў нагрубіаініць гэтай «паненцы», што ўзялася нямаведама адкуль, быў абяззброены.* [Мікалай Астроўскі. Як гартавалася сталь (ч. 1). Пераклад Л. Ганкіна]

#### **4. Точность и вольность в передаче грамматической конструкции.**

Перевод советского времени с русского языка на языки народов СССР и в обратном направлении (а именно переводы этой эпохи лежат в основе нынешнего состава корпуса) отмечен высокой степенью вольности на разных уровнях, и при составлении и анализе параллельного корпуса это постоянно приходится учитывать. Причины этого разнообразны – общая установка ряда переводчиков, роль автора, участвующего в творческом пересоздании «авторизованного» текста, а нередко и цензурные причины.

Разумеется, за внешней «вольностью» могут стоять тонкие семантические различия, и особенно это касается грамматики. Если применить к грамматике предложенные М. Л. Гаспаровым для лексики стихотворного перевода понятия «коэффициент точности» и «коэффициент вольности» [2], то на основании параллельного корпуса можно получить следующие данные. При переводе с русского на белорусский русская конструкция *было* «сохранена» (переведена белорусским плюсквамперфектом) в 60 % случаев (коэффициент точности), а белорусский плюсквамперфект привнесён переводчиком в 21 % случаев (коэффициент вольности). При переводе с белорусского на русский заметно выше оба показателя – коэффициент точности 70 %, вольности 36 %<sup>3</sup>. Это значит, что переводчики на белорусский язык используют плюсквамперфект

<sup>3</sup> Клаузы (простые предложения), в переводе опущенные или вставленные, из подсчётов вовсе исключены.

вообще сдержаннее, будь то в соответствии с русским *было* или нет, чем переводчики на русский – русское *было*.

Представляется, что наблюдаемая картина зависит от того, что нормативный статус белорусского (как и украинского) плюсквамперфекта в советское время не был признан<sup>4</sup>, в результате чего переводчик, разумеется, чувствовал себя сильнее стеснённым в использовании этой формы, чем писатель, создающий оригинальный текст. Возможно, имеет смысл также апеллировать к обусловленной какими-то внутренними причинами ограниченной продуктивности и/или большей факультативности белорусской формы. Не исключено, пока корпус сравнительно невелик, что это результат индивидуальной практики немногих конкретных переводчиков, особенно заметной для значительных по объёму текстов. Так, Алесь Жук (известный также как оригинальный писатель), переведший на белорусский язык «Мастера и Маргариту» М. А. Булгакова, только один раз передал плюсквамперфектом русскую конструкцию с *было*, использованную в оригинале 11 раз; во всех остальных случаях *было* переводчиком последовательно опускается, иногда заменяясь на различные субституты; явно за такой стратегией стоит осознанная позиция. Для решения этой проблемы нужны дополнительные исследования.

Для анализа семантики конструкции представляют определённый интерес модели перевода, замещающие нашу конструкцию или стимулы, вызывающие привнесение конструкции в переводе. Как и в случае с английским и немецким языками (см. [11] и начало данной статьи), стимулы эти могут быть информативны для изучения семантики конструкций. Например, русская конструкция с *было* в белорусском передаётся сочетанием прошедшего времени и наречиями ограниченной продолжительности (*крыху, толькі*), начала действия (*спачатку*), а также наречия *ужо*. Это вполне соответствует выявленным ранее переводным стимулам для *было* в английских и немецких текстах. В соответствии с белорусским плюсквамперфектом в русском появляется несколько иной набор моделей-замен: *уже, -таки, даже, – но*, к сожалению, для статистической оценки тут данных пока недостаточно.

Исследование грамматических значений и конструкций на материале параллельных корпусов – уже давно хорошо зарекомендовавшее себя направление, активно применяемое в славистике. Нет сомнений, что еще предстоит сделать много работы по корпусному изучению белорусского и украинского языков, которые вообще в мире корпусной лингвистики (особенно на фоне других славянских) пока представлены недостаточно.

---

<sup>4</sup> Подробнее см. [7, с. 178]. В академических пособиях и грамматиках 1950–1980-х годов плюсквамперфект, если упоминался вообще, расценивался как «отклонение от литературной нормы», «особенность разговорного стиля» и т. п.; исключением является ТСБМ [13], дающий эту конструкцию без ограничительных помет. У истоков такого подхода стоял, по-видимому, Е. Ф. Карский, занимавший даже более радикальную позицию: «давнопрошедшее время... в современном белорусском языке совершенно утрачено, лишь в редких случаях находим его отражения» [5, с. 376], «в живой белорусской речи... встречается очень редко и как особая форма прошедшего не ощущается» [5, с. 286–287].

## СПИСОК ЛИТЕРАТУРЫ

- 1 Анічэнка, У. В. Формы плюсквамперфекта ў беларускай пісьменнасці XV–XVI ст.ст. / У. В. Анічэнка // Ученые записки Мозырского педагогического института. – Вып. 1. 1957.
- 2 Гаспаров, М. Л. Подстрочник и мера точности / М. Л. Гаспаров // О русской поэзии. Анализы. Интерпретации. Характеристики. – М., 2001. – С. 361–372.
- 3 Громова, М. М. Функционирование форм плюсквамперфекта в говорах средней Пёзы (Архангельская область) / М. М. Громова // Фонетика и грамматика: настоящее, прошедшее, будущее. К 50-летию научной деятельности С. К. Пожарицкой. – М., 2010. С. 200–207.
- 4 Зализняк, А. А. Древнерусские энклитики / А. А. Зализняк. – М., 2008. – 280 с.
- 5 Карский, Е. Ф. Белорусы: Язык белорусского народа / Е. Ф. Карский. – Выпуски 2–3. – М., 1956.
- 6 Князев, Ю. П. Форма и значение конструкций с частицей было в русском языке / Ю. П. Князев // Сокровенные смыслы. Слово, текст, культура. Сборник статей в честь Н. Д. Арутюновой. – М., 2004. – С. 296–304
- 7 Лепешаў, І. Я. Сучасная беларуская літаратурная мова: Спрэчныя пытанні / І. Я. Лепешаў. – Гродна, 2002.
- 8 Пожарицкая, С. К. О семантике некоторых форм прошедшего времени глагола в севернорусском наречии / С. К. Пожарицкая // Revue des études slaves LXIII/4, 1991. – P. 787–799.
- 9 Сичинава, Д. В. «Сдвиг начальной точки»: употребление некоторых глагольных форм в интродуктивной функции // В. Ю. Гусев, В. А. Плунгян, А. Ю. Урманчиева (ред.). Исследования по теории грамматики. Вып. 4: Грамматические категории в дискурсе. – М.: Гнозис, 2004. – С. 241–274.
- 10 Сичинава, Д. В. Стремиться пресекать на корню: русская конструкция с *было* по корпусным данным // Д. В. Сичинава, К. Л. Киселёва и др. (ред.) Корпусные исследования по русской грамматике. М., 2009. – С. 362–396.
- 11 Сичинава, Д. В. Комплексное исследование одноязычного и параллельного корпусов в грамматических исследованиях / Д. В. Сичинава // Труды Международной конференции «Корпусная лингвистика–2011». – СПб.: СПбГУ, 2011. – С. 316–332.
- 12 Сичинава, Д. В. Частицы *было* и *бывало* в русском языке XVIII века / Д. В. Сичинава // Русский язык в научном освещении. – 2012. – № 2.
- 13 Тлумачальны слоўнік беларускай мовы: Т. 1. Мінск, 1977. – 608 с.
- 14 Шкраба, І. Р. Сучасная беларуская мова: марфалогія / І. Р. Шкраба. – Мінск, 2007.
- 15 Barentsen, A. A. The use of the particle БЫЛО in modern Russian / A. A. Barentsen // Dutch Studies in Russian Linguistics, Amsterdam. Vol. 8 (1986). – P. 1–68.
- 16 Parallel texts: using translational equivalents in linguistic typology / M. Cysouw & B. Wälchli // Sprachtypologie und Universalienforschung (STUF) 2007, 60.2. – P. 95–99.

УДК 811.161.3'373.611

А. В. Солахаў (Мазыр)

### СЛОВАЎТВАРАЛЬНЫЯ І СЕМАНТЫЧНЫЯ АСАБЛІВАСЦІ НОВАЙ ЛЕКСІКІ Ў МОВЕ БЕЛАРУСКІХ ВЭБ-ТЭКСТАЎ

*У артыкуле аналізуюцца словаўтваральныя і семантычныя асаблівасці новай беларускай лексікі, выкарыстоўваемай у розных тэкстах, размешчаных на вэб-сайтах.*



Пачатак ХХІ ст. вызначаецца хуткімі сацыяльна-палітычнымі, культурнымі і навукова-тэхнічнымі зменамі, напрамую звязанымі з паскарэннем моўнай дынамікі, якая адлюстроўваецца ў росце словаўтваральных неалагізмаў, у т. л. аказіянальных утварэнняў, запазычанняў з іншых моў, у працэсах дэархаізацыі лексікі беларускай мовы і інш.

Працэс неалагізацыі, які ахапіў розныя лексіка-граматычныя разрады, шырока адлюстроўваецца ў мове тэкстаў, размешчаных на беларускамоўных сайтах сеткі Інтэрнэт. Электронныя сродкі найбольш аператыўна, чым усе іншыя, фіксуюць розныя лексічныя інавацыі. Шматлікія вэб-тэксты адлюстроўваюць непасрэднасць і вольнасць у выбары аўтарамі таго або іншага словаўтваральнага сродку і адпаведна спосабу словаўтварэння.

Разгледзім словаўтваральныя асаблівасці некаторых інавацый у тэкстах розных жанраў і розных аўтараў.

Вось ліст Avieczki пад назвай «die Wiese. – Нос erat in fatis». «Авечка» – школьніца-выпускніца, удзельніца раённых, гарадскіх і рэспубліканскіх алімпіяд па беларускай мове. Яна добра валодае лексічнымі, словаўтваральнымі і стылістычнымі магчымасцямі беларускай мовы. У тэксце яе ліста ўжываецца паўтара дзесятка слоў, якія не фіксуюцца слоўнікамі беларускай мовы і вызначаюцца словаўтваральнымі і семантычнымі асаблівасцямі.

Так, назоўнікі *раён* і *гарад* – раённая алімпіяда школьнікаў па прадметах і гарадская алімпіяда школьнікаў па прадметах – словаўтваральныя наватворы. Яны ўтвораны спосабам дэсуфіксацыі + субстантывацыі прыметнікаў *раённы* і *гарадскі*: *Вечар той нядзелі (леташняй, пасля раёну) быў невыносным: «Ну і што, хай і горшая, ну і што, галоўнае не перамога...»* [1]; *А толькі што.. скончыўся горад. Знаёмыя твары, знаёмы пах, але ўзровень іншы. Заданні кардынальна іншыя* [1].

Назоўнік *рэспубліка* – рэспубліканская алімпіяда школьнікаў па прадметах утвораны тым жа спосабам, што і папярэднія словы. Аднак у гэтым выпадку словаўтваральны ланцужок падаўжаецца: назоўнік *рэспубліка* сам становіцца ўтваральным і скарачаецца, утвараючы наватвор *рэспа* з тым жа значэннем: *На сурвэтцы, якую трэба спаліць і выпіць з шампанскім на НГ, я тады напісала: «хачу трапіць на рэспубліку».* Як і раней я прагнула ведаць. *Вось як закаханы, як вар’ят, ну проста хацела ШМАТ ведаць. Я марыла трапіць у каманду, якую рыхтуюць да рэспы* [1].

У слове *выпаадак* няма апіскі. Аўтар свядома ўстаўляе гук *а*, каб узмацніць сэнс слова, каб падкрэсліць няправільнае яго вымаўленне некаторымі выкладчыкамі ВНУ: *Выкладчыкі, канешне, былі некаторыя цудоўныя. Але галоўнай радасцю быў.. м... Рагойша. Гэта той **выпаадак**, калі словы непатрэбныя* [1].

Назоўнік *будзьмы* – пажаданні, тосты – утвораны спосабам субстантывацыі дыялектнай дзеяслоўнай формы: – *я навучылася выцінаць і рабіць такія штукі, як на паштоўках **будзьмы*** [1].

Звяртае ўвагу прыметнік *разумныя-прыразумныя*. Ён утвораны шляхам складання простаі формы прыметніка і яго формы найвышэйшай

ступені параўнання з парушэннем словаўтваральнай нормы: у беларускай літаратурнай мове форма найвышэйшай ступені параўнання прыметнікаў утвараецца далучэннем прыстаўкі *най-* да формы вышэйшай ступені. Прыстаўка *пры-* – элемент мовы-«трасянке» – адпавядае рускай прыстаўцы *пре-*, як і яна, далучаецца да простаі формы прыметніка *разумны*. Утварэнне такога прыметніка выклікаецца імкненнем аўтара да іроніі: *Вакол стаялі разумныя-прыразумныя дзеці, паўтаралі цытаты «Мой родны кут, як ты мне...», «Ты мне вясною прыснілася...», «Зорка Венера ўзышла...»* [1].

У прыметніку *мегаважны* надзвычай важны‘ першая частка *мега-* паходзіць ад грэч. *tegas* вялікі‘ і звычайна ўжываецца пры ўтварэнні навуковых тэрмінаў. У дадзеным выпадку кампанент *мега-* сінанімічны прыстаўцы *супер* вельмі; надзвычай; звыш», але вызначаецца большай экспрэсіяй і выразнасцю: *Нейкія цёткі спрабавалі выглядаць перад намі **мегаважнымі** асобамі, а атрымоўвалася, як у сельскай школе* [1].

Падобна да адзначанага прыметніка ўтворана і прыслоўе *мегапрыемна* вельмі прыемна‘: – *я пазнаёмілася з харошымі і страшэнна разумнымі дзецьмі. Не скажу, што мне заўжды **мегапрыемна** з імі сустракацца, але яны – новыя* [1].

Ад прыметнікаў, якія абазначаюць колер, у беларускай мове прыслоўі не ўтвараюцца, паколькі яны выражаюць пастаянную статычную прымету, уласціваю толькі прадмету: яна не здольна абазначаць якасць працэсу дзеяння. У тэксце «Авечкі» прыслоўе *ружова* набывае пераноснае значэнне прыемна‘, і таму яго ўжыванне не выклікае пярэчанняў: *У мяне тут пахне ружовымі цюльпанамі. Пахне-та ружова, а ўнутрах неяк жоўта* [1].

Аказіянальны дзеяслоў *забатаніць* старанна вывучаючы што-небудзь, пра многае даведацца; многае вывучыць‘ утвараецца трохсродкавым спосабам – прэфіксацыі + немарфемнага скарачэння + суфіксацыі (< *за-* + *батані/к/* + *-і(ць)*): «*Усё, – вырашыла Авечка, – вось зараз як возьмуся за падручнікі ды як **забатаню!**...*» [1]. У тэксце ўжыты і назоўнік з гэтым каранем – *батанства* веды, розум‘ (< *батан/ік/* + *-ств(а)*). Праўда, ён закрэслены і заменены сінонімам *розуму*: *І нешта ва ўсім гэтым «свяце батанства розуму» ўсё-такі было* [1].

Выклічнік *фублін* утвораны складаннем двух выклічнікаў – *фу* і *блін*: *Але ўжо з 1.09 пачалі ціснуць: «Ну што, сашенька, айда за первым дзіпломом?» **Фублін**. Больш за ўсё гэта не люблю* [1].

У сваім лісце аўтар звяртаецца і да абрэвіяцыі. Назвы пачатковых літар выкарыстаны ім, напрыклад, пры ўтварэнні абрэвіятуры *ВВ* вуснае выказванне на алімпіядзе па беларускай мове‘: *Вядома, разам са снегам раставала і жаданне наведваць гэтую 19-ю школу (дзе ўсё і адбывалася). Пра нейкую серабранскую я наогул маўчу (там было **ВВ**)* [1].

У тэксце, дзе гаворыцца пра ўмовы выкарыстання сайта «Модныя ножкі», побач з агульнаўжывальным назоўнікам *карыстальнік* выкарыстоўваецца аказіянальнае слова з суфіксам *-ач* – *карыстач*: *Як **карыстачу**, Вам дазваляецца праглядаць і загружаць матэрыялы сайта «Модныя ножкі»*

толькі для асабістага некамерцыйнага выкарыстання [2]. Там жа ўжыты складаныя назоўнікі, якія не фіксуюцца лексікаграфічнымі працамі, *вэб-СМІ* *электронныя сродкі масавай інфармацыі*, *гіперспасылка* *спасылка на электронны рэсурс*: *Дазваляецца цытаванне артыкулаў (без выкарыстання фатаграфій) у электронных вэб-СМІ без папярэдняга ўзгаднення з адміністрацыяй праекта «Модныя ножкі» пры ўмове размяшчэння сумесна з цытатай актыўнай прамой гіперспасылкі на цытаваны артыкул* [2].

