

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**

**Выпуск 1**

**Гомель  
УО «ГГУ им. Ф. Скорины»  
2010**

УДК 81'322(045)  
ББК 81.1Я43  
К 637

Редколлегия: В. И. Коваль (ответств. ред.), Е. В. Ничипорчик,  
И. Г. Гомонова

Рецензенты: А. В. Никитевич, доктор филологических наук,  
профессор, заведующий кафедрой русского языка УО «Гродненский  
государственный университет им. Я. Купалы»; П. Е. Ахраменко,  
кандидат филологических наук, доцент, доцент кафедры русского  
языка УО «Мозырский государственный педагогический университет  
им. И. П. Шамякина»

**Компьютерная лингвистика:** научное направление и учебная  
К 637 дисциплина [Текст]: сборник научных статей. Вып. 1 / В. И. Коваль  
(ответств. ред.) [и др.]; М-во образования РБ, ГГУ им. Ф. Скорины. –  
Гомель: ГГУ им. Ф. Скорины, 2010. – 236 с.  
ISBN 978-985-439-506-7

В сборник включены научные статьи по актуальным вопросам компьютерной лингвистики: корпусной лингвистике, машинном переводе, программном обеспечении лингвистических исследований, прикладной лексикографии, использованию компьютерных технологий в учебно-методической деятельности.

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

УДК 81'322(045)  
ББК 81.1Я43

ISBN 978-985-439-506-7

© УО «Гомельский государственный университет им. Ф. Скорины», 2010

# **КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА: ТЕОРИЯ И ПРАКТИКА. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ, ЕЕ СТАНОВЛЕНИЕ И РАЗВИТИЕ**

А. М. Аматов

## **МАШИННАЯ ОБРАБОТКА ТЕКСТА И ПЕРЕВОД: ПРОБЛЕМЫ И ПОИСК РЕШЕНИЙ**

Об «умных машинах» на протяжении нескольких веков писали многие: от философов-картезианцев и писателей-фантастов до кибернетиков. Непосредственно проблемами машинного перевода лингвисты и техники стали вплотную заниматься с тех самых пор, как появились первые компьютеры.

В 1964 году корпорация РЭНД (Т. Гордона и О. Хелмера) сделала прогноз развития науки и техники на ближайшие 50 лет. Было опрошено большое количество экспертов в разных областях знания, а также несколько писателей-фантастов, в число которых входили, в частности, А. Азимов и А. Кларк. Прогнозы (всего их было 25) касались самых разных областей деятельности человека – от полной автоматизации автотранспорта до лечения наследственных заболеваний путём модификации генов. И одним из первых (под номером 4) значилось использование автоматических переводчиков на иностранные языки. Эксперты решили, что эта идея будет практически осуществлена уже в 1972 году, то есть всего через 8 лет после прогнозирования. Основания для оптимизма понятны – ещё в январе 1954 года состоялась первая публичная демонстрация машинного перевода с русского языка на английский («Джорджтаунский эксперимент»). Сообщение об этом событии было опубликовано в журнале “Computers and Automation”, 1954, № 2. Однако уже осенью того же года в СССР появляется реферат этого сообщения, сделанный Д. Ю. Пановым (РЖ ВИНТИ «Математика», 1954, № 10, «Перевод с одного языка на другой при помощи машины: отчет о первом успешном испытании»). Сам Д. Ю. Панов возглавил работы по машинному переводу, и первый опыт перевода с английского языка на русский с помощью машины БЭСМ был получен уже к концу 1955 г. Параллельно в Отделении прикладной математики Математического института АН СССР по инициативе математика А. А. Ляпунова велись работы по машинному переводу математических текстов. Первые программы машинного перевода, разработанные этим коллективом, были реализованы на машине «Стрела» [1], [2], [3].

С тех пор произошла немыслимая эволюция компьютерной техники и информационных технологий. Сейчас в домах и на рабочих столах у многих стоят компьютеры, превосходящие по быстродействию и объёмам ОЗУ и ПЗУ совокупную мощность всех ЭВМ какого-нибудь крупного

исследовательского института 60-х годов, не говоря уже о мультимедийных свойствах, которые в то время не рождались даже в воображении самых смелых писателей-фантастов (достаточно почитать описание компьютеров у признанного мастера жанра А. Азимова). Однако ничего похожего на стремительное развитие не наблюдается в сфере машинного перевода: у электронных переводчиков год от года улучшаются интерфейсы, пополняются словари, появляются дополнительные функции, но вот в основной своей работе – а это, конечно же, перевод текста – они по-прежнему не могут добиться удовлетворительных результатов.

За прошедшие четыре с половиной десятилетия автоматический перевод превратился в отдельную (и вполне коммерческую) индустрию, однако на вопрос, сбылся ли прогноз корпорации РЭНД, однозначный ответ дать не так просто. С одной стороны, первые опыты по автоматическому переводу были проведены раньше 1972 года (и даже раньше 1964 года – времени составления прогноза), и к настоящему времени имеется большое количество электронных программ-переводчиков, включая онлайн-овые. С другой стороны, несмотря на все усилия в этом направлении, следует признать, что вплоть до сегодняшнего дня в распоряжении пользователей нет компьютерных программ, которые, получив на входе связный текст (даже небольшой длины) на одном языке, могли бы давать на выходе также связный текст на другом языке.

Чтобы убедиться в этом, достаточно провести простой эксперимент: введём в программу-переводчик (самую современную, со всеми доступными обновлениями) небольшой текст на русском языке, переведём его с помощью программы на английский, а затем результат при помощи той же программы переведём обратно на русский. В качестве примера возьмём первый абзац данной статьи: (то есть текст без сложных художественных оборотов) и применим к нему вышеописанную процедуру (использовалась онлайн-овая программа-переводчик Google <http://translate.google.com/>). Перевод абзаца на английский язык: *On the “smart cars” for several centuries have written many: from the Cartesian philosophers and science fiction writers to cybernetics. Immediate problems of machine translation linguistics and technology have become closely involved ever since, appeared the first computers.*

А вот перевод полученного текста обратно на русский язык: *На “умные машины” в течение нескольких веков писали много: от декартовых философов и писателей-фантастов с кибернетикой. Срочные проблемы машинного перевода лингвистики и технологии стали активно участвовать с тех пор, как появился первый компьютер.*

Как видим, английский перевод получился неадекватным, а повторный русский перевод вообще имеет очень отдалённое сходство с оригиналом, да и связным текстом его назвать нельзя. Если лексические ошибки (*машины => cars, непосредственно* (наречие!) => *immediate* (прилагательное!) => *срочные* и т. п.) уже затрудняют понимание текста, то грамматические несоответствия делают его просто нечитаемым.