Некаторыя аўтары адлюстроўваюць у сваіх матэрыялах словы, што з'явіліся для абазначэння культурных падзей. Выставу кніг, арганізаваную на плошчы горада Дуброўна, пад уплывам рускай анамастычнай традыцыі назвалі *Бібліяградам* *горадам кніг* (на ўзор былой назвы горада на Няве – *Ленінград*): *Працавалі работнікі Дубровенскай цэнтральнай бібліятэчнай сістэмы на славу. Цуд, які яны стварылі на плошчы ля фантана гараджанам, запомніцца надоўга і яны, дакладна, будуць спадзявацца, што гэтае падарожжа ў Бібліяград першае, але не апошняе* [3].

Для ўзмацнення выказвання аўтары вэб-тэкстаў звяртаюцца да рэдуплікацыі слоў (*любіць-любіць-любіць*): *Калі не здолець разам быць / Я б хацела навучыцца / Цябе любіць-любіць-любіць, / Сапраўдна, безадказна ліцца!* [4]. А вось прыслоўе *сапраўдна*, ужытае ў вершы, з'яўляецца недарэчным (у беларускай мове яму адпавядае прыслоўе *па-сапраўднаму*), ды і сам верш даволі недасканалы.

Наватворы, ужытыя некаторымі маладымі аўтарамі з парушэннямі словаўтваральных нормаў, выклікаюць толькі шкадаванне, выяўляючы бездапаможнасць гэтых асоб у пытаннях словаўтварэння, іх моўную глухату, эстэтычную слепату: *Аднойчы грып свіны па-свінску, / Такое мне падклаў свіннè, / Што насвінячыўся да віску, / Каб мне свіняча не было. / Але свінячая зараза / Свінячыць не перастае, / І рохкае ўсё раз за разам, / Нібы мяне свіннèй заве! / Ах, ты, свінячая істота, / Цябе я адсвінячу – ў раз! / І будзеш ты стаяць у плоце, / Між іншых вірусаў-зараз!* [5].

Нават больш вопытныя з аўтараў могуць дапускаць словаўтваральныя недарэчнасці. Так, у лісце «Пра журфак» ужываецца шасцікампанентнае прыслоўе: *Незалежная прэса на журфаку – гэта «Веснік МТЗ». Бо ад яе нічога не залежыць. Астатняе – фашыстоўска-семіцка-буржуазна-нацыяналістычна-прапагандыска-антыбеларуска зараджаная зараза* [6]. Аўтару, аднак, бракуе меры. Справа тут зусім не ў шматкампанентнасці, а ў тым, што апошні кампанент падобны да назоўніка і не «хоча» ні сэнсава, ні структурна спалучацца са словазлучэннем *зараджаная зараза*. У гэтым выпадку на месцы прыслоўя трэба было б ужыць прыметнік, а дзеепрыметнік *зараджаная* выключыць са сказа: ён лішні.

Такім чынам, у беларускамоўных вэб-тэкстах змяшчаецца разнастайны матэрыял па беларускім словаўтварэнні. Іх аўтары звяртаюцца да розных спосабаў словаўтварэння – часта ўжывальных (суфіксацыя, складанне, абрэвіяцыя і інш.) і рэдкіх (дэсуфіксацыя + субстантывацыя, немарфемнае скарачэнне і інш.). Адны аўтары бліскуча валодаюць словаўтваральнымі,

семантычнымі і стылістычнымі магчымасцямі мовы, другія жадаюць выказацца, але дапускаюць разнастайныя парушэнні норм беларускай мовы, у т. л. і словаўтваральныя, часам утвараючы словы-нязграбы. Такім аўтарам варта было б задумацца не толькі над тым, што сказаць, але і як сказаць. Культура маўлення адлюстроўвае ўнутраную культуру чалавека, з'яўляецца люстэркам яго душы, гармоніі яго думак і парыванняў. Не дарэчна ўтворанае слова калечыць тэкст, збядняе думку, выклікае недавер да аўтара.

#### СПІС ЛІТАРАТУРЫ

- 1 Avieczka. die Wiese. – Hoc erat in fatis / Avieczka // [Электронны рэсурс].– Рэжым доступу: <http://avieczka.livejournal.com/19102.html>. – Дата доступу: 16.10.2012.
- 2 Модныя ножкі. Умовы выкарыстання сайта // [Электронны рэсурс].– Рэжым доступу: <http://www.mod-no.ru/about/?l=bel&terms>. – Дата доступу: 17.10.2012.
- 3 Шавяка, В. Падарожжа ў Бібліяград / В. Шавяка // [Электронны рэсурс].– Рэжым доступу: <http://www.dubrovno.by/?p=15912.html>. – Дата доступу: 17.10.2012.
- 4 Ясючэня, А. Прызнанне / А. Ясючэня // [Электронны рэсурс].– Рэжым доступу: <http://yastasya.com/2010/09/stihi/belaruska>. – Дата доступу: 17.10.2012.
- 5 Янкоўскі, А. Вершы / А. Янкоўскі // [Электронны рэсурс].– Рэжым доступу: <http://www.vershy.ru/content>. – Дата доступу: 17.10.2012.
- 6 Uskoff, Jouri/ Пра журфак / Jouri\_Uskoff // [Электронны рэсурс].– Рэжым доступу: [http://jouri\\_uskoff.livejournal.com/44164.html](http://jouri_uskoff.livejournal.com/44164.html). – Дата доступу: 19.10.2012.

УДК 81 322:004.65

А. Ю. Станкевич (Гродно)

### ИСПОЛЬЗОВАНИЕ СТРОКОВЫХ ФУНКЦИЙ ПРИ РАБОТЕ С ЛИНГВИСТИЧЕСКОЙ КАРТОТЕКОЙ MS ACCESS

*Приводятся условия ориентированных на студентов-филологов задач применения встроенных и пользовательских строковых функций Ms Access. Даются комментированные решения таких задач.*

Строковые функции Ms Access являются эффективным инструментом обработки данных лингвистической картотеки. Ниже приводим условия задач на применение строковых функций и комментированные примеры решений. Для компактной записи решений даем SQL-код запроса; важно, что в режиме SQL аргументы многоместных функций отделяются друг от друга знаком «,», а в режиме конструктора – знаком «;».

*ЗАДАЧА № 1. Дано:* поле Voc таблицы DICTIONARY (тип данных: текстовый; значение этого поля для каждой записи – заголовочное слово словарной статьи; единый стиль оформления заголовочных слов не выдержан: часть вокабул набрана с первой прописной буквы, часть – полностью прописными буквами, часть – строчными буквами).

*Задание:* преобразовать записи в поле Voc так, чтобы все заголовочные слова были набраны прописными буквами (иначе говоря – были набраны в верхнем регистре).

*Решение-1:*

```
UPDATE DICTIONARY SET DICTIONARY.Voc = UCase([Voc]);
```

*Решение-2:*

```
UPDATE DICTIONARY SET DICTIONARY.Voc = StrConv([Voc],1);
```

*Комментарий.* В решении-1 используется строковая функция UCase, возвращающая заданную строку, преобразованную в верхний регистр. В решении-2 используется строковая функция StrConv, возвращающая заданную строку, преобразованную в соответствии со значением второго аргумента (здесь равен 1), в верхний регистр.

Вариацией задачи № 1 может стать задача на преобразование символов строки в нижний регистр через функцию LCase или функцию StrConv (со значением второго аргумента, равным 2).

*ЗАДАЧА № 2. Дано:* поле Author таблицы KWIC (тип данных: текстовый; значение этого поля для каждой записи – набранные через пробел имя, отчество и фамилия автора; в записях поля Author могут быть следующие ошибки: имя и/или отчество и/или фамилия автора набраны со строчной буквы).

*Задание:* преобразовать записи в поле Author так, чтобы имя, отчество и фамилия были набраны с первой прописной буквы.

*Решение-1:*

```
UPDATE KWIC SET KWIC.Author = StrConv(LCase([Author]),3);
```

*Решение-2:*

```
UPDATE KWIC SET KWIC.Author = StrConv(StrConv([Author],2),3);
```

*Комментарий.* Задача № 2 вводится для того, чтобы продемонстрировать специфику функции StrConv и (что более важно) – чтобы продемонстрировать прием использования в качестве аргумента некоторой функции значения, возвращаемого вложенной функцией.

*ЗАДАЧА № 3. Дано:* поле Word таблицы TAB\_TEXT (тип данных: текстовый; значение этого поля для каждой записи – словоформа некоторого текста, словоформы могут повторяться).

*Задание:* получить обратный словарь словоформ, исключив из него повторы.

*Решение:*

```
SELECT TAB_TEXT.Word  
FROM TAB_TEXT  
GROUP BY TAB_TEXT.Word, StrReverse([word])  
HAVING (((TAB_TEXT.Word) Is Not Null))  
ORDER BY StrReverse([word]);
```

*Комментарий.* Группировка исключает из выдачи повторяющиеся элементы. Обратный порядок достигается сортировкой по полю с реверсированными строками, возвращаемыми функцией StrReverse. Реверсированная

строка не показывается в выдаче. Поскольку возникает синтаксическая ошибка, если функция StrReverse получает как аргумент пустую строку, исключение пустых строк (фрагмент SQL-кода: Is Not Null) обязательно.

Одной из вариаций задачи № 3 может быть задача на получение обратного частотного словаря (допустим, поле Word содержит начальные формы множества слов, образованных суффиксальным способом, и исследуется словообразовательная активность суффиксов); решение:

```
SELECT TAB_TEXT.Word, Count(TAB_TEXT.Word) AS Freq
FROM TAB_TEXT
GROUP BY TAB_TEXT.Word, StrReverse([word])
HAVING (((TAB_TEXT.Word) Is Not Null))
ORDER BY StrReverse([word]), Count(TAB_TEXT.Word) DESC;
```

*ЗАДАЧА № 4. Дано:* поле SingPlur таблицы PARADYGM (тип данных: текстовый; каждая запись поля содержит две формы: форму единственного и форму множественного числа некоторой лексемы, формы разделены знаком «//»).

*Задание:* распределить содержимое поля SingPlur по двум полям: для каждой записи в поля Sing и Plur записать формы единственного и множественного числа соответственно.

*Решение:*

```
SELECT Paradygm.SingPlur, Left([SingPlur],InStr([SingPlur],"/")-1) AS Sing, Mid([SingPlur],InStr([SingPlur],"/")+2) AS Plur
FROM Paradygm;
```

*Комментарий.* В решении используются следующие строковые функции: а) функция Left, возвращающая определенное вторым аргументом количество знаков с левого края строки, указанной первым аргументом; б) функция InStr, возвращающая номер позиции первого вхождения подстроки, определенной вторым аргументом, в строку, определенную первым аргументом; в) функция Mid, возвращающая из строки, указанной первым аргументом, подстроку с позиции, указанной вторым аргументом. Используются вложенные функции.

*ЗАДАЧА № 5. Дано:* поле Voc таблицы DICTIONARY (тип данных: текстовый; значение этого поля для каждой записи – заголовочное слово словарной статьи; единый стиль оформления заголовочных слов не выдержан: в записи встречаются начальные пробелы).

*Задание:* исключить начальные пробелы в поле Voc.

*Решение:*

```
UPDATE DICTIONARY SET DICTIONARY.Voc = LTrim([Voc]);
```

*Комментарий.* В решении используется строковая функция LTrim, возвращающая копию указанной строки без начальных пробелов.

*ЗАДАЧА № 6. Дано:* код пользовательской функции DelSpace, удаляющей начальные, конечные и двойные пробелы (знак пробела обозначен так: Chr(32)):

```
Function DelSpace(str)
```

```
If Len(str) > 0 Then str = Trim(str)
Do While Len(str) > 0 And InStr(1, str, Chr(32) & Chr(32)) > 0
    str = Replace(str, Chr(32) & Chr(32), Chr(32))
```

```
Loop
DelSpace = str
End Function
```

SQL-код запроса с пользовательской функцией DelSpace, очищающей от некорректных пробелов поле Def таблицы DICTIONARY, таков:

```
UPDATE DICTIONARY SET DICTIONARY.Def = DelSpace([Def]);
```

*Задание:* пользуясь справкой Ms Access, найти описание встроенных функций Trim, Len, InStr, Replace, инструкций If ... Then и Do ... Loop и проанализировать синтаксис пользовательской функции.

*Комментарий.* На продвинутом этапе обучения можно дать основы создания пользовательских функций. В пользовательской функции DelSpace использованы следующие встроенные текстовые функции: а) функция Trim, возвращающая копию указанной строки без начальных и конечных пробелов; б) функция Len, возвращающая число знаков в строке; в) функция InStr (описана в комментарии к задаче № 4); г) функция Replace, возвращающая строку, где указанная вторым аргументом часть строки (здесь – два пробела) заменена строкой, указанной третьим аргументом (здесь – один пробел).

*ЗАДАЧА № 7. Дано:* код пользовательской функции DelSpace.

*Задание:* изменить функцию так, чтобы она возвращала строку, очищенную от знаков перевода строки и двойных пробелов; разделители слов обозначать так: знак пробела: Chr(32), знак перевода строки: Chr(10).

*Решение:*

```
Function ReplaceSymZ(str)
Do While Len(str) > 0 And InStr(1, str, Chr(10)) > 0
    str = Replace(str, Chr(10), Chr(32))

Loop
Do While Len(str) > 0 And InStr(1, str, Chr(32) & Chr(32)) > 0
    str = Replace(str, Chr(32) & Chr(32), Chr(32))

Loop
ReplaceSymZ = str
End Function
```

Хотя формулировка задач содержит имена только задействованных в решении полей, желательно, чтобы студенты работали с таблицами, поля которых соответствуют возможному множеству атрибутов некоторого лингвистического объекта. Так, таблица DICTIONARY, являющаяся упрощенным представлением структуры словарной статьи (представлением через одну таблицу, а не системой связанных таблиц), помимо описанных в условиях задач поля Voc и ключевого поля Code, может содержать следующие поля: для грамматических и стилистических помет: поля Gram и Styl (тип данных обоих полей: текстовый, желательно: задана подстановка),

для толкования и иллюстраций: поля Def и Ex (тип данных обоих полей: поле Memo). Таблица KWIC, помимо описанных в условиях задач поля Author и ключевого поля Code, может содержать следующие поля: для левого и правого контекстов: поля LContext и RContext (тип данных обоих полей: текстовый), для опорного слова: поле Pattern (тип данных: текстовый), для служебных реверсированных полей, используемых при обратной сортировке левого / правого контекстов и опорного слова – поля RevLContext, RevRContext, RevPattern соответственно (тип данных каждого из трех полей: текстовый); заметим, что сходную, но с большим числом полей таблицу студенты получают при экспорте в формат xls фрагмента выдачи поиска по НКРЯ [2].

Знание строковых функций, во-первых, сделает более эффективной работу студентов-филологов с лингвистической картотекой, во-вторых – подготовит студентов, обучающихся по направлению «Компьютерное обеспечение», к изучению аппарата строковых методов Python.

#### СПИСОК ЛИТЕРАТУРЫ

1 Ms Access: справка [Электронный ресурс] / Microsoft Corporation. – Режим доступа: <http://office.microsoft.com/ru-ru/access-help/FX010064691.aspx>. – Дата доступа: 20.09.2012.

2 Национальный корпус русского языка [Электронный ресурс] / НКРЯ. – Режим доступа: <http://www.ruscorpora.ru/>. – Дата доступа: 20.09.2012.

УДК 811.163.1'322:37.091.3:004.891

Т. А. Трохова, Е. И. Холявко (Гомель)

### **ЭКСПЕРТНАЯ СИСТЕМА ОБУЧЕНИЯ СТАРΟΣЛАВЯНСКОМУ ЯЗЫКУ**

*Предлагается аргументация необходимости разработки экспертной системы и использования ее в процессе обучения старославянскому языку. Рассматривается структура обучающей экспертной системы, ее основные функции и назначение, приводятся логические модели различных видов знаний. Выполнен анализ предметной области, выбраны методы заполнения базы знаний.*

Старославянский язык традиционно считается одним из труднейших предметов, изучаемых студентами первого курса, обучающимися по специальности «Русская филология». Опыт преподавания старославянского языка в высшей школе позволяет сделать вывод о том, что временные рамки курса не дают возможности уделить должного внимания системному пониманию курса, поэтому данный подход рассматривается параллельно с изучением основ языковой реконструкции.



Изучение столь важной и сложной филологической дисциплины совпадает с адаптационным периодом в студенческой жизни первокурсников. Как свидетельствуют данные опроса старшекурсников, среди факторов, препятствующих успешному овладению предметом, в первую очередь назывались повышенная тревожность, бытовая неустроенность, неумение рационально организовать самостоятельную работу над заданным материалом, недостаточность текущего контроля, отсутствие мотивации, а затем – отсутствие необходимой школьной языковой базы, сложность и теоретичность курса. Следует считаться также с сокращением количества учебных часов, общим требованием активизации самостоятельной работы студента, необходимостью ее организации и стимулирования.

Одним из возможных решений проблемы является выведение на качественно новый уровень возможностей студентов для самостоятельного приобретения знаний в области старославянского языка как системы. К сожалению, необходимые традиции самообразования не сложились. Но в то же время очевидны рост числа компьютеров в персональном пользовании и естественное стремление молодежи овладеть компьютерной техникой на профессиональном уровне. Это благоприятное обстоятельство позволяет ставить задачи повышения образовательного уровня посредством создания компьютерных учебных университетских курсов и экспертных обучающих систем. Следует отметить и такое положительное обстоятельство, что подобный подход к самообразованию не будет являться новым либо неожиданным для наших учащихся: в коммерческой продаже пользуется спросом множество мультимедийных программ изучения языковых курсов. Но в этом перечне практически отсутствуют мультимедийные программы, позволяющие приобрести знания и навыки освоения старославянского языка. Поэтому возможное решение проблемы видится в создании эффективной обучающей компьютерной системы, построенной по принципу экспертных систем и систем искусственного интеллекта.

Структура экспертной системы строится из следующих основных блоков: база знаний; машина вывода; интеллектуальный редактор базы знаний; интерфейс пользователя [1]. База знаний является основой любой экспертной системы, она содержит знания в конкретной предметной области, представленные в виде информационных компонентов различной логической структуры. Структура знаний носит двойственный характер: с одной стороны – знания должны предлагаться пользователю в удобном и понятном для него виде, а с другой стороны – логическая структура знаний во внутреннем представлении должна соответствовать требованиям машины вывода. Интеллектуальный блок экспертной системы – машина вывода – предназначен для построения цепочек выводов, моделирующих ход рассуждений экспертов на основании знаний, имеющихся в базе знаний экспертной системы по конкретному запросу пользователя. В состав машины вывода входит решатель и подсистема объяснений, которая позволяет пользователю проследить трассировку всего процесса получения ответа на тот или иной вопрос

с указанием использованных фрагментов базы знаний. Интеллектуальный редактор базы знаний представляет собой программный компонент, позволяющий эксперту и инженеру по знаниям заполнять базу знаний в интерактивном диалоговом режиме. Интерфейс пользователя – это программный компонент, позволяющий пользователю обращаться с запросами к экспертной системе и получать результаты поиска в удобной и понятной форме. Особое внимание следует уделить структуре представления знаний в экспертных системах. Согласно теории искусственного интеллекта знания делятся на фактографические и процедурные. Если для хранения фактографических знаний удобно применить фреймовые модели, а в случае преобладания в знаниях численных составляющих – модели баз данных, например, реляционные, то для хранения процедурных знаний применяются продукционные модели, строящиеся на основе правил productions «ЕСЛИ-ТО», модели на основе нейронных сетей или ориентированных гиперграфов. В одной базе знаний экспертной системы могут сочетаться различные модели представления знаний, машина вывода в этом случае должна учитывать особенности работы с каждой из логических структур знаний.

Основным этапом в разработке экспертной системы является квалифицированный анализ предметной области, в качестве которой для данной обучающей системы был выбран старославянский язык.

Курс старославянского языка открывает цикл историко-лингвистических дисциплин, закладывающих основы фундаментальной лингвистической подготовки филолога-русиста, именно здесь осваиваются теоретические принципы и практические навыки сравнительно-исторического анализа. С этой учебной дисциплиной тесно связаны курсы исторической грамматики русского языка, истории русского литературного языка, а также сравнительной грамматики славянских языков, изучение которых невозможно без основательного знания курса старославянского языка – своего рода лингвистического введения в изучение истории русского и других славянских языков. Курс старославянского языка сохраняет свою самостоятельность и имеет особое содержание, которое не может быть передано другими курсами. Без знания старославянского языка, его культурной, общественной функции, без знания истории русского литературного языка как сложного взаимодействия языка церковнославянского, генетически связанного со старославянским, и живого народного русского языка невозможно понять всю сложность современного русского литературного языка, его семантической структуры, лексико-грамматической системы, стилистической организации.

Курс старославянского языка содержит сведения о древнем книжно-письменном языке славян, его системе, его культурной роли в жизни славян, о его функции общеславянского литературного языка в древний период и о его значении в истории литературных языков южных и восточных славян в более позднюю историческую эпоху. Поэтому изучение старославянского языка имеет не только лингвистическое, но и общепилологическое значение, являясь своеобразным введением в славянскую филологию.

Изучение старославянского языка обеспечивает глубокое понимание языковых закономерностей, которые определяют структуру и функционирование языковой системы в прошлом и настоящем, позволяет ориентироваться в ее изменениях на современном этапе и в определенной степени прогнозировать направления ее развития в будущем. Все это означает научное овладение языком, что составляет одну из главных задач обучения студентов специальности «Русская филология».

Разделы базы знаний построены в соответствии с традиционными частями курса: введение, фонетика и фонология, морфология, синтаксис и лексика. Для каждого раздела учтены особенности представления знаний. Если во введении преобладают фактографические знания, то в последующих разделах актуализируется проблема сохранения взаимосвязи языковых элементов на разных хронологических срезах. В качестве логической структуры в этом случае выбраны продукционные модели знаний.

Создание обучающей экспертной системы должно базироваться на основных логико-лингвистических моделях обучения. В зависимости от требований к обучающей системе моделями обучения могут служить обучение путем самостоятельного открытия, обучение при наблюдении примеров, обучение путем объяснений. Разрабатываемая экспертная система совмещает в себе все три модели обучения.

При работе с обучающей экспертной системой студенту предоставляется возможность изучить теоретический материал по классифицированным разделам курса, рассмотреть частные случаи воплощенных языковых закономерностей, получить разъяснение базовых терминов и комментариев типовых фактов, который можно использовать как модель для анализа аналогичных ситуаций. В процессе диалога с экспертной системой студенту предлагается ряд вопросов, для ответа на которые требуется самостоятельное воссоздание логических связей и формулирование выводов. Выполнение этих заданий идет в пошаговом режиме, который позволяет при необходимости получать направляющую помощь. Несмотря на то что классическая экспертная система не содержит функции контроля знаний, в обучающую систему может быть включен набор тестов по отдельным разделам и курсу в целом.

Экспертная система обучения старославянскому языку призвана не только представить студенту исчерпывающий теоретический материал по данному курсу, но и с помощью удобного интерфейса путем ведения методически продуманного диалога направлять формирование системного понимания взаимосвязанных языковых фактов, подсказывая в нужные моменты типовые варианты и модели уже существующих решений.

Предлагаемая экспертная система вполне вписывается в современную парадигму университетского образования. Ее внедрение позволит активизировать регулярную учебную деятельность студента, будет способствовать получению системных качественных знаний, своевременной коррекции методики преподавания и усилит мотивацию обучения.

## СПИСОК ЛИТЕРАТУРЫ

1 Джексон, П. Введение в экспертные системы / П. Джексон. – 3-е изд. – М.: Издательский дом «Вильямс», 2001. – 624 с.

УДК 37.091:81:005.935.33

К. Л. Хазанава (Гомель)

### **АБ ВЫКАРЫСТАННІ ЭЛЕКТРОННАГА ТЭСЦІРАВАННЯ ПРЫ ВЫКЛАДАННІ ЛІНГВІСТЫЧНЫХ ДЫСЦЫПЛІН**

*Вьяўляецца значэнне электроннага тэсціравання пры выкладанні лінгвістычных дысцыплін, вызначаюцца метады і спосабы стварэння электронных тэстаў, асаблівасці падрыхтоўкі тэставых заданняў. Абгрунтаўваецца неабходнасць вывучэння метадыкі стварэння электронных тэстаў для студэнтаў-філолагаў.*

На памежжы 20–21 стагоддзяў камп'ютар, ці, дакладней, персанальны камп'ютар (знакаміты РС – *personal computer*) трывала замацаваўся ў чалавечым жыцці і заканамерна трапіў у навучальны працэс. У сучаснай адукацыі разнастайныя камп'ютарныя і інфармацыйныя тэхналогіі сталі сродкам інтэнсіфікацыі і паляпшэння вучэбнага працэсу. Камп'ютар як найсучаснейшая прылада чалавечай дзейнасці, трапляючы ў навучанне, якасна змяняе і павялічвае магчымасці захавання і прымянення ведаў чалавекам.

Сваё пачэснае месца маюць найноўшыя тэхналогіі і пры выкладанні лінгвістычных дысцыплін. Прыклад гэтага – электронныя падручнікі па разнастайных галінах мовазнаўства, па гісторыі ўсходнеславянскіх моў, дзе ўтрымліваюцца тэкст, малюнкi, фотаздымкі, маецца пэўнае гукавое суправаджэнне (што асабліва важна для засваення фанетыкі і арфаэпіі), нават асобныя відэаматэрыялы, а таксама тэставыя заданні, інтэрнэт-спасылкі на навуковую літаратуру або старадрукі.

Развіццё камп'ютарнай тэхнікі і інфармацыйных тэхналогій у сучасны момант адбываецца настолькі інтэнсіўна, што ўжо зараз для стварэння электронных тэставых матэрыялаў выкладчыку-філолагу не трэба шукаць пэўных метадычных дапаможнікаў і падрыхтаваных электронных варыянтаў распрацовак тэстаў. Намнога прасцей і хутчэй стварыць электронныя тэсты самастойна паводле гатовых і наяўных у значнай колькасці праграм для стварэння электронных тэстаў. Такія тэсты будуць абапірацца на вучэбны тэарэтычны і фактычны матэрыял, неабходны для выкладчыка ў дадзены момант вучэбнага працэсу.

Менавіта для набыцця падобных навыкаў (у ліку многіх іншых) у Рэспубліцы Беларусь быў уведзены ў практыку навучання напрамак – «Камп'ютарная філалогія». Падрыхтоўка спецыялістаў па камп'ютарнай

лінгвістыцы вядзецца з 1996 года у межах адпаведнай спецыялізацыі ў БДУ, МДЛУ, ГрДУ і адпаведнага напрамку ў ГДУ імя Ф. Скарыны.

Увогуле тэсціраванне – гэта адзін з раздзелаў дыягностыкі. Тэсціраванне прымяняецца ў тэхніцы, медыцыне, псіхіятрыі, адукацыі для вызначэння прыдатнасці аб'екта тэсціравання для выканання тых ці іншых функцый. Якасць тэсціравання і сапраўднасць яго вынікаў у значнай ступені залежыць ад тэстара.

Паняцце *тэст* звычайна разумеецца як стандартызаваны, кароткі, абмежаваны ў часе іспыт, прызначаны для ўстанаўлення колькасных і якасных індывідуальных адрозненняў. *Педагагічны тэст* – гэта сістэма заданняў узрастаючай цяжкасці, спецыфічнай формы, якія дазваляюць якасна і эфектыўна вымяраць узровень і ацаніць структуру падрыхтаванасці навучэнцаў.

Шырокае прымяненне тэсціравання ў адукацыі пачалося ў ЗША яшчэ ў канцы 19 ст. Нягледзячы на распрацоўку расійскімі псіхолагамі ў першай трэці 20 ст. педагагічнай тэсталогіі, у 30-я гады навуку аб тэстах аб'явілі буржуазнай, яе мэты – рэакцыйнымі, а самі тэсты – педалагічнымі скажэннямі. Актыўнае ўвядзенне тэстаў у адукацыйны працэс краін постсавецкай прасторы пачалося толькі на мяжы 20–21 стагоддзяў.

Асноўнай перавагай тэсціравання як больш якаснага спосабу вымярэння ўзроўню ведаў, з чым звычайна асацыіруецца выніковы тэставы бал, з'яўляецца аб'ектывізацыя ацэньвання, што дасягаецца праз стандартызаваную працэдуру правядзення, праверкі паказчыкаў якасці заданняў і тэстаў цалкам. Тэсціраванне прапануе роўныя ўмовы для ўсіх іспытантаў, выключаючы як вядомы «чалавечы фактар», суб'ектывізм выкладчыка, так і «нчаслівы (або наадварот, калі не «нашанцавала» з экзаменацыйным білетам) выпадак», элемент выпадковасці пры выцягванні білета. Разам з тым тэст больш аб'ёмны за традыцыйны экзамен інструмент, бо тэсціраванне можа ўключаць у сябе заданні і па некаторых раздзелах, і па ўсіх тэмах курсу, і таму дазваляе ўстанавіць узровень ведаў і па прадмеце ў цэлым, і па асобных яго раздзелах.

Важнай перавагай тэсціравання з'яўляецца яго эканамічнасць, якая вынікае з адзінаразовай распрацоўкі інструментарыя тэсціравання і наступнага шматразовага яго выкарыстання. Акрамя таго, правядзенне тэсціравання для сярэднестатыстычнай групы з 25–30 чалавек займае 1,5–2 гадзіны, а вусны экзамен – больш за 4 гадзіны. Яшчэ больш інтэнсіўным па часавых і матэрыяльных затратах з'яўляецца камп'ютарнае тэсціраванне.

Аднак нельга не адзначыць, што стварэнне якасных тэставых заданняў – доўгі і працаёмкі працэс. Распрацоўка тэставых заданняў для правядзення камп'ютарнага, або электроннага, тэсціравання патрабуе складанай папярэдняй падрыхтоўкі.

Матэрыялы кантрольна-вымяральных падсістэм электронных сродкаў навучання для правядзення тэсціравання распрацоўваюцца ў адпаведнасці з патрабаваннямі да ведаў, уменняў і навыкаў іспытантаў. Пасля распрацоўкі тэхналогій тэсціравання праводзіцца экспертная ацэнка

якасці кантрольна-вымяральных матэрыялаў і праверка іх адпаведнасці зместу навучання, а таксама ацэнка паўнаты пакрыцця патрабаванняў вымяральнымі матэрыяламі. У працэсе правядзення тэсціравання з выкарыстаннем распрацаваных электронных сродкаў навучання адбываецца вызначэнне і шкаліраванне вынікаў вымярэння, прывядзенне іх да адзінай сістэмы ацэньвання, параўнанне вынікаў, фармуляванне вывадаў па якасці навучання з выкарыстаннем электронных сродкаў навучання.

У навуковай літаратуры, прысвечанай праблемам падрыхтоўкі электроннага тэсціравання, вылучаецца велізарная колькасць патрабаванняў для тэставых заданняў. Асноўнымі з іх з'яўляюцца:

- зразумелая мова заданняў і варыянтаў адказаў;
- нешматлікасць слоў у заданні (каля 10 слоў);
- нешматлікасць варыянтаў адказаў (да 5–6);
- адсутнасць заданняў з адназначным адказам (так, не);
- значная колькасць заданняў (колькасць заданняў і надзейнасць тэксту знаходзяцца ў прама прапарцыянальнай залежнасці);
- адсутнасць правакацыйных пытанняў;
- рознасць месца правільнага адказу ў кожным заданні;
- адсутнасць пытанняў, якія даюць перавагі іспытантам пэўнага полу, нацыянальнасці, расы.

Пасля складання тэставых заданняў, выкладчык паўстае перад выбарам праграмы, якая дазволіць правесці камп'ютарнае тэсціраванне. Перш за ўсё для стварэння тэсту можна выкарыстаць офісныя дадаткі. Напрыклад, MS Power Point прапануе магчымасці для стварэння тэстаў з адным правільным варыянтам адказаў і з некалькімі правільнымі адказамі і ўводам адказаў. Недахопам тэсту, створанага ў падобным фармаце, з'яўляецца тое, што іспытант не бачыць, на якія пытанні даў памылковыя адказы, што абавязкова патрабуе пасля правядзення такога тэсціравання яго падрабязнае абмеркаванне і аналіз.

Электронныя тэсты магчыма ствараць у электронных табліцах. Такі спосаб маецца ў выкарыстанні MS Excell.

Нават агульна вядомая, даступная, шырока прызначаная і распаўсюджаная MS Word таксама ўтрымлівае патэнцыял для стварэння электронных тэстаў пры звароце для панелі інструментаў – «формы».

І натуральна, падрыхтоўка тэставага інструментарыя наўрад ці можа абысціся без знакамітага сусветнага павуціння. Сучасная інтэрнэт-прастора прапануе разнастайныя віды тэставых канструктараў.