В этой связи есть основания считать, что столь скромные успехи в разработке программного обеспечения для перевода текстов не связаны с плохой работой программистов или компьютерных техников, а являются результатом плохой проработки этой проблемы с лингвистической стороны. Ведь программисты наверняка смогли бы написать хорошую программу для перевода (они же пишут хорошие программы для решения самых разных научных и практических задач), если бы они чётко знали, *на что именно* нужно запрограммировать компьютер. Но для этого надо, чтобы система языка (точнее, как минимум двух языков) была формализована настолько, чтобы её можно было представить в виде последовательности команд, которые необходимо выполнить для получения лексически и грамматически верного высказывания, а затем эти высказывания можно было бы связать в цельный текст.

Понятно, что одного знания слов для выражения мысли явно недостаточно, ведь любое высказывание представляет собой упорядоченную структуру, и значение высказывания не сводится к семантике входящих в него слов. Соответственно, к словарю необходимо подходить так, как он того заслуживает, а именно – как к совокупности единиц, обладающих номинативным (назывным) значением. И в этом отношении электронным переводчикам нужно отдать должное – с функциями электронного словаря они справляются очень хорошо.

Грамматика изучает строение языка как системы, в её «юрисдикции» находятся, в сущности, все отношения и связи, которыми эта система характеризуется. Соответственно, сами эти связи и отношения часто называют грамматикой того или иного языка. Другими словами, помимо звукового строя (фонетика) и значения корней знаменательных слов (лексика) всё остальное в языке – грамматика. Даже такие аспекты, как просодика (интонация, ударения, паузы) и значения некорневых элементов слова лежат в сфере интересов грамматики. И именно эта сфера остаётся наименее проработанной применительно к автоматическому анализу текста и машинному переводу.

В свете представлений о лексике и грамматике рассмотрим вопрос о выражении мысли средствами языка – при переводе именно это имеет первостепенное значение. При взгляде на уровни языковой структуры становится понятно, что законченная мысль может выражаться единицами не ниже предложения. Основная функциональная особенность предложения по сравнению с нижележащими единицами языка – предикативность, т. е. его способность соотносить своё содержание с действительностью, а без этого мысль выразить невозможно. Соответственно, вопросом особой важности при переводе высказывания с одного языка на другой становится установление предикативной связи в предложении-оригинале и нахождение её соответствия в языке перевода. Если это сделать удастся, то перевод в общих чертах можно считать состоявшимся.

В этой связи наибольшую сложность и важность представляет собой синтаксическая категоризация глаголов, поскольку именно глагол выполняет основную синтаксическую функцию, входя в ядро предиката. Для глаголов такие количественные категории определяют количество и состав актантов, взаимодействуя с которыми он образует синтаксические структуры. При присвоении глаголу той или иной синтаксической категории учитываются трансформационные возможности предиката, но не его внутренняя семантика. При таком подходе получается, что один и тот же глагол обладает некоторым набором синтаксических категорий, образующих парадигму и реализуемых в конкретном типе синтаксической конструкции. Наличие у глагола ряда количественных категорий не может представлять особой проблемы (например, стоит ли рассматривать эти единицы как один глагол или как разные). Ведь, в конце концов, наличие у одной и той же единицы ряда лексико-семантических вариантов ни у кого нареканий не вызывает.

При таком подходе процедура анализа предложения строится по следующему алгоритму:

1. Поиск предикативного ядра (глагол).
2. Выбор первой количественной категории глагола.
3. Построение модели предложения на основе данной категории.
4. Проверка соответствий остальных лексем выбранной модели.
5. При отсутствии соответствия – выбор следующей количественной категории глагола и переход к шагу 3.

Для перевода с одного языка на другой после завершения данной процедуры выполняются следующие операции:

6. Выбор из языка перевода глагольной лексемы, соответствующей по значению глаголу данной количественной категории в языке оригинала.
7. Построение модели предложения в языке перевода на основе количественной категории выбранного глагола.
8. Подбор значений лексем, входящих в синтаксическую модель.

Разумеется, этот алгоритм потребует большого количества переборов разных вариантов – как значений, так и, что сложнее, моделей построения предложений. Однако для современных компьютеров большое количество вычислений едва ли будет представлять слишком сложную задачу. Главная проблема – лингвистическая, и заключается она в создании подобия словаря, в котором вместо значений лексем будут заданы их количественные категории.

Подводя итог, следует отметить, что проблемы создания программ-переводчиков отнюдь не исчерпываются одной лишь категоризацией лексики, да и вообще в этой отрасли есть целый ряд вопросов, не связанных напрямую с лингвистикой. Однако разработка алгоритма порождения и интерпретации высказывания, позволяющего преодолеть её негативные последствия, представляются важным шагом на пути создания «умной машины». Началом же разработки такого алгоритма вполне может стать

формализация естественного языка через выявление набора дискретных грамматических категорий, присущих каждому элементу лексикона того или иного языка.

#### СПИСОК ЛИТЕРАТУРЫ

1 Кузнецов, П. С. Основные проблемы машинного перевода / П. С. Кузнецов, А. А. Ляпунов, А. А. Реформатский // Вопросы языкознания. – 1956. – № 5. – С. 48–54.

2 Ляпунов, А. А. Использование вычислительных машин для перевода с одного языка на другой / А. А. Ляпунов, О. С. Кулагина // Природа. – 1955. – № 8. – С. 18–22.

3 Панов, Д. Ю. Автоматизация перевода с одного языка на другой / Д. Ю. Панов, А. А. Ляпунов, И. С. Мухин // Сессия по научным проблемам автоматизации производства. – М. : Изд. АН СССР. – 1956.