Генератар тэстаў стварае тэсты ў фармаце HTML [1]. Можна самім задаваць колькасць пытанняў, варыянтаў адказаў. Улічваецца і складанасць пытанняў. Тэсты ствараюцца MS Explorer 6.0 і вышэй, а затым працуюць пад кіраўніцтвам любога браўзера, які падтрымлівае JavaScript.

Цікавай і своеасаблівай электроннай абалонкай для будовы тэстаў з'яўляецца праграма *Hot Potatoes* «гарачая бульба» [2]. Праграма распрацавана ў 1997–2003 гадах у Цэнтры інфармацыйных тэхналогій

у гуманітарнай адукацыі Універсітэта Вікторыі (Канада). Аўтары праграмы – Сцюарт Арнэйл (Stewart Arneil), Марцін Холмс (Martin Holmes), Хілары Стрыт (Hilary Street).

–“Гарачая бульба” – гэта інструментальная праграма-абалонка, якая дае магчымасць ствараць інтэрактыўныя заданні і тэсты для кантролю і самакантролю навучэнцаў. З дапамогай праграмы можна стварыць 10 тыпаў практыкаванняў і тэстаў па разнастайных дысцыплінах з выкарыстаннем тэкставай, графічнай, аўдыа- і відэаінфармацыі. Заданні захоўваюцца ў фармаце вэб-старонкі, і для іх выкарыстання неабходны вэб-браўзар.

У склад праграмы ўваходзяць 5 блокаў праграм для складання заданняў і тэстаў розных відаў. Кожны з блокаў можа выкарыстоўвацца як самастойная праграма: *JQuiz* – віктарына – пытанні з множным выбарам адказу; *JCloze* – запаўненне пропускаў; *JMatch* – устанаўленне адпаведнасцей; *JMix* – узнаўленне паслядоўнасці; *JCross* – крыжаванка.

Такім чынам, праграма прапануе магчымасці для творчасці тэставых заданняў і закрытага, калі кожнае пытанне суправаджаецца гатовымі варыянтамі адказаў, з якіх неабходна выбраць адзін ці некалькі правільных, і адкрытага тыпу, калі на кожнае пытанне іспытант павінен прапанаваць свой адказ: дапісаць слова, словазлучэнне, сказ, знак, формулу.



Такім чынам, правядзенне камп'ютарнага тэсціравання адкрывае перад выкладчыкам новыя магчымасці і творчыя шляхі для педагагічнай працы. Такі метады навучання ў арганічным сінтэзе з іншымі традыцыйнымі і, магчыма, больш звычнымі, спрыяе павышэнню якасці адукацыйнага працэсу. Правядзенне камп'ютарнага тэсціравання, акрамя таго, з'яўляецца больш –“мяккім” інструментам у псіхалагічных адносінах, бо ствараюцца роўныя ўмовы для ўсіх іспытанцаў, выкарыстоўваецца адзіная працэдура і адзіныя крытэрыі ацэнкі, што прыводзіць да зніжэння перадэкзаменацыйных нервовых напружанняў. У выніку іспытант падчас экзамену ў выглядзе камп'ютарнага тэсціравання трапляе ў сітуацыю зносін сам на сам з камп'ютарам, што для сучаснага маладога пакалення (школьнікі, навучэнцы, студэнты) стала прывычным, знаёмым, а для часткі – нават улюбёным спосабам правядзення часу.

## СПИС ЛИТАРАТУРЫ

- 1 [Электронны рэсурс]. – Рэжым доступу: [www.testobr.narod.ru/doc/generator.zip](http://www.testobr.narod.ru/doc/generator.zip)
- 2 [Электронны рэсурс]. – Рэжым доступу: <http://web.uvic.ca/hrd/hotpot>

УДК 81'32

Б. Э. Хакимов (Казань)

### **ПРОБЛЕМЫ ГРАММАТИЧЕСКОЙ АННОТАЦИИ В ДИАЛЕКТНЫХ КОРПУСАХ ТЕКСТОВ (НА ПРИМЕРЕ МИШАРСКОГО ДИАЛЕКТА ТАТАРСКОГО ЯЗЫКА)<sup>5</sup>**

*В статье рассматриваются концептуальные проблемы разработки моделей и методов корпусно-ориентированной грамматической аннотации текстов на мишарском диалекте татарского языка в контексте создания универсальных многоязычных лингвистических моделей. Исследуется характер грамматической вариативности в мишарском диалекте, оцениваются потенциальные возможности автоматизации корпусной разметки.*

Корпусные исследования в тюркских языках начались сравнительно недавно, о чем свидетельствует небольшое количество по-настоящему репрезентативных и аннотированных корпусов текстов на тюркских языках. В то же время, с начала 90-х годов XX века ведутся различные исследования татарского языка в компьютерном аспекте. Разработаны концепции машинного фонда татарского языка и внедрения татарского языка в компьютерные технологии [2, 3]. В настоящее время идея машинного фонда татарского языка трансформируется и конкретизируется в направлении концепции национального корпуса татарского языка, который должен включать и диалектный материал.

Мишарский (западный) диалект татарского языка, распространенный не только в Республике Татарстан, но и за ее пределами, представлен системой многочисленных говоров, в ряде случаев существенно отличающихся как от литературного татарского языка, так и друг от друга [5]. Анализ и классификация этих различий позволяет найти пути формализованного представления морфологических характеристик произвольного текста на мишарском диалекте в электронном машиночитаемом формате.

Проблемы подготовки корпусов национальных языков и диалектов так или иначе группируются вокруг следующих центральных задач:

- задача создания коллекции текстов;
- задача обеспечения репрезентативности и сбалансированности корпуса;
- задача автоматизации процесса аннотации.

---

<sup>5</sup> Работа выполнена при поддержке РГНФ (грант 12-14-16035).



Не останавливаясь подробно на первых двух проблемах, очевидно имеющих свою специфику при работе с диалектным материалом, обратимся непосредственно к концептуальным вопросам разработки методов и систем автоматизированной обработки языковых данных для систем корпусных лингвистических исследований. В данном аспекте, помимо собственно диалектологической метаразметки, большой интерес представляет проблема грамматической разметки диалектных текстов, в которых лексическая, морфологическая и синтаксическая вариативность ставит под сомнение саму идею единой системы формализованной аннотации для конкретного диалекта, ведь таковая подразумевает наличие определенного стандарта. Анализ данной ситуации для мишарского диалекта татарского языка позволяет сделать некоторые выводы.

Разработанные к настоящему моменту морфологические модели татарского языка уже задают определенные стандарты представления татарской агглютинативной морфологии [1, 6]. В частности двухуровневая модель татарской морфологии Д. Ш. Сулейманова, Р. А. Гильмуллина и др. реализована в виде базы фонологических и морфотактических правил, построенной на основе глагольных и номинативных парадигм, и определяющей взаимосвязи между основой и аффиксальными группами языка. Татарский язык имеет две основные словоизменительные парадигмы, описывающие порядок присоединения к основе аффиксальных классов в определенной последовательности [1].

Принадлежность к агглютинативному типу определяет особенности морфологического строения слова в татарском языке: аффиксы последовательно присоединяются к корню, каждое грамматическое значение, как правило, выражается отдельным аффиксом, аффиксы однозначны и регулярны, т. е. способны присоединяться ко всем словам определенной части речи [7]. В татарском языке морфологические характеристики легко распознаются при автоматическом анализе аффиксального состава словоформы.

Таким образом, ориентация на аффиксальную структуру слова является отличительной особенностью средств автоматической обработки текстов на тюркских языках. Не менее важна полнота описания с представлением всех возможных альтернативных разборов и разумный баланс между объективной реальностью языка и традициями грамматической теории. Наряду с вариативностью грамматической структуры в тюркских языках и диалектах четко прослеживается и единая система, что делает в определенной степени возможным универсальное формализованное описание и упрощает создание многоязычных ресурсов и сравнительное исследование родственных языков. Соответственно, исследование диалектов в подобном аспекте приобретает дополнительную актуальность.

Для мишарского диалекта татарского языка существуют различные подходы к описанию грамматических явлений и представлению речи в текстовой форме. С одной стороны, это традиции, сложившиеся в татарской диалектологии за последние десятилетия [5]. Согласно этим традициям,

тексты записываются в транскрипции на основе стандартной татарской орфографии с добавлением некоторых специальных символов. Грамматическая структура мишарского диалекта описывается в трудах по татарской диалектологии в соответствии со сложившимися в татарской лингвистике подходами и, как правило, в сравнении с литературным языком. С другой стороны, в ряде последних фундаментальных исследований используется типологический подход к описанию и объяснению языковых явлений, универсальная общетипологическая терминология и латинизированная транскрипция на основе разработанной Международной фонетической ассоциацией универсальной системе символов [4]. Учитывая существенные и порой принципиальные различия данных, целесообразным представляется параллельное использование обеих систем и реализация возможности такого параллельного использования.

Далее рассмотрим потенциальные возможности автоматизации процесса грамматической аннотации для мишарского диалекта татарского языка. Центральным элементом системы автоматизированной грамматической аннотации является морфологический анализатор (парсер). Разработанный для литературного татарского языка морфологический анализатор содержит словарь основ и словарь аффиксов. В словаре основ все лексемы подразделяются на 4 морфологических типа. В свою очередь, все морфологические типы подразделяются на морфонологические типы в зависимости от того, какие алломорфы аффиксальных морфем присоединяются к данной основе [6].

Для применения подобного анализатора по отношению к диалектным текстам необходима его адаптация. Принципиальная возможность такой адаптации вполне обоснованно может быть поставлена под сомнение. Однако мишарский диалект имеет практически идентичную с татарским литературным языком систему морфологии, различия в основном имеют фонологическую природу либо представлены некоторыми специфичными грамматическими формами. Таким образом, адаптация морфоанализатора может быть достигнута путем ввода специфичных фонологических правил и морфонологических типов, а также составления диалектологического словаря основ. Морфотактические правила и модели словоизменительных парадигм в таком случае не подлежат серьезной корректировке, но потребуются изменения в наборах аффиксальных морфем.

Так как на уровне отдельных говоров диалектная вариативность представлена в меньшей степени, наиболее перспективной в этом направлении является разработка частных формально-грамматических моделей говоров с потенциальной возможностью дальнейшего обобщения на уровне целого диалекта. Подобные исследования представляют интерес не только непосредственно для татарской диалектологии, но и дают материал для дальнейших исследований в области построения многоязычных лингвистических моделей.

## СПИСОК ЛИТЕРАТУРЫ

1 Suleymanov, D. Sh., Guilmouline R. A., Guilmouline A. A. Tatar phonological rules as a base of two-level morphological analyzer, in Proceedings of LP'2000, ed. V. Palek and O. Fujimura: 495–504 p., The Karolinum Press, Prague.

2 Бухараев, Р. Г. К концепции внедрения татарского языка в компьютерные технологии / Р. Г. Бухарев, Д. Ш. Сулейманов // В сб. "Интеллект. Язык. Компьютер", вып. 2. Казань: Изд-во КГУ, 1995. – С. 8–19.

3 Бухараев, Р. Г. К концепции Машинного Фонда Республики Татарстан / Р. Г. Бухарев, Ф. С. Сафиуллина, Д. Ш. Сулейманов и др. // Татарский язык и новые информационные технологии. Серия: Интеллект. Язык. Компьютер. – Вып. 2. – Казань: Изд-во Казан. ун-та, 1995. – С. 20–35.

4 Мишарский диалект татарского языка: очерки по синтаксису и семантике / под ред. Е. А. Лютиковой, К. И. Казенина, В. Д. Соловьева, С. Г. Татевосова. – Казань: Маагриф, 2007. – 383 с.

5 Рамазанова, Д. Б. Вопросы татарской диалектологии / Д. Б. Рамазанова. – Казань, 2008. – 272 с.

6 Сулейманов, Д. Ш. Структурно-функциональная компьютерная модель татарских морфем / Д. Ш. Сулейманов, А. Р. Гатиатуллин. – Казань: ФЭН, 2003. – 220 с.

7 Татарская грамматика. Т. II. Морфология / Ред. М. З. Закиев, Ф. А. Ганиев, К. З. Зиннатуллина. – Казань: Татарское кн. изд-во, 1993. – 397 с.

УДК 81.161.1'271.12

Н. И. Шабулдаева (Гомель)

## ЯЗЫКОВАЯ НОРМА И НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА

*В статье рассматриваются материалы Национального корпуса русского языка при исследовании тенденций развития литературной нормы русского языка на примере имен существительных анемон/анемона, имеющих колебания в грамматическом роде.*

Языковая норма – явление историческое. Это давно стало аксиомой. Нормы изменяются из-за внутренних языковых процессов и экстралингвистических факторов. Определённое языковое явление носителями данного языка принимается как литературное (чаще всего под влиянием наиболее образованной части общества, а также средств массовой информации) и закрепляется в различных специальных словарях и справочниках. В свою очередь носители языка обращаются к этим справочникам и уточняют, правильно ли они используют то или иное языковое явление. Авторы справочников пишут, например: «Такое-то слово (или такая-то его форма) более употребительно». Но возникает вполне оправданный вопрос: «Каким образом автор этого источника узнал, что это действительно так?»

Мы поставили перед собой цель с помощью Национального корпуса русского языка выяснить на примере одного слова, можно ли всецело положиться на данные справочников по языковым нормам.

В качестве примера нами взято слово, имеющее колебание в роде, – *анемон – анемона*.

Д. Э. Розенталь в «Справочнике по правописанию и литературной правке», изданном под редакцией И. Б. Голуб в Москве в издательстве «Айрис-пресс» в 2004 году, пишет: «*анемон – анемона* (более распространена вторая форма; так, в «Словаре иностранных слов» первый вариант вообще не приводится)» [1, с. 193]. Паспортные данные названного словаря не названы.

В «Справочнике по русскому языку: Правописание. Произношение. Литературное редактирование», написанном в соавторстве Д. Э. Розенталем, Е. В. Джанджаковой и Н. П. Кабановой, изданном в Москве в издательстве «Айрис-пресс» в 2005 году (указано, что текст печатается в новой редакции по изданию «Справочник по правописанию, произношению и литературному произношению» 4-е изд., испр. – М.: «ЧеРо», 2001), указано по поводу этой пары следующее: «более употребительна первая форма» [2, с. 356].

«Словарь трудностей русского языка» Д. Э. Розенталя и М. А. Теленковой за 1985 год на первом месте ставит форму *анемон* [3, с. 22].

Энциклопедическое издание «Природа Белоруссии. Популярная энциклопедия», выпущенная в Минске в 1989 году издательством «Белорусская советская энциклопедия» имени П. Бровки» приводит форму *анемона* «ветреница» [4, с. 260].

Толковые словари последних лет также придерживаются последовательности *анемон – анемона*, из чего мы можем понять, что форме мужского рода отдаётся предпочтение: *Анемон* и *Анемона*. М., ж. 1. Травянистое растение семейства лютиковых с цветками различной окраски; ветреница ([dic.academic.ru](http://dic.academic.ru) > Толковый словарь Ефремовой); *АНЕМО́Н* -а; м.; *АНЕМО́НА*, -ы; ж. [греч. *anemos* – ветер]. «То же». ([dic.academic.ru](http://dic.academic.ru) > Большой толковый словарь русского языка. – 1-е изд-е: СПб.: Норин) [5]. К месту сказать, что вышеуказанные словари приводят не все реально существующие значения этого слова. Кроме того, что это травянистое растение семейства лютиковых, это также ядовитая актиния, или анемон, морское метровое животное-растение. (Леонид Саксон. Принц Уэльский // «Октябрь», 2001)<sup>6</sup>.

Для того чтобы ответить на вопрос, какую же форму этого слова предпочитают носители русского языка, мы обратились к данным Национального корпуса русского языка. В результате введения этих форм в поисковую систему выданы следующие сведения: *анемон* – 56 документов, 89 вхождений; *анемона* – 53 документа, 79 вхождений. К сожалению, и в том и другом случае НКРЯ выдаёт формы мужского, женского рода и множественного числа, не дифференцируя их. Это затрудняет решение поставленной задачи. В результате отбора форм вручную было установлено, что форма мужского рода *анемон* употреблена 15 авторами, форма женского рода *анемона* – 13 авторами. Значительно чаще это слово

<sup>6</sup> Данные Национального корпуса русского языка.

используется в форме множественного числа. Нами эти формы в расчёт не берутся, кроме форм родительного падежа множественного числа, по которым можно определить, какой род этого слова хочет использовать автор: *А на склонах сопок – целые колонии луговой герани, бледно-розовых анемон* (жен. род), *студено-синих колокольчиков...* (Вадим Бурлак. Хранители древних тайн (2001)) – *Ковер из диких анемонов* (муж. род) *затягивал подножие сопок, но за рекой все как-то резко и грозно менялось.* (Леонид Юзефович. Князь ветра (2001)).

По данным НКРЯ впервые это слово дважды употребил Ф. И. Буслаев в труде «Об эпических выражениях украинской поэзии» в 1850 году: *Эта могильная сон-трава, открывающая человеку во сне тайны, принадлежит к роду Анемона (Anemone patens, pulsatilla).* (Ф. И. Буслаев. Об эпических выражениях украинской поэзии (1850)). Как видим, первой была форма мужского рода.

Обе формы за 159 лет (с 1850 по 2009 года) употреблены следующее количество раз: мужской род – 24 раза, женский род – 22 раза. Как видим, они почти совпадают.

Каким же образом эти формы распределялись по годам? Следующим это слово как существительное женского рода трижды использовал Николай II: *Во время прогулки набрали целую корзину анемон.* (Николай II. Дневники 1894–1896 (1894–1896)). Почти через пятнадцать лет И.И. Мечников три раза использовал также форму женского рода: *Наблюдали морскую анемону, относящуюся к виду Actinia mesembryanthemum, которая прожила 66 лет* (И. И. Мечников. Этюды оптимизма (1907–1913)). В 20–40-е годы XX века М. М. Пришвин, Б. Л. Пастернак и К. Г. Паустовский в своих произведениях воспринимают это слово также как существительное женского рода: *Я вернулся домой не с убитыми птицами, а с букетом анемон, волчьего лыка, тончайших веточек с золотистыми сережками.* (М. М. Пришвин. Дневники (1926)); *На дне сухого подполья разрывчато, как солнце, горели четыре реччатые молнии, и, соперничая с лампами, безумствовали в огромных лоханях, отобранные по колерам и породам, жаркие снопы пионов, желтых ромашек, тюльпанов и анемон.* (Б. Л. Пастернак. Охранная грамота (1930)); *Это был неизвестно откуда поавший в наш сад и расцветший второй раз цветок лесной анемоны.* (К. Г. Паустовский. Книга о жизни. Далекие годы (1946)).

В эти же годы В. П. Катаев и М. А. Осоргин приводят его в форме мужского рода: *Я стоял в джунглях папоротников, цитрусов, пальм, анемонов, парниковой сирени* (В. П. Катаев. Цветы (1936)); *Мне было жалко, что белые весенние цветы в засушенном виде всегда желтеют: ...лесной анемон, прелестный сибирский княжик...* (М. А. Осоргин. Времена (1942)).

В дальнейшем, с 1950 по 1995 годы, в различных изданиях приводится форма мужского рода *анемон*. Выявлено восемь источников, в которых слово *анемон* используется по одному разу. Среди них обнаружены работы разных жанров:

1) художественные произведения: а) поэзия: *С долин Даурии гористой Возьму роскошный Анемон, Статис роскошный и душистый И снежной белизны Пион.* (Л. К. Чуковская. Декабристы (1950–1951)); б) проза: *Ученый почувствовал себя в аквариуме, и ощущение, что он сам – морской анемон, сознание не смогло опровергнуть.* (И. Губерман. Борьба с безумием // «Химия и жизнь», 1968); *Стояла неважная для рыбалки погода – тихая, яркая и засушливая, но проселок был еще по-весеннему плотным и легким, и на опушках сосновых подлесков то и дело попадались колонии анемонов.* (Константин Воробьев. Вот пришел великан (1971)); *Он слушал и писал свои гармонизации, лучшие из которых он посвящал мечте своей молодости – девушке именем Ми-Бемоль, летавшей в хороводе крылатых подруг в пачках, похожих на цветки анемона, и улыбавшейся ему из этого сонма крылатых, который звался – балет.* (А. И. Цветаева. Сказ о звонаре Московском (1976)); *Она, точно маргаритка, распускается с наступлением ночи и, точно анемон, от легкого дуновения свертывает свои лепестки и умирает, если к ней прикоснешься.* (Анна Ларина (Бухарина). Незабываемое (1986–1990));

2) научные статьи: *Так хищник узнает, что у ярко окрашенного животного неприятный вкус или что оно может ужалить, как, например, морской анемон.* (Л. Солодкин. Краски живого // «Химия и жизнь», 1966); *Корневище и стебель прострела содержат ядовитое вещество анемонин и продукт его расщепления анемоноль (их названия происходят еще от одного имени этого растения – анемон).* (Г. Тафинцев. Сибирский подснежник // «Химия и жизнь», 1970); *В ней обнаружили 25 морских анемонов, 75 голотурий, 5 двустворчатых раковин, другие живые существа.* (Н. Ерохина. Загадочные свечения океана // «Техника – молодежи», 1990).

Сергей Жемайтис в одной работе использовал это слово девять раз. Оно стоит в форме множественного числа, но родительный падеж указывает на мужской род: *Внизу рдели тигровые звезды, уничтожая аллеи водорослей, клумбы анемонов, коралловые полипы, рыб, моллюсков, крабов – все неисчислимое население рифа.* (Сергей Жемайтис. Большая лагуна (1977)).

В первое десятилетие XXI века начинает преобладать форма женского рода анемона. Она зафиксирована с 2001 по 2009 годы восемь раз. Данные НКРЯ приводят только примеры из нехудожественной прозы: *В английском садике высажены многолетние травянистые лесные, теневыносливые растения – аризема, хоста, ландыш, ветреница (анемона), медуница* (Лилиан Плотникова. Зеленый музей (2002) // «Ландшафтный дизайн», 2002.09.15). А форма мужского рода анемон употреблена только три раза в 2001 и 2002 годах: *Духи со свежими нотками анемона и ветивера.* (Елена Голованова. Одна-единственная часть суши (2002) // «Домовой», 2002.05.04).

Какой же вывод мы можем сделать? Какую форму должен предпочесть носитель языка, заинтересованный в правильности своей речи? На наш взгляд, при составлении различного рода справочников в настоящее время крайне целесообразно использовать данные НКРЯ, которые помогут уточнить, какая форма слова и в каких стилях речи преобладает.

Исследованное слово *анемон* – *анемона* в связи с редким употреблением вероятнее всего так и будет иметь колебание в роде.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Розенталь, Д. Э. Справочник по правописанию и литературной правке / Д. Э. Розенталь; под ред. И. Б. Голуб. – 9-е изд. – М.: Айрис-пресс, 2004. – 368 с.
- 2 Розенталь, Д. Э. Справочник по русскому языку: Правописание. Произношение. Литературное редактирование. / Д. Э. Розенталь, Е. В. Джанджакова, Н. П. Кабанова. – М.: Айрис-пресс, 2005. – 768 с.
- 3 Розенталь, Д. Э. Словарь трудностей русского языка. / Д. Э. Розенталь, М. А. Теленкова. – 4-е изд., стереотип. – М.: Рус. яз., 1985. – 704 с.
- 4 Анемона // Природа Белоруссии. Популярная энциклопедия / под ред. И. П. Шамякина. – Минск: изд-во «Белорусская советская энциклопедия» имени П. Бровки, 1989. – 598 с.
- 5 Яндекс. Словари [Электронный ресурс]. – Режим доступа: [<http://slovari.yandex.ru/>] – Дата доступа : 11.10.2012.

УДК 37.091.3(=581):004.77

А. Г. Шукайлова (Гомель)

### **ИСПОЛЬЗОВАНИЕ ИНТЕРНЕТ-РЕСУРСОВ В ПРОЦЕССЕ ПРЕПОДАВАНИЯ КИТАЙСКОГО ЯЗЫКА В ГИМНАЗИИ**

*В статье рассматриваются возможности использования сети Интернет в процессе обучения китайскому языку. Делается вывод о том, что на сегодняшний день преподаватель китайского языка должен уметь хорошо ориентироваться в огромном количестве интернет-ресурсов, которые обеспечивают овладение иностранным языком в единстве с культурой его носителей.*

В век компьютеризации и информатизации сложно переоценить значение Интернета для современного человека, поскольку «всемирная паутина» обладает безграничными информационными ресурсами и не менее впечатляющими услугами. Несомненно, Интернет проник во все без исключения сферы деятельности человека. Не стала исключением и профессиональная деятельность современного учителя. Всё чаще поднимается вопрос об использовании новых информационных технологий при обучении школьников иностранному, а в частности и китайскому языку. В данном случае имеются в виду не только новые технические средства, но и новые формы и методы преподавания, новый подход к процессу обучения китайскому языку.

Востребованность китайского языка в современном мире сейчас очевидна для всех. Предпринимаются шаги по расширению прямых связей между белорусскими и китайскими школами, начинают разрабатываться совместные образовательные программы по межшкольному культурному общению. В связи с этим показателен опыт ГУО «Гимназия № 10 г. Гомеля».

В рамках проекта «Китай становится ближе», целью которого является формирование коммуникативных компетенций через включение учащихся в личностно и общественно значимую деятельность по освоению культуры традиционного и современного Китая, была открыта секция белорусско-китайской дружбы «БелКи» Клуба Международной Дружбы Государственного учреждения образования «Гимназия № 10 г. Гомеля». Группа модераторов участвовала в создании web-страницы для общения с китайскими сверстниками: [http://vk.com/bel\\_ki](http://vk.com/bel_ki).

Цели и содержание обучения китайскому языку школьников ориентированы на «соизучение» языков и культур. Основная цель обучения состоит в формировании поликультурной многоязычной личности, способной к межкультурному общению [1]. Для реализации данной цели на уроках китайского языка нельзя недооценивать использование интернет-ресурсов. Во-первых, можно включать аутентичные материалы сети (как текстовые, так и звуковые) в содержание урока, то есть интегрировать их в программу обучения. Приведем примеры некоторых ресурсов: «Китайский язык. 50 текстов для чтения. Начальный и средний уровни» [2]; «Международное радио Китая» [3]; «Тематические уроки китайской лексики» [4] и др.

Несомненно, ресурсы глобальной сети можно использовать для получения информации по страноведческой тематике, так как в существующих учебниках этот материал часто преподносится не так интересно, красочно и иногда и без наглядности. Интернет делает возможным интерактивное путешествие по Китаю, посещение парков, осмотр памятников и любых других достопримечательностей. Например: «Китай – все о Китае, достопримечательности Китая, китайская кухня» [5]; «Достопримечательности Китая» [6] и др.

Ресурсы сети Интернет также можно использовать для систематического изучения курса китайского языка дистанционно под руководством преподавателя. Например, «Уголок МАО: Учим иероглифы быстро и легко с помощью системы МАО. Уроки китайского языка он-лайн. Софт для изучения китайского языка» [7]; «Китайский язык он-лайн» [8] и др. Например, последний из названных сайтов предлагает широчайший выбор интернет-услуг по изучению языка: сведения по грамматике, тесты, упражнения, тематические словари, учебные тексты, пословицы и поговорки, изучение иероглифики, новости на китайском языке. Сайт оказывает ряд он-лайн услуг: HSK он-лайн, китайский телеканал он-лайн, а также он-лайн консультации, которые проводит сотрудник Торгово-промышленной палаты провинции Хэйлунцзян Се Вэнькуй. Также на сайте размещена информация о возможностях обучения в китайских университетах и о вакансиях (обязательное знание китайского языка). В разделе «Полезное» можно найти более 300 фото самых живописных мест Китая, распределённых по тематическим группам: Храм лазурных облаков, чайный домик Лао Шэ, площадь Тяньаньмэнь, парк Дракона и др. Таким образом, сайт предоставляет широкие возможности для самостоятельного изучения китайского языка и культуры под руководством учителя.



В соответствии с учебной программой Министерства образования Республики Беларусь для достижения основной цели обучения иностранным языкам необходимо овладение учащимися иноязычной коммуникативной компетенцией [9]. Ведущим средством достижения этой цели является коммуникативный системно-деятельностный подход. Данный подход как нельзя лучше мотивирован: его задача заинтересовать ребят в изучении иностранного языка посредством накопления и расширения их знаний и опыта. Школьники должны быть готовы использовать язык для реальной коммуникации вне занятий, например, во время посещения страны изучаемого языка, во время приёма иностранных гостей дома, в процессе переписки. В соответствии с коммуникативным системно-деятельностным подходом интернет создает уникальную возможность для изучающих иностранный язык пользоваться аутентичными текстами, общаться с носителями языка, то есть он создает естественную языковую среду.

Обучать речевой деятельности можно лишь в общении, живом общении. А для этого нужен партнер. Компьютерная программа, CD-диск, какими бы интерактивными при этом они ни были, могут обеспечить лишь квазиобщение, т. е. общение с машиной, а не с живым человеком. Исключение составляют компьютерные телекоммуникации, когда ученик вступает в живой диалог (письменный или устный) с реальным партнером – носителем языка. Конечно же, вход в глобальную сеть непосредственно на уроке практически невозможен, так как в школах пока лишь малое количество кабинетов имеют прямой доступ к Интернету. Тем не менее, используя информационные ресурсы Сети, интегрируя их в учебный процесс, можно достичь немалых успехов в реализации коммуникативного системно-деятельностного подхода. Ведь использование сети Интернет в учебном процессе дает возможность почувствовать, что такое реальное общение, помогает лучше узнать китайский язык с помощью зарубежных друзей, способствует взаимопроникновению двух культур, обогащает учеников межкультурными знаниями и умениями. Благодаря электронной почте и различным тематическим чатам учащиеся могут пообщаться со сверстниками – носителями языка. Они расширяют свой словарный запас, в первую очередь, обиходную лексику.

Базовый набор услуг может включать в себя:

- электронную почту (e-mail);
- телеконференции (usenet);
- видеоконференции;
- возможность публикации собственной информации, создание собственной домашней странички (homepage) и размещение ее на Web-сервере;
- разговор в сети (Chat) и др.

Таким образом, на сегодняшний день преподаватель китайского языка должен уметь хорошо ориентироваться в огромном количестве интернет-ресурсов, которые, обеспечивают овладение иностранным языком в единстве с культурой его носителей, а применение компьютерных технологий следует рассматривать как один из эффективных способов организации учебного процесса.

## СПИСОК ЛИТЕРАТУРЫ

- 1 Учебная программа для общеобразовательных учреждений с белорусским и русским языками обучения. Китайский язык III–XI классы. [Электронный ресурс]. – Режим доступа: <http://www.adu.by/index.php/uchitelu/uchebnye-programmy/item> – Дата доступа: 07.10.12.
- 2 Китайский язык. 50 текстов для чтения. Начальный и средний уровни [Электронный ресурс]. – Режим доступа: <http://www.twirpx.com/file/401189> – Дата доступа: 07.10.12.
- 3 Международное радио Китая [Электронный ресурс]. – Режим доступа: <http://russian.cri.cn/other/chinese/xhy.htm>– Дата доступа: 07.10.12.
- 4 Тематические уроки китайской лексики [Электронный ресурс]. – Режим доступа: <http://www.internetpolyglot.com/lessons-zh-ru>– Дата доступа: 07.10.12.
- 5 Китай – все о Китае, достопримечательности Китая, китайская кухня [Электронный ресурс]. – Режим доступа: <http://dreamvoyage.ru/publ/1-1-0-27>– Дата доступа: 07.10.12.
- 6 Достопримечательности Китая [Электронный ресурс]. – Режим доступа: <http://www.turinfo.ru/attractions/kitay/>– Дата доступа: 07.10.12.
- 7 Уголок МАО: Учим иероглифы быстро и легко с помощью системы МАО. Уроки китайского языка он-лайн. Софт для изучения китайского языка [Электронный ресурс]. – Режим доступа: [www.umao.ru/](http://www.umao.ru/)– Дата доступа: 07.10.12.
- 8 Китайский язык он-лайн [Электронный ресурс]. – Режим доступа: <http://studychinese.ru/>– Дата доступа: 07.10.12.
- 9 Учебная программа для общеобразовательных учреждений с русским языком обучения. Иностранные языки III–XI классы. [Электронный ресурс]. – Режим доступа: <http://www.adu.gov.by/sm.aspx?guid=24463> – Дата доступа: 07.10.12.

УДК 801.519.816

И. Н. Юдкин-Рипун (Киев, Украина)

### **ЗАДАЧИ РЕФЕРИРОВАНИЯ ДРАМЫ ДЛЯ ПОСТРОЕНИЯ КОРПУСНОГО ТЕЗАУРУСА ПЕРИФРАСТИЧЕСКИХ ВЫРАЖЕНИЙ**

*Предложен алгоритм описания перифрастики драматического текста, отличающийся тем, что вносятся поправки на преобразование прямой речи партнеров диалога в текст наблюдателя, являющийся источником для построения тезауруса.*

Предложенное В. М. Савицким новое понимание идиоматики, основанное на интерпретируемости (в частности, переводимости) выражений [3, с. 37], открывает перспективы для применения корпусного подхода. В частности, идиоматизация имеет смысл по отношению к тому корпусу текстов, в котором выражение функционирует. Например, предложение «*если бы вы знали меня, то знали бы и моего отца*» (Евангелие от Иоанна, 8, 19) может рассматриваться как «разовое высказывание» (по В. Г. Адмони) или как иллюстрация условного наклонения, однако в отнесении к корпусу

евангельских текстов оно не только превращается в идиому, но и становится теологическим тезисом, породившим различные толкования. Особенно актуализируется корпусная обусловленность идиоматической интерпретации выражений в драматических текстах, когда общие места благодаря референции раскрывают скрытый идиоматический смысл, как, например, в известной фразе Банко из шекспировского «Макбета» *“there will be rain tonight”* (сегодня ночью будет дождь), где предсказание погоды обретает смысл пророчества о кровопролитии.