А. В. Зубов

### КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА В БЕЛАРУСИ: ЭВОЛЮЦИЯ ЗАДАЧ

Впервые термин «компьютерная лингвистика» (Computation Linguistics) появился в начале 60-х годов XX столетия в США, где в 1962 году была создана «Ассоциация машинного перевода и компьютерной лингвистики». Русский термин «компьютерная лингвистика» вошел в активный обиход взамен существовавшего ранее термина «вычислительная лингвистика» в 1989 году, когда почти одновременно вышли из печати две публикации [1; 2].

Как видно, с того времени прошло уже много лет. Однако до сих пор нет единого общепринятого определения понятия «компьютерная лингвистика». Её задачи пересекаются с задачами, решаемыми в таких направлениях лингвистики как «вычислительная лингвистика», «инженерная лингвистика», «математическая лингвистика», «лингвистические основы информатики», «прикладная лингвистика». В известной работе «Основы компьютерной лингвистики» её автор, Ю. Н. Марчук, отмечает, что компьютерная лингвистика изучает все «аспекты связи языка, мышления как непосредственной действительности, и моделирования этой действительности с помощью компьютерных программ» [3, с. 23]. О сложности точного определения понятия «компьютерная лингвистика» говорит, например, тот факт, что в относительно недавно вышедшей книге Санкт-Петербургского ученого С. А. Коваля [4] треть книги (около 50 стр.) посвящено трактовке этого понятия и доказательствам того, чем компьютерная лингвистика отличается от прикладной лингвистики, инженерной лингвистики, алгоритмической лингвистики, корпусной лингвистики.

Компьютерная лингвистика возникла в результате взаимодействия различных научных дисциплин и социальных потребностей. Профессор Р. Г. Пиотровский выделяет пять таких факторов [5]:

1. Потребности самого языкознания. Расширение и углубление лингвистических исследований «настоятельно потребовало создания надежной экспериментальной методики, которая давала бы возможность не только собирать и трактовать непосредственно наблюдаемые лингвистические факты, но также исследовать глубинные, скрытые от прямого наблюдения лингвистические объекты и связи».

2. Стремление математиков и программистов использовать языковой материал для приложения и экспликации новых идей.

3. Всё время усиливающиеся контакты кибернетики с семиотическими и системными исследованиями, направленными на создание систем искусственного интеллекта, имитирующих речемыслительную деятельность человека.

4. Запросы информационно-вычислительной индустрии и потребности автоматизации управления промышленностью, транспортом, войсками.

5. Необходимость оптимизации и повышения эффективности процесса обучения языкам.

Анализ задач, решаемых компьютерной лингвистикой в уже упомянутых работах, а также в большом числе последующих публикаций, позволяет сделать вывод о том, что компьютерная лингвистика – это направление в лингвистике, связанное с получением *новой* информации о тексте, речи и языке в целом с использованием компьютеров, математических методов и методов информационного моделирования.

Сюда можно отнести следующие задачи:

1. Автоматическое построение словарей слов и словосочетаний.
2. Создание алгоритмов и программ автоматического анализа слов, словосочетаний и предложений.
3. Разработка методов атрибуции текстов и речи, методов распознавания стилей и жанров текстов и речи.
4. Создание компьютерных систем автоматического реферирования и аннотирования текстов.
5. Создание компьютерных систем перевода текстов с одного языка на другой.
6. Создание компьютерных систем понимания текста и речи.
7. Создание компьютерных систем генерации текстов и речи.
8. Создание систем, моделирующих речевую деятельность человека – компьютерных диалоговых систем.

В Беларуси началом работ по компьютерной лингвистике можно считать 1965 год. Именно тогда на основании решения Президиума Академии Наук СССР был направлен на кафедру общего языкознания Минского государственного педагогического института иностранных языков (МГПИИЯ) «для чтения лекций, оказания помощи в проведении научных



работ и подготовки научных кадров» доктор филологических наук из Ленинграда Р. Г. Пиотровский. В том же году аспирантами на эту кафедру по специальности «структурная, прикладная и математическая лингвистика» были рекомендованы два математика-программиста, окончившие МГУ им. М. В. Ломоносова и Минский государственный педагогический институт им. М. Горького. Они в сотрудничестве с преподавателями-лингвистами МГПИИЯ и другими уже опытными программистами г. Минска стали проводить на компьютере «Минск-22», установленном в МГПИИЯ, первые работы по автоматическому составлению частотных словарей по текстам различных подязыков и частотных списков словосочетаний. Уже в 1969 году в МГПИИЯ был выпущен сборник научных работ «Статистика текста. Том 1», где были опубликованы 15 статей по результатам автоматического создания частотных списков слов и словосочетаний различных языков.

В последующих научных сборниках, выпущенных в МГЛУ [6; 7; 8; 9; 10; 11; 12; 13; 14; 15], задачи использования компьютера в лингвистических исследованиях все более усложнялись. Так, последовательно решались такие задачи:

1970-й год: автоматический перевод; автоматическая сегментация английских и немецких предложений; синтез видовых форм русского глагола;

1972-й год: автоматический синтаксический анализ; автоматическое устранение лексической неоднозначности;

1977-й год: формализация содержания англоязычных текстов технических описаний; формализация структуры рекламного текста и его порождение.

Аналогичные исследования проводились кафедрой информатики и прикладной лингвистики МГПИИЯ и на хоздоговорных условиях по заказу различных городов Советского Союза (Минска, Ленинграда, Свердловска и др.). Например, такие [16; 17; 18; 19; 20].

В эти же годы на кафедре экспериментальной фонетики МГПИИЯ (ныне – МГЛУ) успешно работала над проблемами компьютерной лингвистики, связанными с синтезом устной речи, кандидат филологических наук, профессор Елена Борисовна Корневская. Ее работы в соавторстве с сотрудником Института технической кибернетики Академии наук Беларуси, доктором технических наук Борисом Мефодиевичем Лобановым, позволили создать синтезатор «ФОНЕМОФОН-4», а в последующем – ряд других систем синтеза речи [21].

С 1966 года активно работала в Минске научно-исследовательская группа по проблемам искусственного интеллекта. Возглавлял ее доктор филологических наук, профессор Виктор Владимирович Мартынов, который успешно занимался проблемами формализации семантики текстов в целях их автоматической обработки. Теоретическими аспектами формализации семантики для создания технологий проектирования

интеллектуальных систем все эти годы занимается кафедра искусственного интеллекта Белорусского государственного университета информатики и радиоэлектроники, руководимая доктором технических наук, профессором Владимиром Васильевичем Голенковым.

Решая различные задачи, связанные с использованием методов компьютерной лингвистики, кафедра информатики и прикладной лингвистики МГЛУ активно сотрудничала с профессором кафедры прикладной лингвистики Белорусского государственного университета Владимиром Александровича Карповым и кандидатом филологических наук, доцентом этой же кафедры Анастасией Ивановной Головня.

По проблемам создания машинного фонда белорусского языка и корпуса текстов белорусского языка кафедра информатики и прикладной лингвистики МГЛУ постоянно сотрудничает с кандидатом филологических наук, доцентом, заведующей кафедрой общего и славянского языкознания Гродненского государственного университета Людмилой Васильевной Рычковой.

В наши дни в МГЛУ ежегодно выполняются госбюджетные научно-исследовательские работы по проблемам компьютерной лингвистики. Одно из основных направлений научных работ кафедры в последние годы связано с проблемами корпусной лингвистики. Созданы два параллельных (англо-русский и франко-русский) тегированные корпуса текстов, включающие тексты по лингвистике (автоматическая обработка текстов), поэзии и деловой документации. Создается совместно с Институтом языка и литературы имени Якуба Колоса и Янки Купалы Национальной Академии наук Республики Беларусь большой корпус текстов белорусского языка, который включает 1 млн. словоупотреблений белорусских текстов и по 200 000 словоупотреблений русско-белорусских, англо-белорусских и немецко-белорусских текстов. В связи с тем, что некоторые дисциплины в школах и спецкурсы в вузах читаются на русском и белорусском языках, кафедра информатики и прикладной лингвистики с 2008 года начала работать над созданием русско-белорусского корпуса учебных текстов.

Второе направление исследований кафедры информатики и прикладной лингвистики МГЛУ в области компьютерной лингвистики связано с моделированием на компьютере творческой деятельности человека. Созданы алгоритмы и программы, позволяющие автоматически создавать англоязычные деловые письма, рекламные тексты, сказки, пословицы и поговорки на английском и французском языках. Создана также компьютерная программа, моделирующая стихотворные тексты, которые использовал в качестве песен известный французский певец Шарль Азнавур. В настоящее время кафедра работает над созданием алгоритмов и программ, моделирующих стихи С. Есенина и Джорджа Байрона.

## СПИСОК ЛИТЕРАТУРЫ

- 1 Городецкий, Б. Ю. Компьютерная лингвистика и моделирование языкового общения / Б. Ю. Городецкий // Новое в зарубежной лингвистике. Вып. XXIV. Компьютерная лингвистика. – М., 1989.
- 2 Белоногов, Г. Г. Компьютерная лингвистика в системе НТИ / Г. Г. Белоногов, Ю. Г. Зеленков // Вопросы информационной теории и практики. Сб. научных статей. – М., 1989. – № 58.
- 3 Марчук, Ю. Н. Основы компьютерной лингвистики / Ю. Н. Марчук. – М. : МПИ, 2000. – 226 с.
- 4 Коваль, С. А. Лингвистические проблемы компьютерной морфологии / С. А. Коваль. – Санкт-Петербург: СПбГУ, 2005. – 147 с.
- 5 Пиотровский, Р. Г. Инженерная лингвистика и теория языка. / Р. Г. Пиотровский. – Л. : Наука, 1979. – 112 с.
- 6 Статистика текста. Том II. Автоматическая переработка текста: сб. науч. ст. / МГПИИЯ; редкол.: А. И. Киселевский (отв. ред.) [и др.]. – Минск, 1970. – 652 с.
- 7 Частные вопросы автоматического анализа текстов: сб. науч. ст. / МГПИИЯ; редкол.: А.В. Зубов (отв. ред.) [и др.]. – Минск, 1972. – 395 с.
- 8 Лингвостатистика и автоматический анализ текстов: сб. науч. ст. / МГПИИЯ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 1973. – 458 с.
- 9 Вопросы общей и прикладной лингвистики: сб. науч. ст. / МГПИИЯ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 1975. – 297 с.
- 10 Методы анализа текстов: сб. науч. ст. / МГПИИЯ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 1975. – 226 с.
- 11 Автоматический анализ текстов: сб. науч. ст. / МГПИИЯ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 1976. – 198 с.
- 12 Лингвистическое моделирование коммуникативных систем: сб. науч. ст. / МГПИИЯ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 1983. – 263 с.
- 13 Проблемы компьютерной лингвистики: сб. науч. ст. / МГЛУ; редкол.: Р. Г. Пиотровский (отв. ред.) [и др.]. – Минск, 1997. – 178 с.
- 14 Компьютерная лингвистика и обучение языку: сб. науч. ст. / МГЛУ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 2000. – 219 с.
- 15 Актуальные проблемы компьютерной лингвистики: сб. науч. ст. / МГЛУ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 2005. – 318 с.
- 16 Автоматический семантико-синтаксический анализ специальных текстов: отчет о НИР (закл.) / Минск. гос. пед. инст. ин. языков; рук. темы Л. Е. Машкина. – Минск., 1977. – 52 с.
- 17 Разработка лингвистического обеспечения закрытых естественных языков («земля-воздух») и построение модели функционирования такого языка в условиях ч/м диалога: отчет о НИР / Минск. гос. пед. инст. ин. языков; рук. темы О. А. Нехай. – Минск., 1978. – 102 с.
- 18 Лингвистическое обеспечение информационно-поисковых систем: отчет о НИР / Минск. гос. пед. инст. ин. языков; рук. темы А.В. Зубов. – Минск., 1980. – 30 с.
- 19 Создание алгоритмов и программ функционирования русско-английского, англо-русского, русско-французского, французско-русского словарей в памяти ЭВМ малого объема: отчет о НИР / Минск. гос. пед. инст. ин. языков; рук. темы В. Н. Билан. – Минск, 1981. – 442 с.
- 20 Алгоритмизация процедуры перевода немецких, французских, итальянских, испанских, английских и русских реплик разговорно-бытового характера: отчет о НИР / Минск. гос. пед. инст. ин. языков; рук. темы О. А. Нехай. – Минск., 1983. – 228 с.