Особый класс идиоматики составляют перифрастические выражения, представляющие обозначаемый объект через описание отсылающих к нему частичных и косвенных признаков. Замена прямого указания перифрастическим описанием составляет необходимый компонент идиолекта драмы и лирики. Например, бегство Подколесина через окно в финале гоголевской «Женитьбы» является совершенно очевидным перифрастическим обозначением смерти, будучи соотнесено с мотивом разбитого зеркала в начале драмы. Слова Арбенина к Нине (из лермонтовского «Маскарада») *«нынче я доволен был тобой»* оказываются перифрастическим обозначением свершившегося отравления в свете того, что известно зрителю. Для драмы перифрастические описания становятся ведущим средством идиоматизации и требуют специального изучения.

Задача составления тезауруса подобных перифрастических выражений и связанных с ними мотивов для корпуса драм предполагает, в свою очередь, подготовку специально препарированных текстов, поскольку по непосредственному описанию невозможно определить смысловые сдвиги. В частности, перед поиском элементов тезауруса необходимо составление свертки текста для определения референций выделенного перифрастического выражения, описание сюжетной схемы текста. Еще на заре компьютерной эпохи И. П. Севбо отметила двойственность такой процедуры, расщепляющей текст на пропозициональные и перечислительные (списочные) структуры: «Реферат, как правило, состоит из законченных мыслей, содержащих субъект и предикат. Аннотация, напротив, может содержать лишь перечень субъектов» [4, с. 94]. Практика так называемых дайджестов для прозаических жанров и либретто для музыкального театра предоставляет материал для разработки соответствующего программного обеспечения.

Особенность драматического текста, построенного по принципу перлокутивного речевого акта (диалога, фактически обращенного к третьему лицу – аудитории), состоит в невозможности обеспечить его связность только за счет его внутренних ресурсов. Партнеры диалога, высказывают только частичные сведения и потому предполагают текст компетентного арбитра, обеспечивающего недостающие звенья мотивировки высказываний. Такую компетенцию в античной драме частично демонстрировал хор, выступавший в роли всезнающего повествователя, однако и там обмен реплик предполагает дополнение комментариями для адекватного воспроизведения смысла. Например, в финале «Эдипа» Сенеки слова героя, восхваляющие

тму как прощение, понятны только с учетом известного из предыдущего сообщения Вестника, описывающего ослепление. Поэтому обычная процедура свертки текста применительно к драме нуждается в поправке: перед тем как резюмировать, необходимо прокомментировать текст.

Эта поправка, в свою очередь, предполагает представления реплик в качестве того, что Р. Ковальски предложил описывать как так называемые клаузы, то есть посылки, предполагающие заключения [1, с. 21]. Поскольку именно выводы наблюдателя являются решающими для связности драматического текста, то и реплики диалога раскрывают смысл именно в качестве таких посылок для заключений на основе известной наблюдателю информации. Например, в заключительной сцене Гамлета решение матери героя Гертруды выпить предназначенный ему бокал с большой долей вероятности указывает на стремление спасти сына, которое не выводимо непосредственно из реплик: *—Queen: Here, Hamlet, take my napkin, rub thy brows; / The Queen carouses to thy fortune, Hamlet! / Hamlet: Good madam! King: Gertrude, do not drink. Queen: I will, my lord; I pray you, pardon me. King (aside): It is the poisoned cup; it is too late*". В предыдущем акте слова Короля в разговоре с Лаэртом о намерении отравить кубок были произнесены сразу перед появлением Гертруды, что подсказывает мысль о том, что они были ею услышаны. Эта информация позволяет из приведенных реплик сделать дополнительные заключения: зная о подготовленном отравлении, Гертруда намеревается спасти сына и осуществляет то, о чем говорится в репликах. В свете таких заключений слова Гертруды обретают перифрастическое значение: *выпить за удачу* (to carouse to fortune) значит *спаси*, *дать сыну платок вытереть брови* обретает смысл как *вытереть лоб (от пота)*, так и *вытереть глаза (от слез)*, а просьба о прощении указывает на знание о совершаемом самоубийстве.

Преобразование диалога в структуру гипотаксиса с репликами в качестве посылок для собственных заключений наблюдателя связано и с необходимостью выявления актуальных предикатов (рем актуального членения предложений), предполагающее использование методов представления данных. Так, описание текста фреймами, фиксирующими стандартные ситуации, дает только потенциальную синтаксическую структуру, предустановленную самими слотами фрейма, тогда как для выявления актуальной структуры необходимо учесть намерения участников коммуникации. Поэтому, в частности, для драматического текста представляется полезным учитывать модель ТЕКСТ – СМЫСЛ – ЦЕЛЬ, разработанную Э. В. Поповым, где предполагается, что «основная задача общения состоит в том, чтобы определить цели участников общения» [2, с. 119]. Учет актуальной структуры, связанной с намерениями, особенно важен для адекватности описания драматического текста, представляющего акт общения. Например, в чеховской «Чайке» в заключительном (перед самоубийством) объяснении Треплева в его реплике «я не верую и не знаю, в чем мое призвание» ремой оказывается отрицание, что подтверждается параллелизмом к предыдущей

реплике Заречной «*я верую ..., и когда думаю о своем призвании, то не боюсь жизни*». Негатив становится тут перифрастическим описанием того, что свершится через несколько минут, так что «отсутствие призвания» означает фактически отсутствие воли к жизни.

С учетом указанных особенностей драматического текста был разработан алгоритм представления данных для перифрастики. Предварительно препарированный текст включает материал реплик в качестве подчиненных предложений гипотаксиса, построенного на основании компетентности наблюдателя, владеющего информацией о разворачивании событий драмы. Такой текст реплик партнеров диалога, дополненный комментариями наблюдателя и представленный в гипотаксисе, становится объектом стандартного описания по методу семантических сетей, которое позволяет выявить потенциальную структуру. Наряду с текстуальной перспективой, представленной таксисом предикатов, используется предложенное Г. Р. Яуссом понятие горизонта как порога различимости семантических признаков [5, с. 466], где обобщаются представления об омонимии и синонимии. Например, игра с изменением горизонта наблюдается в эффекте омонимического расщепления, как чеховском «Вишневом саду» (действие 4) в обмене репликами Трофимова («... *я могу проходить мимо вас, я силен и горд*») и Лопатина («... *а жизнь знай себе проходит*»), где высвечиваются контрастные смыслы одного и того же глагола «проходить».

Решающим шагом к представлению перифрастики является переход от потенциальной к актуальной структуре и к определению ремы каждого отдельного высказывания. Существенно, что в этом переходе предикатами становятся элементы вышеупомянутого горизонта, выступающие потенциально как сопутствующие обстоятельства, детализирующие высказывания. Например, в драме А. И. Сумбатова-Южина «Джентльмен» (д. 4, явление 5) в восклицании Кэтт «*Будь вы в самом деле женаты, я стала бы вашей любовницей*», завершающей размолвку с ее приятелем Остужевым после его противоречивых заявлений (в частности, «*только свободно я могу любить беззаветно*») ремой оказывается оборот «в самом деле», указывающий на лживость этих заявлений. В свою очередь, высказывания самого Остужева оказываются перифрастическими выражениями эгоистических склонностей его характера, что засвидетельствовано еще одной репликой Кэтт «*никто меня не спросит, чем же я хочу быть*». Здесь ремой оказывается местоимение первого лица, противопоставленное эгоцентрической речи партнера, а выражение «никто не спросит» оказывается перифрастической заменой «вам безразлично».

Таким образом, создание свертки драмы (либретто, дайджеста) позволяет проинтерпретировать текст реплик партнеров диалога и выявить его актуальную структуру. На этой основе становится возможным обнаружить перифрастическое использование выражений для обозначения подтекста и составить соответствующий тезаурус.

## СПИСОК ЛИТЕРАТУРЫ

- 1 Ковальски, Р. Логика в решении проблем / Р. Ковальски. – М.: Наука, 1990 – (Проблемы искусственного интеллекта) – 280 с.
- 2 Попов, Э. В. Общение с ЭВМ на естественном языке / Э. В. Попов. – М.: Наука, 1982 – 360 с.
- 3 Савицкий, В. М. Основы общей идиоматики / В. М. Савицкий. – М.: Гнозис, 2006 – 208 с.
- 4 Севбо, И. П. Структура связного текста и автоматизация реферирования / И. П. Севбо. – М.: Наука, 1969 – 136 с.
- 5 Яус, Ганс Роберт. Досвід естетичного сприйняття і літературна герменевтика / Ганс Роберт Яус. – К.: Основи, 2011 – 624 с.

УДК 303.022:81'42

Е. И. Ядченко (Минск)

### **ПОТЕНЦИАЛ КОМПЬЮТЕРНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ MAHQDA В ЛИНГВИСТИЧЕСКОМ ИССЛЕДОВАНИИ**

*Объектом исследования в данной статье является компьютерное программное обеспечение MaxQDA. На примере анализа дискурса мигрантов демонстрируются основные функции программы, релевантные для лингвистического исследования: функция построения тематических деревьев, функция лексического поиска, функция анализа лексической частотности.*

Одним из новых направлений в исследовании языка и мышления явился в последние десятилетия *дискурс-аналитический подход* [1], [2], [3]. Объектом исследования лингвистического дискурс-анализа выступает текст в социальном контексте, в его информативной и интерактивной данности, в единстве его «контентной» (собственно содержательной) и контекстной (ситуационной) составляющих [3]. В такой перспективе содержание текста, отныне дискурсии, представляет собой феномен, выходящий за рамки текстового пространства как объективного данного. А основным методом его изучения становится *реконструкция* с учётом как лингвистических, так и экстралингвистических, прагматических, социокультурных, психологических и других параметров.

Создаваемое как вспомогательное средство для качественного анализа данных по принципу «Обоснованной теории» [4], [5], современное компьютерное программное обеспечение MaxQDA позволяет осуществлять анализ (относительно) больших корпусов данных, сокращая временные затраты исследователя. Данное программное обеспечение является многофункциональным и довольно гибким, оно может применяться в различных гуманитарных исследованиях – в социологии, психологии, педагогике, лингвистическом анализе текста [6, с. 8], так как работает с широко используемой

категорией «кодирование», однако принципиально не задаёт каких-то критериев. Следовательно, код может объединять в себе самые разные по объёму содержательные единицы – от одной лексемы до нескольких абзацев текста.

Мы используем компьютерное программное обеспечение MaxQDA как вспомогательное средство при лингвистическом анализе дискурса мигрантов. Наша процедура анализа включает два основных этапа: (1) структурный анализ содержания дискурсии и (2) функциональный анализ содержания дискурсии.

1. На этапе **структурного анализа дискурсии** мигранта мы используем *функцию построения тематических деревьев* [6, с. 37]. Тематическое дерево дискурсии представляет собой некую модель содержания данной дискурсии с точки зрения её тема-рематического развития. Данный анализ рассматривается нами как подготовительный этап для последующего функционального анализа, при котором тематические деревья используются как легко управляемый гипертекст. Для этого мы осуществляем кодирование дискурсии.

На операциональном уровне кодирование осуществляется следующим образом: в поле «Список документов» импортируется документ в формате word либо pdf, с которым исследователь собирается работать (в левом верхнем углу, текст «Лина»). При наведении курсора на данный документ и подтверждении двойным нажатием левой клавишей мыши текст высвечивается в поле «Документ» (справа сверху) появляется текст. Именно это поле и является основным рабочим. При помощи мыши исследователь выделяет отрезок текста и присваивает ему код. Каждый код высвечивается рядом с текстом, как если бы исследователь работал с бумагой и маркером. При этом возможна цветовая маркировка кодов, что в дальнейшем облегчает поиск нужных отрезков по тексту и анализ (Рисунок 1).

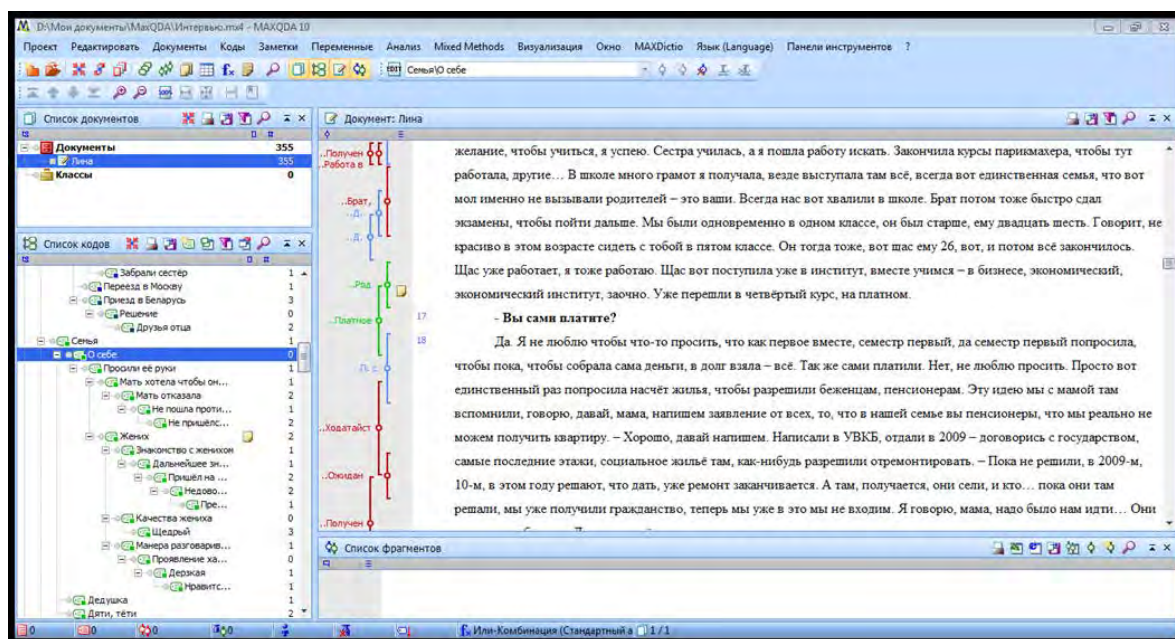


Рисунок 1. Кодирование и структурирование тем

Все коды подаются в виде списка в поле «Список кодов» (слева внизу), который может быть упорядочен в процессе кодирования. Так, возможно переименование отдельных кодов по принципу уточнения, а также структурирование системы кодов как иерархической лестницы.

Графа «Визуализация» предлагает техники, при помощи которых возможно графическое оформление системы кодов в виде определённых графиков. Так, например, тематическую структуру интервью возможно представить в виде МахМар (рисунок 2). При этом расстановка графических объектов на экране производится заново (не автоматически). Данное тематическое дерево при помощи простейших операций может быть экспортировано в графический документ и при необходимости распечатано в формате фотографий.

Функция «Заметки» позволяет в процессе анализа текста «останавливаться» на определённых позициях и «прикреплять» к ним в виде памяток необходимые комментарии, пояснения (рисунок 3), которые в дальнейшем могут использоваться на последующих этапах анализа.