21 Корневская, Е. Б. Модели синтеза мелодического контура русских и английских фраз / Е. Б. Корневская, Б. М. Лобанов // Автоматическое распознавание слуховых образов. – Киев, 1982.

22 Мартынов, В. В. Кибернетика. Семиотика. Лингвистика / В. В. Мартынов. – Минск : Наука и техника, 1966. – 146 с.

І. Л. Капылоў, У. А. Кошчанка

## **КОРПУС ТЭКСТАЎ ЯК АСНОВА АПТЫМІЗАЦЫІ ПАДРЫХОТЎКІ ФУНДАМЕНТАЛЬНАГА “ТЛУМАЧАЛЬНАГА СЛОЎНІКА БЕЛАРУСКАЙ МОВЫ”**

У межах дзяржаўнай падпраграмы навуковых даследаванняў “Беларуская мова і літаратура ў кантэксце цывілізацыйнага развіцця Рэспублікі Беларусь: гісторыя, сучасны стан, тэндэнцыі” на 2011–2015 гады запланавана заданне “Распрацоўка тэарэтычных асноў, фарміраванне лексікаграфічнай базы і стварэнне аўтаматычнага аналізатара тэкстаў для новага фундаментальнага тлумачальнага слоўніка беларускай мовы і зводнага слоўніка беларускіх народных гаворак”. Перад лексікаграфамі Інстытута мовы і літаратуры імя Якуба Коласа і Янкі Купалы НАН Беларусі стаіць надзвычай складаная задача распрацаваць канцэпцыю і асноўныя тэарэтычныя палажэнні ўкладання новага фундаментальнага акадэмічнага тлумачальнага слоўніка. Для вырашэння дадзенай задачы неабходна ў першую чаргу канчаткова вызначыцца з храналагічнымі межамі сучаснай беларускай мовы і адпаведна – з крытэрыямі адбору крыніц, на аснове якіх будзе сфарміравана лексікаграфічная база тлумачальнага слоўніка. Існыя на сённяшні дзень акадэмічныя тлумачальныя слоўнікі (“Тлумачальны слоўнік беларускай мовы” ў 5-ці тамах шасці кнігах пад рэдакцыяй К. К. Атраховіча (Крапівы) (1977–1984 гг., рэстр – 93 500 слоў); “Тлумачальны слоўнік беларускай літаратурнай мовы” пад рэдакцыяй М. Р. Судніка і М. Н. Крыўко) (1996 г.), фіксуюць лексіку адносна невялікага храналагічнага перыяду – лексіку выданняў, апублікаваных з 1951 па 1996 год. За межамі слоўнікаў засталіся гістарычна знакавыя перыяды ў гісторыі беларускай мовы: у прыватнасці, перыяд выдання газеты “Наша Ніва”, на старонках якой, выпрацоўваліся арфаграфічныя, лексічныя і граматычныя нормы новай беларускай літаратурнай мовы, фарміраваўся яе фразеалагічны фонд, адбывалася станаўленне функцыянальных стыляў. У перыяд з 1996 года да сённяшніх дзён у лексічнай сістэме беларускай мовы адбыліся значныя змены, абумоўленыя як інтралінгвістычнымі, так і экстралінгвістычнымі фактарамі. На жаль, паслядоўны, сістэматычны ўлік новых лексічных адзінак і аналіз стану лексіка-семантычнай сістэмы сучаснай беларускай мовы з канца 80-х гадоў XX стагоддзя па розных прычынах амаль не праводзіўся, за выключэннем дысертацыйнага даследавання В. І. Уласевіч (Лексічныя працэсы ў беларускай мове 90-х гадоў XX ст.” (Мінск, 2002) і

“Слоўніка новых слоў беларускай мовы” В. І. Уласевіч, Н. М. Даўгулевіч. Новыя лінгвістычныя слоўнікі, што з’явіліся ў апошнія гады, у большасці сваёй абапіраюцца на лексічную базу “Слоўніка беларускай мовы” пад рэдакцыяй акадэміка М.В. Бірылы (1987 г.) і ў поўнай меры не адлюстроўваюць лексічныя змены беларускай мовы. У першую чаргу гэта абумоўлена тым, што кожны асобны даследчык або ўкладальнік не ў стане ахапіць і прааналізаваць вялікія аб’ёмы інфармацыі без выкарыстання сучасных камп’ютарных тэхналогій. У такой сітуацыі адзіным магчымым выйсцем з’яўляецца стварэнне Нацыянальнага корпусу беларускай мовы, які б уключаў у сябе тэксты рознай стылёвай і жанравай прыналежнасці і ахопліваў бы ўсе перыяды новай беларускай літаратурнай мовы. Корпус тэкстаў павінен стаць асновай працы над тлумачальным слоўнікам.

Папярэдне плануецца, што аб’ём тлумачальнага слоўніка ў друкаванай версіі будзе складаць прыкладна 15 тамоў у традыцыйным выдавецкім фармаце. “Сучасны агульнамоўны тлумачальны слоўнік, – на думку В. П. Лемцюговай, – павінен арыентавацца на як мага паўнейшае адлюстраванне жывога, актуальнага, рэальна існуючага інвентару. У ім павінен быць задзейнічаны ўвесь вопыт, усе рэсурсы і набыткі ў галіне мовы розных пакаленняў” [1, с. 113]. Слоўнік павінен адпавядаць на шырокім запатрабаваннем сучаснага чытача, у ім павінна быць зафіксавана як мага поўна актуальная лексіка. Корпус тэкстаў дасць магчымасць аб’ектыўна паказаць лексічныя рэсурсы мовы, у тым ліку той пласт новых лексічных адзінак, які сфарміраваўся ў беларускай мове ў апошнія дзесяцігоддзі ў выніку працэсаў дэархаізацыі, неалагізацыі і запазычання з іншых моў.

Выклікае цікавасць і з’яўляецца надзвычай карысным для беларускіх укладальнікаў слоўнікаў вопыт сербскіх лексікографаў. На думку аўтараў “Слоўніка сербскахарвацкай літаратурнай і народнай мовы”, адным з асноўных прынцыпаў пры стварэнні такога тыпу лексікаграфічных крыніц павінен стаць прынцып, згодна з якім укладальнікі слоўніка не павінны дапускаць у яго “толькі тыя словы, якія, на іх думку, маюць права на існаванне ў сучаснай літаратурнай мове. Паколькі такі падыход укладальнікаў быў бы індывідуальным і залежаў ад іх ведаў, густу і разумення” [2, с. XXIII]. Такі падыход аўтары слоўніка тлумачылі ў першую чаргу тым, што іх літаратурная мова знаходзіцца “ўсё яшчэ ў стане станаўлення і таму пакуль нельга сцвярджаць, якія словы маюць выключнае права на прысутнасць.” Па словах укладальнікаў “Слоўніка сербскахарвацкай літаратурнай і народнай мовы”, “слоўнікі і граматыкі не ствараюць літаратурную мову, а абавязак укладальнікаў – апісаць літаратурную мову і прафесійна паказаць, наколькі новыя набыткі знаходзяцца а ў сугуччы з духам народнай мовы” [3, с. XXIII].

Нявырашаным застаецца пытанне адлюстравання дыялектызмаў у тлумачальным слоўніку. Традыцыйна дыялектная лексіка не ўключаецца ў такога тыпу лексікаграфічныя крыніцы. Разам з тым, рэгіянальная лексіка часта служыць рэзервам літаратурнай мовы: шырока выкарыстоўваецца ў

творах мастацкай літаратуры, прычым не толькі ў мове персанажаў, але і ў аўтарскай мове.

У сучаснай лексікаграфіі не існуе адзіных падыходаў і патрабаванняў да апісання сінкрэтычных і варыятыўных фразеалагічных выразаў.

У новым слоўніку будзе прадстаўлена спецыяльная лексіка (тэрміны, наменклатура, прафесіяналізмы) розных тэрміналагічных сістэм. Адпаведна павінны быць распрацаваны крытэрыі іх адбору.

Пры распрацоўцы слоўнікавага артыкула аўтары будуць арыентавацца на сучасную метадалогію і ўлічваць напрацоўкі, якімі валодае сучасная лексікалагічная тэорыя і лексікаграфічная практыка. Асабліва важнай з'яўляецца задача семантычнага апісання лексікі, якая прадугледжвае дастатковы поўны ахоп лексічных значэнняў слоў з мэтай паказу намінатыўных і экспрэсіўных магчымасцей мовы, сістэмных сувязей лексічных адзінак, выяўленне этапаў фарміравання лексічных значэнняў слова і сэнсавай структуры слова ў цэлам.

Граматычнае апісанне кожнага слова будзе абапірацца на дасягненні сучаснай тэорыі граматыкі.

Новыя навукова абгрунтаваныя прынцыпы павінны прымяняцца пры адборы і прэзентацыі ілюстрацыйнага матэрыялу да слоўнікавых артыкулаў. Пры гэтым, як слушна адзначае В. П. Лемцюгова, “трэба не проста падаць слова ў кантэксце, а падабраць такія цытаты, якія б правільна раскрывалі змест лексічнай адзінкі, давалі б уяўленне пра стылістычныя і граматычныя асаблівасці слова, паказвалі б мажлівасці яго вобразнага ці пераноснага ўжывання, спалучальнасць з іншымі словамі”. [2, с. 194, 195].

Кардынальнага перагляду патрабуе сістэма памет у тлумачальным слоўніку.

Такім чынам, галоўная задача фундаментальнага тлумачальнага слоўніка новага тыпу заключаецца не толькі ў тым, каб адлюстравалі ўсе наяўныя ў мове словы і, абапіраючыся на кантэкст, даць іх значэнне і паказаць фармальна-граматычныя, словаўтваральныя, стылістычныя ўласцівасці слоўнікавага фонду сучаснай беларускай мовы.

На сучасным этапе развіцця лінгвістыкі і камп'ютарных тэхналогій немагчыма праводзіць навуковыя даследаванні, ствараць граматыкі і складаць слоўнікі без выкарыстання корпусаў, паколькі пісьменна сфарміраваны корпус не толькі ў значнай ступені садзейнічае аб'ектыўнай і змястоўнай распрацоўцы слоўнікавых артыкулаў, але з'яўляецца аб'ектыўным адлюстраваннем сучаснага стану і асаблівасцей функцыянавання мовы.

Пры падрыхтоўцы новага тлумачальнага слоўніка неабходна ўлічыць еўрапейскі вопыт, паколькі многія аднамоўныя і перакладныя слоўнікі еўрапейскіх моў грунтуюцца на базе корпусаў. Напрыклад, англамоўныя *The Oxford English Corpus* (аб'ём слоўніка – 2 мільярды слоў), *COBUILD Corpus*, *Longman Lancaster English Corpus*, *British National Corpus* і іншыя.

Ёсць значныя напрацоўкі і ў славянскай лексікаграфіі: *Korpus Języka Polskiego Wydawnictwa Naukowego PWN* (аб’ём – 40 мільёнаў слоў) і інш. Таксама варта згадаць *Нацыянальны корпус рускай мовы*, на падставе, якога створаны шэраг слоўнікаў (<http://www.dict.ruslang.ru>).

Пры ўсіх станоўчых баках выкарыстання корпусаў пры ўкладанні слоўнікаў сітуацыя ў гэтай сферы застаецца неадназначная. Як адзначае аўтар манаграфіі “*Lexicography of English*”, “пры ўкладанні слоўнікаў многіх моў цяпер выкарыстоўваюцца электронныя корпусы. Але некаторыя слоўнікі грунтуюцца на невялікіх корпусах, створаных для вузкіх патрэб, або ўкладаюцца ўвогуле без аніякіх корпусаў з той прычыны, што продажы гэтых слоўнікаў не пакрыюць расходы на стварэнне і эксплуатацыю корпусу. Англамоўная супольнасць больш багатая, бо прадае больш слоўнікаў на англійскай мове, чым іншыя моўныя супольнасці на сваіх мовах. Але гэта хутчэй выключэнне, а не правіла. Нават у англамоўным свеце розныя выдаўцы праяўляюць розную ступень зацікаўленасці [да стварэння корпусаў]: брытанскія больш за амерыканскіх, а выдаўцы навучальных слоўнікаў больш за выдаўцоў іншых кніг [4, с. 351]. Гэта заўвага цалкам справядлівая і ў дачыненні да беларускай сітуацыі.