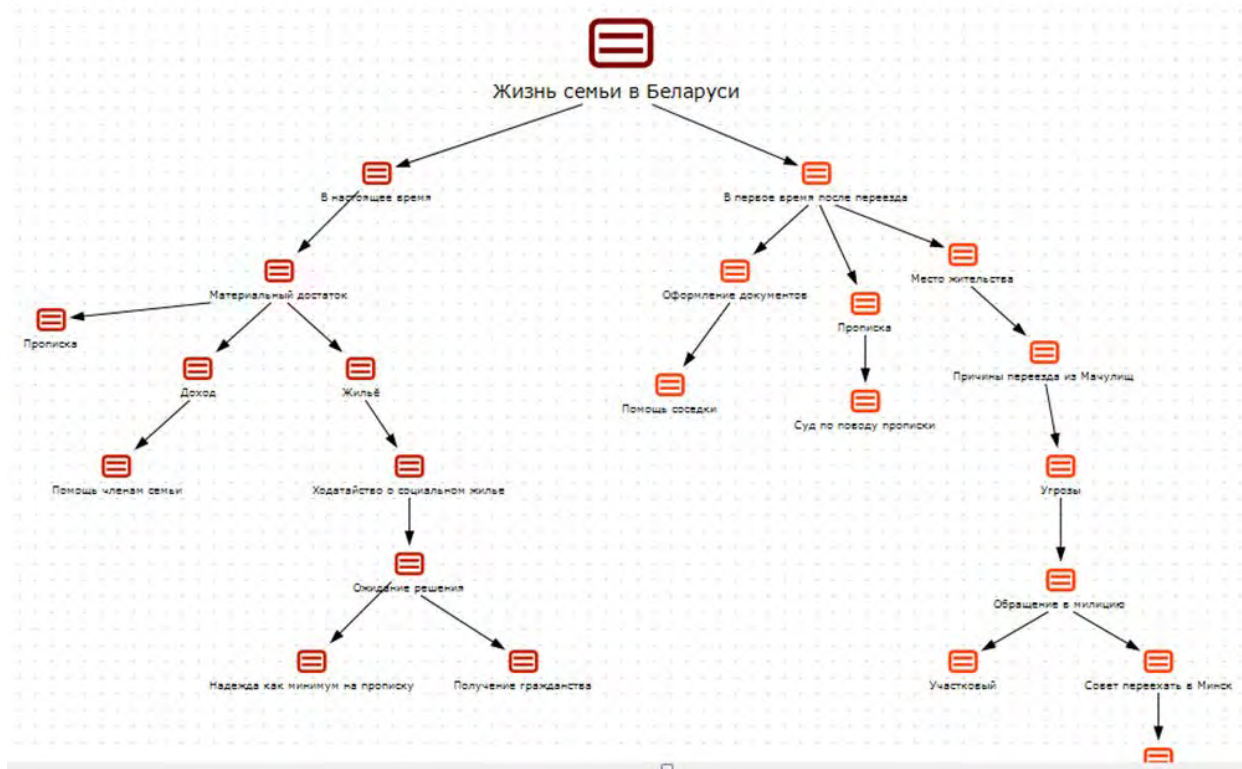


Рисунок 2. Тематическое дерево

2. На этапе **функционального анализа дискурсии** мигранта осуществляется выявление содержательного наполнения тем дискурсии. Наряду с качественным анализом содержательного наполнения тем мы учитываем и количественную реализацию, поскольку она может сигнализировать дискурсивную.



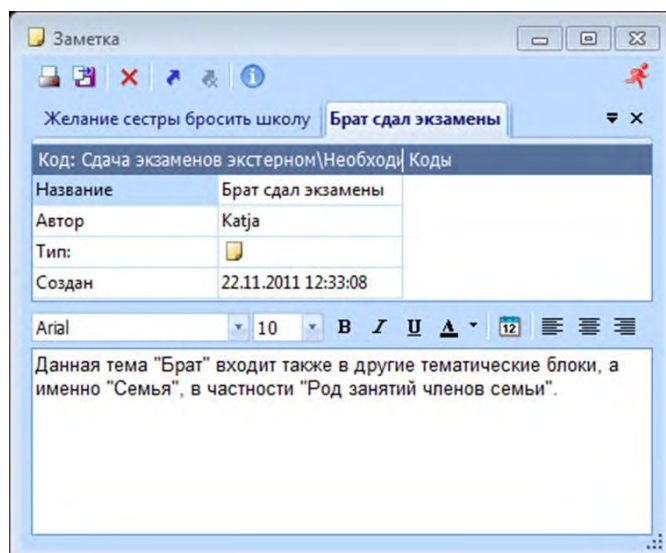


Рисунок 3 – Заметки

значимость определённых феноменов, получивших реализацию в дискурсе. На данном этапе становятся актуальными следующие функции программы MaxQDA:

*Функция лексического поиска по тексту.* При вводе лексического компонента в строку поиска отыскиваются все позиции данного слова и указывает абзац, в котором оно присутствует (рисунок 4). Одновременно происходит выделение данного слова в тексте [7, с. 41].

*Функция быстрого анализа частотности слов.* При выборе данной функции в виде таблицы подаётся список всех слов из текста (рисунок 2). Из этого списка мы можем выбрать интересующий нас компонент по самому слову (в алфавитном порядке), процентному соотношению искомого слова ко всему тексту либо длине слова (например, при поиске местоимений, междометий) [7, с. 43].

Документ	Поиск	Страница	Абзац
Фатима	побили	1	2
Фатима	побили	1	2
Фатима	побили	1	7
Фатима	побили	1	7
Фатима	побили	1	11
Фатима	побили	1	11
Фатима	побили	1	11
Фатима	побили	1	11
Фатима	побили	1	11
Фатима	побили	1	11
Фатима	побили	1	12
Фатима	побили	1	30
Фатима	побили	1	30
Фатима	побили	1	59

Рисунок 4. Лексический поиск

Слово	Длина слова	Частота	%	Фатима
поближе	7	1	0,01	1
побили...	7	1	0,01	1
побили	6	12	0,13	12
побальзай	9	1	0,01	1
по-ноему	8	7	0,08	7
по-другому	10	1	0,01	1
по	2	15	0,16	15
плохо	5	6	0,07	6
племянница	10	1	0,01	1
племянники	10	1	0,01	1
племянник	9	2	0,02	2

Рисунок 5. Анализ частотности слов

## СПИСОК ЛИТЕРАТУРЫ

- 1 Дейк, ван Т. А. Язык. Познание. Коммуникация: пер. с англ. / Т. А. ван Дейк. – М.: Прогресс, 1989. – 312 с.
- 2 Ingo, H. Diskurslinguistik nach Foucault: Theorie und Gegenstände / H. Ingo, Warnke (Hg.). – Berlin, New York : Walter de Gruyter, 2077. – 278 s.
- 3 Ухванова-Шмыгова, И. Ф. Методология исследования политического дискурса: Актуальные проблемы содержательного анализа общественно-политических текстов / И. Ф. Ухванова-Шмыгова, А. А. Маркович, В. Н. Ухванов // под общ. ред. И. Ф. Ухвановой-Шмыговой. Выпуск 3. – Мн.: «Технопринт», 2002. – 360 с.
- 4 Страусс, А. Основы качественного исследования / А. Страус, Дж. Корбин // Обоснованная теория: процедуры и техники. – М.: УРСС, 2001. – 253 с.
- 5 Strübing, Jörg. Grounded Theory / J. Strübing. – Wiesbaden: VS Verlag, 2004. – 182 S.
- 6 Kuckartz, Udo. Einführung in die computergestützte Analyse qualitativer Daten / Udo Kuckartz. – Wiesbaden: VS Verlag für Sozialwissenschaften, 2005. – 255 s.
- 7 Kuckartz, U. Qualitative Datenanalyse computergestützt / U. Kuckartz, H. Grunenberg, A. Lauterbach // Methodische hintergründe und Beispiele aus der Forschungspraxis. – Wiesbaden: VS Verlag, 2004.

УДК 811.161.3

Е. Ф. Асенчик (Гомель)

### **ИСПОЛЬЗОВАНИЕ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В КОМПЛЕКСНОМ АНАЛИЗЕ ЛЕКСИЧЕСКИХ ЕДИНИЦ**

*В статье рассматривается комплексный анализ лексемы грех, осуществляемый с помощью программы «Грамматический словарь» и интернет-ресурсов. Делаются выводы о необходимости использования подобных видов анализа для получения наиболее полной информации о слове, для прослеживания многочисленных связей слова, в том числе межъязыковых, для статистических данных об употреблении слова в речи.*

Современное общество, науку невозможно представить без новейших информационных технологий (ИТ). Современная лингвистика, а также области гуманитарного знания, непосредственно с ней связанные, весьма успешно пользуются различными разработками и достижениями в области ИТ. Активно развиваются такие новые и наиболее перспективные направления языкознания, как корпусная лингвистика, где исследовательский процесс невозможен без современных компьютерных антологий – языковых баз данных; также квантитативная лингвистика, исследующая язык при помощи статистических методов и тесно связанная с корпусной и компьютерной лингвистикой, поставляющими важные эмпирические данные.

Применение ИТ при решении вопросов, связанных с передачей смыслового содержания и формы языковых единиц одного языка средствами другого, весьма популярно. На данном этапе развития современной

переводческой деятельности компьютерные технологии позволяют достаточно легко преодолеть языковой барьер.

В условиях немалого разнообразия ИТ встает вопрос о создании технологий, достаточно компактных, удобных в использовании и в то же время содержащих в себе комплексную информацию, допустим, о лексической единице языка. Подобные задачи достаточно успешно решает программа «Грамматический словарь».

Традиционные (бумажные) грамматические словари – это словари, которые содержат сведения о морфологических и синтаксических свойствах слова. Грамматические словари включают слова, расположенные в прямом или обратном алфавитном порядке. Принципы отбора и объем сведений о слове различны в зависимости от назначения и адресата каждого грамматического словаря.

**Электронный Грамматический словарь русского языка** (версия 8.54.7746) предназначен для быстрого поиска и удобного просмотра информации в морфологической базе данных русского языка, в частности таблиц склонения и спряжения, списков синонимов и антонимов, а также сведений о словоизменении и словообразовании. Он объединяет в себе особенности словаря-программы (устанавливается на персональный компьютер и может быть использован без привлечения Интернета) и онлайн-словаря (имеет гиперссылки, следовательно, обращается к Интернету), а также использует материалы интернет-энциклопедии «Википедия».

Рассмотрим **возможности грамматического словаря** [1]:

- Программа выполняет **быстрый поиск** словарных статей по первым буквам любой грамматической формы слова, поиск по маске с использованием символов \* и ?, нечеткий поиск похожих слов среди более чем 164 тысяч русских словарных статей.

- Словарь описывает **словоизменение**: склонение существительных (45 тысяч статей, 442 тысячи форм), прилагательных и причастий (54 тысячи статей, 1 млн 560 тысяч форм), спряжение глаголов (18 тысяч статей, 217 тысяч форм), формы наречий (7000 статей) и числительных для русского и английского языков.

- **Тезаурус** содержит информацию о связях слов, включая синонимы (60 тысяч), антонимы (более 5000), уменьшительные формы (более 3000) и словообразовательные связи (более 400 тысяч), переводы, гипонимы и гиперонимы.

- **Разбивка на слоги** слов русского языка.
- **Ёфикация** – отображение слов с буквой ё.
- **Употребление слов**: частота, короткие предложения с подсветкой вхождения, отметки для нецензурных, грубых и пр. слов, ударения.
- **Поиск слов в лексиконе**.

Программа выполняет поиск словарных статей по мере ввода пользователем первых букв, постепенно уточняя выбор. При этом обновляется список найденных словарных статей, двойной щелчок мыши в котором выводит страницу соответствующей статьи.

Словарь позволяет искать слова по маске с использованием специальных символов ? и \*. Символ ? означает одну любую букву, а символ \* - любую цепочку букв. Таким образом, маска **грех\*** позволяет найти существительные *грех, греховность, греховодник* и *грехопадение*, прилагательное *греховный*, глаголы *грех* (безличный), *грешить, греховодничать*, наречие *греховно*.

Еще один вид поиска - **нечеткий**, в ходе которого будут найдены слова, приблизительно соответствующие введенному образцу. Этот режим включается с помощью ввода символа ~ в начале образца, например ~**грех**. В результате программа находит слова, содержащие в своём составе сочетание букв «гр», «гор», «рех»: *огрех, угорь, сгрести, огреть, игра, дуреха*.

### Разметка страниц словарных статей

Словарная статья состоит из:

- 1) шапки статьи;
- 2) таблицы склонения или спряжения со всеми формами;
- 3) примеров употребления;
- 4) информации из тезауруса: синонимы, антонимы и так далее;
- 5) ссылки на «Викисловарь».

Шапка словарной статьи содержит набор отметок о морфологических свойствах, общих для всех грамматических форм слова:

**Грех** *сущ. неодуш. муж. р. грех 31497*

Первой подаётся базовая форма словарной статьи с ударением (если слово неодносложное), затем – имя грамматического класса (в нашем примере – существительное), далее – основные грамматические категории (для существительного *одушевлённость/неодушевлённость, род*), для русских глаголов, причастий и деепричастий выводится информация о *переходности и виде*.

Также в шапке выводится информация о делении основной формы слова на слоги и ориентировочной частоте употребления слова (в нашем примере – *31497*)

Раздел информации из тезауруса содержит примеры употребления слова в тексте, а также разнообразные связи слова с другими словами и соответственно словарными статьями программы, в том числе – уменьшительные и увеличительные формы, синонимы и антонимы, однокоренные и всевозможные производные слова:

*Употребление:* Грех предаваться унынию, когда есть другие **грехи**. Христианство считает **грехом** и безумием любое обращение к диаволу в колдовстве, гаданиях и других суевериях. Книга Откровения заканчивается пророчеством о том, что победа Бога над **грехом** увенчает тяжёлую

борьбу. В них доктор богословия выступал против продажи индульгенций и власти Папы над отпущением **грехов**.

*Уменьшительные формы:* грешок

*Синонимы:* проступок, прегрешение

*Словообразование – причастие:* грешивший, грешащий, погрешивший, согрешивший

*Словообразование – прилагательное:* безгрешный, грешный, греховный

*Словообразование – наречие:* греховно

*Словообразование – существительное:* грешник, грехопадение

*Словообразование – глагол:* грешить, согрешить

*Переводы:* evil<sub>англ</sub> sin<sub>англ</sub>

Выводимые в этом разделе слова оформлены как гиперссылки, то есть при клике мышью на них выводят страницу соответствующей статьи.

Можно переключить программу в режим отображения **диаграммы** для связанных слов, текущее слово отображается в центре, связанные с ним слова кликабельны.

Следующим этапом идёт ссылка на **Викисловарь** [2] (при нажатии на ссылку информация открывается в браузере, а не в окошке программы, т. е. для дальнейшей работы требуется наличие интернета). Информация из Викисловаря отражает наличие у слова *грех* **антонима** *благодетение*, **гиперонимов** *нарушение*, *поступок*, **гипонимов** *ложь*, *убийство*, *воровство*, *богохульство*. Также приведены значения слова, его употребление (со ссылкой на НКРЯ) и этимология (со ссылкой на этимологический словарь М. Фасмера):

1. *религ.* нарушение (действием, словом, мыслью) воли Бога, религиозно-нравственных предписаний, правил ◆ **Смертный грех**.

2. *перен.* предосудительный поступок ◆ что начала, то кончу; уж мне **грех** тебя оставить; вот тебе мужицкое платье: скинь-ка свой балахон... М. Ю. Лермонтов, «Вадим», 1833 г. (цитата из Национального корпуса русского языка [3])

**Этимология:** Происходит от праслав. формы *grěxъ*. Скорее всего, слав. *grěxъ* связано с *греть* с первонач. знач. «жжение (совести)»; <...> (Использованы данные словаря М. Фасмера).

Далее приводятся **фразеологизмы** и **устойчивые сочетания**, содержащие исследуемое слово: *от греха подальше*, *и смех и грех*, оформленные как гиперссылки, при нажатии на которые открываются соответствующие статьи.

Т. о. налицо системный анализ слова (грамматический, семантический, словообразовательный, этимологический и проч.), достигаемый взаимодействием различных компьютерных технологий, в т.ч. электронных программ и интернет-ресурсов.

## СПИСОК ЛИТЕРАТУРЫ

- 1 Грамматический словарь: Компьютерная лингвистика. [Электронный ресурс] – Режим доступа: <http://www.solarix.ru/>.
- 2 Грех: материал из Викисловаря. [Электронный ресурс] – Режим доступа: <http://ru.wiktionary.org/wiki/грех>.
- 3 Национальный корпус русского языка. [Электронный ресурс] – Режим доступа: <http://www.ruscorpora.ru/index.html>.

УДК 81'3

Е. Е. Иванов, В. А. Старосельский (Могилев)

### **УЧЕБНАЯ АВТОМАТИЗИРОВАННАЯ СИСТЕМА ЛЕКСИКО-ГРАММАТИЧЕСКОГО АНАЛИЗА ТЕКСТА (при изучении английского языка на начальном этапе)**

*Описываются принципы разработки, структура и способы использования компьютерной программы лексико-грамматического анализа английского текста в учебных целях.*

Целью исследования было создание и апробация компьютерной программы для поиска и выделения частей речи в текстах на английском языке в учебных целях. Программа должна предоставить возможность автоматизированного выделения и отнесения слов, которые содержатся во введенном пользователем тексте, к тем или иным частям речи. Кроме того, программа должна предусматривать добавление в базу данных каждой части речи таких слов, которые невозможно или затруднительно идентифицировать в машинном режиме.

Для создания программы был использован язык программирования Microsoft Visual Basic, который позволяет быстро и легко организовать пользовательский интерфейс программы, а также позволяет организовать простую работу с файлами и базами данных.