На сённяшні дзень у Беларусі няма корпусаў, якія б адпавядалі патрэбам будучага тлумачальнага слоўніка. Існуе некалькі корпусных праектаў, якія могуць стаць асновай для будучага Нацыянальнага корпусу: Вялікі корпус беларускай мовы (Мінскі дзяржаўны лінгвістычны ўніверсітэт) і корпус навуковых тэкстаў беларускай мовы *Corpus Albaruthenicum* (Беларускі нацыянальны тэхнічны ўніверсітэт). Абодва корпусы ствараюцца пры ўдзеле спецыялістаў з Інстытута мовы і літаратуры НАН.

*Вялікі корпус беларускай мовы* (далей *Корпус*) у сучасным фармаце складаецца з уласна Корпусу беларускай мовы і некалькіх паралельных корпусаў (англа-беларускага, нямецка-беларускага, руска-беларускага), выкананых ў адзіным стандарце.

Плануецца, што кожны з паралельных корпусаў будзе ўтрымліваць прыблізна 300 000 словаўжыванняў. Асноўны корпус на першым этапе (да 2010 г.) будзе мець 1 млн. словаўжыванняў. Такі абмежаваны аб’ём корпусу запланаваны адмыслова, паколькі існуе неабходнасць вырашыць спрэчныя пытанні, звязаныя з пашпартызацыяй, граматычнай разметкай і г. д. Тэксты ў *Корпусе* змяшчаюцца цалкам, а не ў выглядзе фрагментаў.

*Corpus Albaruthenicum* выконваўся ў рамках еўрапейскага праекта “*BalticGrid-II*”. На сённяшні дзень аб’ём корпусу складае каля 400000 слоў са знятай аманіміяй і ўключае тэксты з розных галін мовазнаўства, літаратуразнаўства, эканомікі, фізікі, хіміі, генетыкі, медыцыны і г.д. Для яго быў распрацаваны анлайнавы сэрвіс, які прадастаўляе выбаркі інфармацыі з Корпусу. Утыліта дазваляе шукаць словы ў іх зыходнай форме (удрукаванай карыстальнікам), словы з іх словаформамі і камбінацыі слоў (уклучна з іх словаформамі), выкарыстоўваючы лагічны аператар

AND і ўлічваючы адлегласць паміж словамі. Пошук дазваляецца ў межах сказа, абзаца і ўсяго тэксту. Сэрвіс даступны па адрасе <http://www.grid.bntu.by/corpus/>.

У дадзены момант вядзецца праца па ўдасканаленні сэрвісу, будуць уключаны новыя магчымасці пошуку.

Корпусы маюць падобную схему разметкі і змяшчаюць наступныя характарыстыкі:

- агульную інфармацыю пра тэкст;
- базавую структурную анатацыю тэксту: главы, загаловкі, змесціва тэксту да ўзроўню асобных слоў;
- граматычную інфармацыю.

Абодва корпусы размячаліся пры дапамозе *Лексіка-граматычнай базы беларускай мовы*, што распрацоўваецца ў Інстытуце мовы і літаратуры НАН Беларусі.

Існуе папярэдняя дамоўленасць стваральнікаў абодвух корпусаў наконт іх аб'яднання ў адзін, што і стане асновай для будучага корпусу, які будзе выкарыстоўвацца пры ўкладанні “Глумачальнага слоўніка беларускай мовы”.

Першая задача, якая паўстане перад укладальнікамі – гэта стварэнне электроннай базы тэкстаў. На сённяшні дзень існуючая электронная база тэкстаў утрымлівае мастацкія, публіцыстычныя, навуковыя і газетныя тэксты агульным аб'ёмам 20 млн. слоў, але гэтага недастаткова і трэба працягваць яе папаўненне не толькі за кошт апублікаваных тэкстаў, але ўключаць і неапублікаваныя тэксты (блогі, форумы і інш.), а таксама прыклады вуснага маўлення з радыё і тэлебачання. Неабходнай умовай павінна стаць абавязковая перадача выдавецтвамі беларускамоўных тэкстаў для Нацыянальнага корпусу.

Таксама ўкладальнікі ўлічваюць той факт, што класічныя слоўнікі ў папяровым фармаце маюць істотныя абмежаванні і не ў стане даць поўны ілюстрацыйны матэрыял у слоўнікавым артыкуле. Таму плануецца паралельна рыхтаваць больш поўны электронны варыянт слоўніка. Асновай для ўкладання Корпусу а таксама папяровай і электроннай версій “Глумачальнага слоўніка беларускай мовы” павінна стаць *Лексіка-граматычная база беларускай мовы*.

На сённяшні дзень ЛГБ складаецца з граматычнага (уключна са слоўнікам кіравання), марфемнага, словаўтваральнага, арфаэпічнага і глумачальнага слоўнікаў. Кожны слоўнік уяўляе сабой асобную базу даных.

Граматычны слоўнік мае прыблізна 130 000 парадыгм (1 958 000 словаформаў). У слоўніку адлюстроўваецца інфармацыя пра націск, часткова – кіраванне і граматычныя прыкметы лексем. Скланавыя формы маюць таксама падрабязнае апісанне.

Марфемны слоўнік мае 62 000 слоў і складзены такім чынам, што адлюстроўвае не толькі сінхранію, але і дыяхранію. Словаўтваральны слоўнік таксама мае 25 000 слоў і змяшчае інфармацыю пра



вытворнасць/невывтворнасць, утваральную аснову, словаўтваральныя сродкі, змены, якія адбываюцца пры ўтварэнні новага слова.

Спалучэнне марфемнага і словаўтваральнага слоўнікаў дазваляе аўтаматычна генераваць словаўтваральныя ланцужкі і гнёзды, размяркоўваць іх па групам у адпаведнасці з іх этымалогіяй, тыпам змянення і інш. Такі завершаны электронны слоўнік можа стаць крыніцай для даследавання ў галіне марфемікі, словаўтварэння, семантыкі.