В любой пользовательской компьютерной программе ключевую роль играет интерфейс. Очень важно, чтобы он был прост и понятен рядовому пользователю, что упрощает работу пользователя с программой.

При создании программы лексико-грамматического анализа текста к интерфейсу были предъявлены следующие требования: наглядность, простота, понятность, габариты. Пользовательский интерфейс программы получил следующий вид (см. Рисунок 1).

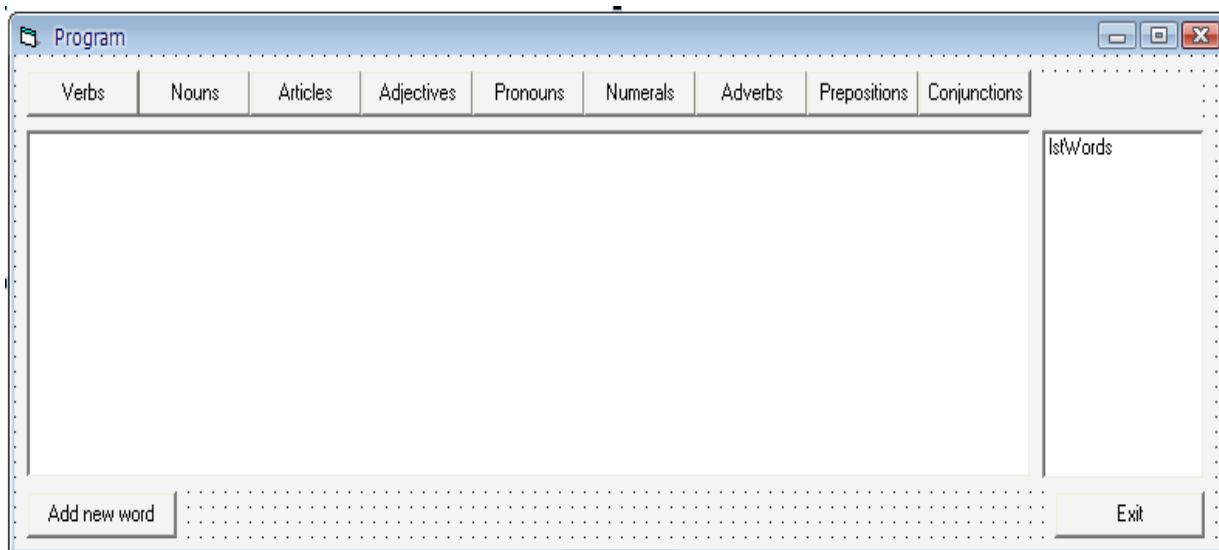


Рисунок 1

Интерфейс программы состоит из нескольких частей. Одна часть включает в себя верхнюю часть окна, где располагаются основные кнопки, выполняющие соответствующие их названию функции данной программы. В нижней части окна находится кнопка «Add New Word», с помощью которой, пользователь сможет добавлять слова в базу данных. Основную часть экрана занимает текстовое поле, где располагается вводимый пользователем текст на английском языке, из которого осуществляется выборка частей речи. При нажатии на одну из кнопок все части речи, соответствующие ее названию, из основного текстового окна переносятся в окно справа.

Все элементы интерфейса имеют понятные названия, находятся на привычных для пользователя местах и не должны вызывать затруднений во время использования программы.

Поскольку программа предназначена для учебных (обучающих) целей (поиск и квалификация частей речи), то вместе с ней одновременно могут использоваться и другие программы, например, текстовый редактор или интернет-браузер. Поэтому интерфейс данной программы имеет небольшие габариты, что позволяет эффективно работать одновременно с двумя или более программами.

У учебной программы интерфейс не должен визуальнo раздражать или надоедать при ее длительном использовании. При оформлении интерфейса программы были использованы стандартные оттенки, которые апробированы и широко используются во многих программах операционной системы Windows.

Основная задача данной учебной программы состоит в определении и дифференциации различных частей речи в заданном тексте, который должен быть вставлен пользователем в текстовое поле. С помощью средств Visual Basic были созданы функции, которые содержат все возможные регулярные признаки для каждой части речи в английском языке (например, наличия артикля у существительных, окончания *-ing* у глаголов, показателей степеней

сравнения *-er* и *-est* у прилагательных и т. п.). Каждое слово подлежит обработке этой функцией, после чего его относят к соответствующей части речи.

Разумеется, в английском языке, как и во многих других языках мира, есть слова, которые не имеют регулярных грамматических показателей своей частеречной принадлежности, например, заимствования из других языков (*data*, *phenomena* и т. д.). При употреблении этих слов в речи, правда, есть случаи, когда перед ними стоит артикль, но это бывает не всегда. Для этих слов создается специальная база данных.

Вывод необходимой информации на экран происходит при нажатии на соответствующую кнопку («Verbs», «Nouns», «Adverbs», «Numerals» и т. д.). При этом происходит следующее: (1) текст в текстовом поле записывается в оперативную память, после чего его можно использовать для своих нужд; (2) текст делится на слова (в оперативной памяти программы); (3) каждое слово относится к определенной части речи (в оперативной памяти программы); (4) в правое текстовое поле из оперативной памяти выводятся слова, относящиеся к запрашиваемой части речи.

В том случае, когда машина затрудняется отнести слово к какой-либо части речи, то программа обращается к базе данных частей речи. Если ответ положительный, то часть речи будет определена, если отрицательный, то функция определить часть речи и ввести данное слово в соответствующий раздел базы данных переходит к пользователю (так осуществляется, с одной стороны, пополнение базы данных, а с другой, учебная цель).

Добавление новой записи производится нажатием кнопки «Add New Word». Затем пользователь видит диалоговое окно добавления новой записи. От пользователя требуется ввести английское слово, выбрать к какой части речи оно относится, после чего нажать кнопку «ОК» (см. Рисунок 2). Если пользователь введет английское значение уже существующего слова, то он будет проинформирован сообщением, что такая запись уже существует.

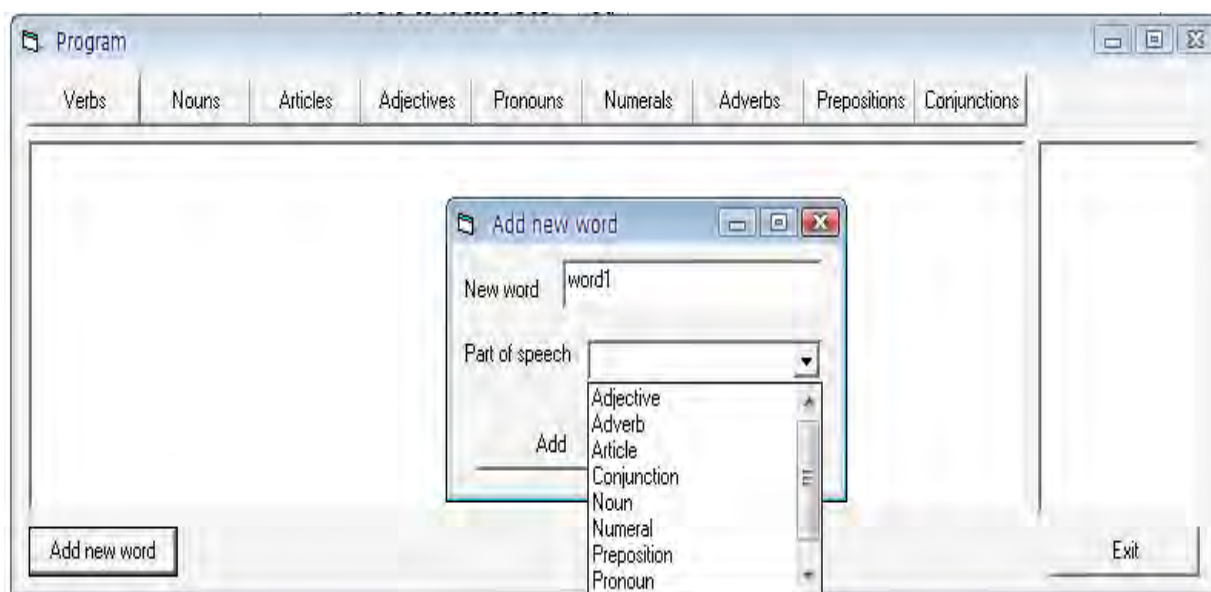


Рисунок 2



Данный процесс похож на ведение «бумажного» словаря на занятиях по практике устной и письменной речи.

Практическая значимость разработки компьютерной программы лексико-грамматического анализа английского текста в учебных целях состоит в возможности более широкого применения автоматизированных операций при обработке языковых данных при изучении английского языка как иностранного на начальном этапе в высшей и средней школе.

УДК 811.161.3

С. Ф. Иванова (Магілёў)

### **ТРАНСФАРМАЦЫІ Ў МАЎЛЕННІ КРЫЛАТЫХ ВЫРАЗАЎ З РУСКАМОЎНЫХ ЛІТАРАТУРНЫХ КРЫНІЦ (на матэрыяле антыцытат з тэкстаў Байнэта)**

*Разглядаюцца парадыйныя, іранічныя і жартоўныя трансфармацыі ў маўленні крылатых слоў з рускамоўных літаратурных крыніц на матэрыяле антыцытат з тэкстаў Байнэта.*

У інтэрнэт-маўленні крылатыя словы ўжываюцца як без змянення іх сэнсу і знешняй формы, так і з пэўнымі змяненнямі, у мадыфікаваным выглядзе. Змяненні звычайна не закранаюць змест крылатых слоў, паколькі яго захоўванне з'яўляецца неабходнай умовай функцыянавання крылатых слоў як інтэртэкстаў. У тых жа выпадках, калі трансфармацыя знешняй формы вядзе да змянення сэнсу крылатых слоў, утвараюцца новыя адзінкі, якія або выконваюць намінацыйную функцыю (скіраваны на адлюстраванне нейкай з'явы, падзеі, служаць для наймення нейкіх рэалій ці выражэння пачуццяў і г. д.), або ствараюцца выключна дзеля жарту, іранічнага асэнсавання рэчаіснасці, пародыі на зыходную адзінку, змест яго тэкставай крыніцы ці светапогляд яе аўтара. Такія парадыйныя, іранічныя і жартоўныя трансфармацыі крылатых выразаў адносяцца да жанру т. зв. «**нтыцытат**» [1] (як адной з разнавіднасцей «**нтыпрыказак**» у шырокім сэнсе [2]) і досыць распаўсюджаны ў цяперашні час.

Ніжэй прапануецца невялікая падборка антыцытат, якія ўтварыліся ў выніку парадыйнай, іранічнай ці жартоўнай трансфармацыі тых крылатых слоў з рускамоўных літаратурных крыніц, што ўжываюцца ў сучаснай беларускай літаратурнай мове [3].

**Жыццё – гэта гульня. У дурня** ⇐ *Што наша жыццё?* – Гульня < руск. *Что наша жизнь?* – *Игра* – выслоўе з оперы «Никова дама» (1890) П. І. Чайкоўскага (1840–1893), лібрэта да якой напісаў М. І. Чайкоўскі (1850–1916), паводле аднайменнай аповесці (1833) А. С. Пушкіна (1799–1837).

**Жыццё даецца чалавеку адзін раз, і то выпадкова** ⇐ *Жыццё даецца <чалавеку толькі> адзін раз <, і пражыць яго трэба так, каб не было сорамна за дарэмна пражытыя гады> < руск. Самое дорогое у человека – это жизнь. Она даётся ему один раз, и прожить её надо так, чтобы не было мучительно стыдно за бесцельно прожитые годы – афарызм з рамана «Как закалялась сталь» (1932–1934) М. А. Астроўскага (1904–1936).*

**Зязюля хваліць пеўня, а певень топча курыцу** ⇐ *Зязюля пеўня ўсхваляе, Таму што хваліць ён яе < руск. Кукушка хвалит петуха За то, что хвалит он кукушку – радкі з байкі «Кукушка и петух» (1834) І. А. Крылова (1769–1844).*

**Народжаныя поўзаць нярэдка кіруюць палётамі** або **Народжаны поўзаць лятаць не можа, затое можа стаць птушкай высокага палёту** або **Народжаны лётаць поўзаць не можа** ⇐ *Народжаны поўзаць лётаць не можа < руск. Рождённый ползатъ – летать не может! – афарызм з верша ў прозе «Несня о соколе» (1895) М. Горкага (А. М. Пешкава, 1868–1936).*

**Пакахайце мяне чорненькую, а беленькую мяне ўсякі пакахае.** **Мулатка** ⇐ *Пакахайце нас чорненькімі, а беленькімі нас ўсякі пакахае < руск. Полюби нас чёрненькими, а беленькими нас всякий полюбит – выслоўе з паэмы «Мёртвые души» (1842, II, 2) М. В. Гоголя (1809–1852).*

**Пры адчыненым акне ў Еўропу сон здаравейшы** ⇐ *У Еўропу прасячы акно < руск. в Европу прорубить окно – выраз з паэмы «Медный всадник» (1833) А. С. Пушкіна (1799–1837).*

**Рукапісы не гараць. А вось ганарары знікаюць бяследна** ⇐ *Рукапісы не гараць < руск. Рукописи не горят – афарызм з рамана «Мастер и Маргарита» (1929–1940) М. А. Булгакава (1891–1940).*

**Усе няшчасці чалавека ад таго, што ён створаны для шчасця** ⇐ *Чалавек народжаны для шчасця, як птушка для палёту < руск. Человек рождён для счастья, как птица для полёта – афарызм з аповядання «Парадокс» (1894) У. Г. Караленкі (1853–1921).*

**Чалавек! – гэта гучыць горда, пакуль не пачынае заплятацца язык** ⇐ *Чалавек – <гэта> гучыць горда < руск. Человек! Это звучит гордо! – афарызм з п'есы «На дне» (1902) М. Горкага (А. М. Пешкава, 1868–1936).*

**Чым менш жанчыну мы кахаем – тым болей часу нам для сну** ⇐ *Чым больш жанчыну мы кахаем, тым спадабацца ёй лягчэй < руск. Чем больше женщину мы любим, тем легче нравимся мы ей – афарызм з рамана ў вершах «Евгений Онегин» (1823–1831, 4, VII) А.С. Пушкіна (1799–1837).*

Найбольш прадуктыўнамі спосабамі трансфармацыі крылатых слоў з'яўляюцца змяненні іх лексічнага складу і сінтаксічнай структуры (на аснове субстытуцыі, інверсіі, элімінацыі і г. д.), разгортванне пэўнага вобраза, далучэнне новага тэксту з мэтай іранічнага або жартаўлівага каментарыя іх зместу, увядзенне (у тым ліку і ў перафразаванай форме) у той ці іншы кантэкст, а таксама камбінацыі названых спосабаў.

## СПІС ЛІТАРАТУРЫ

1 Іванова, С. Ф. Антыцытаты ў сучаснай беларускай мове / С. Ф. Іванова, Я. Я. Іваноў // *Phraseologische Studien: Dynamische Tendenzen in der slawischen Phraseologie / Internationales Slawistenkomitee (Phraseologische Kommission)*; red. D. Balakova und H. Walter. – Greifswald: E.M.A.-Universität, 2010. – S. 174-183.

2 Іванова, С. Ф. Парадыйныя, іранічныя і жартоўныя трансфармацыі крылатых выказаў (да складання слоўніка беларускіх антыпрыказак) / С. Ф. Іванова, Я. Я. Іваноў, В. А. Шпакоўская // *Даследаванні па германскай і славянскай філалогіі = Acta Germano-Slavica: зб. навук. артыкулаў / пад рэд. Я. Я. Іванова.* – Магілёў: УА «МДУ імя А. А. Куляшова», 2011. – Вып. 4. – С. 52–61.

3 Крылатыя выразы ў беларускай мове: тлумачальны слоўнік / пад рэд. С. Ф. Івановай. – Магілёў: МДУ імя А. А. Куляшова, 2006. – Ч. 2: 3 рускамоўных літаратурных і фальклорных крыніц XII–XX стст. – 208 с.

*Научное издание*

# **КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА:**

## **НАУЧНОЕ НАПРАВЛЕНИЕ И УЧЕБНАЯ ДИСЦИПЛИНА**

**Сборник научных статей**

Основан в 2010 году

**Выпуск 2**

Подписано в печать 3.12.2010. Формат 60x84 1/16.  
Бумага офсетная. Ризография. Усл. печ. л. 13,7.  
Уч.-изд. л. 15,0. Тираж 100 экз. Заказ № 614.

Издатель и полиграфическое исполнение:  
учреждение образования  
«Гомельский государственный университет  
имени Франциска Скорины»  
ЛИ № 02330/0549481 от 14.05.2009.  
ЛП № 02330/0150450 от 03.05.2009.  
Ул. Советская, 104, 246019, г. Гомель.