Арфаэпічны слоўнік уключае 60 000 слоў. У аснову фанетычнай транскрыпцыі пакладзены ўніверсальны *Міжнародны фанетычны алфавіт* (<http://www.langsci.ucl.ac.uk/ipa/>)

Такім чынам, фундаментальны акадэмічны тлумачальны слоўнік беларускай мовы па сваёй канцэпцыі і структуры запланаваны як лексікаграфічная крыніца новага пакалення, а асновай яго аптымізацыі, папаўнення і ўдасканалення павінен стаць Нацыянальны корпус беларускай мовы.

#### СПІС ЛІТАРАТУРЫ

1 Лемцюгова, В. П. Беларуская лексікаграфія. Перспектывы і далейшы прагрэс / В. П. Лемцюгова // Сучасныя праблемы беларускага мовазнаўства. матэрыялы навуковай канферэнцыі, прысвечанай 75-годдзю Інстытута мовазнаўства імя Якуба Коласа НАН Беларусі, 2 лістапада 2004 г. Мінск : Права і эканоміка, 2005. – С. 113–115.

2 Лемцюгова, В. П. Беларуская акадэмічная лексікаграфія: стан, праблемы / В. П. Лемцюгова // Сучасныя праблемы беларускай лексікалогіі і лексікаграфіі. Матэрыялы міжнароднай навуковай канферэнцыі. Мінск : Права і эканоміка, 2006. – С. 188–195.

3 Речник српскохрватског књижевног и народног језика. – Књига I. Београд, 1959. – 814 с.

4 Béjoint, Henri. Lexicography of English / Henri Béjoint // New York: Oxford University Press, 2010. – 371 p.

В. И. Коваль

## КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА В ГОМЕЛЬСКОМ УНИВЕРСИТЕТЕ: СТАНОВЛЕНИЕ И ПЕРСПЕКТИВЫ

Компьютерная лингвистика в Гомельском государственном университете имени Ф. Скорины начала развиваться с 2005 года: именно в этом году Совет университета принял решение об открытии соответствующей специализации в рамках специальности «Русская филология». Несомненно, определяющую роль при этом сыграло то обстоятельство, что в ряде ведущих университетов нашей страны – Белорусском, Минском лингвистическом и Гродненском – к этому времени уже был накоплен немалый опыт в подготовке «компьютерных лингвистов», издавалась научная и учебно-методическая литература по различным направлениям компьютерной лингвистики, проводились научные конференции и даже защищались кандидатские диссертации. Что же касается филологического факультета

Гомельского университета, то вплоть до недавнего времени компьютер использовался нами лишь как удобная печатно-множительная техника, а также как инструмент получения официальной информации из компьютерной сети университета, научной и учебно-методической – из сети Интернет, как канал связи с помощью электронной почты.

Впервые о возможной работе в сфере использования компьютерных технологий в лингвистических исследованиях автор этих строк услышал от коллеги из Гродненского университета доцента Людмилы Васильевны Рычковой, предложившей еще в 1998 году, «на заре компьютерной эры», сотрудничество в деле подготовки машинного фонда белорусского языка. Дальше общих разговоров (с нашей стороны) «сотрудничество», однако, не пошло, о чем сейчас мы можем лишь сожалеть: драгоценное время было безвозвратно упущено, а компьютерные технологии в сфере лингвистических исследований между тем стремительно развивались. Очевидно, что лингвистика, как и гуманитарные науки в целом, имеют вполне реальное «компьютерное измерение». Л. В. Рычкова указывает также на важность применения компьютерных технологий в учебно-методической работе: «Использование ЭВМ и компьютеризированных учебных материалов позволяет обеспечить управляемую и оперативно контролируемую внеаудиторную самостоятельную работу обучаемых: освободить преподавателя от рутинной работы и сократить затраты учебного времени на вспомогательные виды работ» [1, с. 65]. Об успехах гродненских «компьютерных лингвистов», их тесных и разнообразных связях с единомышленниками не только из Беларуси (из БГУ и МГЛУ), но из России, Украины и Польши приходилось немало слышать, узнавать из научных журналов и информационных материалов.

Немалую «вдохновляющую» роль в зарождении компьютерной лингвистики в ГГУ имени Ф. Скорины сыграл (к сожалению, рано ушедший из жизни) профессор кафедры прикладной лингвистики БГУ Владимир Александрович Карпов – талантливый организатор науки, нестандартно мыслящий ученый, самобытный поэт. В разговоре с автором этих строк он, в частности, выражал твердое намерение (если не сам лично, то путем привлечения своих учеников – преподавателей кафедры) прочесть в Гомельском университете курс лекций по компьютерной лингвистике, оказать необходимую практическую поддержку.

Большую помощь на подготовительном этапе оказал нам и известный специалист в области компьютерной лингвистики – доктор филологических наук, профессор Александр Васильевич Зубов, заведующий кафедрой информатики и прикладной лингвистики Минского государственного лингвистического университета. Александр Васильевич не только подробно рассказал о «предметном наполнении» специальности «Компьютерная лингвистика» в своем вузе, о тематике курсовых, дипломных работ, подготовленных (и подготавливаемых) под его руководством диссертаций, но и поделился собственными публикациями, учебными программами различных курсов: «Алгоритмизация лингвистических задач», «Лингвистические информационные ресурсы»,

«Методы автоматической обработки текста», «Проблемы порождения и понимания текста и речи», «Прикладные системы обработки информации», «Современные технологии создания рекламы и презентаций» и др.

Со своей стороны, преподаватели кафедры русского, общего и славянского языкознания Гомельского университета, по инициативе которой и была открыта специализация «Компьютерная лингвистика», целенаправленно создавали методическую базу в новой для себя сфере: была изучена доступная научная и учебно-методическая литература, подготовлены учебные и рабочие программы, повышение квалификации преподавателей кафедры в Республиканском институте высшей школы так или иначе «увязывалось» с компьютерной лингвистикой. Одновременно проводилась профориентационная работа в школах Гомеля и области, рассылались информационные письма в районные отделы образования, что способствовало осознанному выбору абитуриентами новой для них специализации.

Естественно, что на начальном этапе становления специализации преподаватели кафедры обучали студентов использованию самых простых компьютерных программ для обработки языка и текста: PowerPoint, Vaal-Mini, WordStat и др.; студенты знакомились также с приемами компьютерного дизайна (программа CorelDraw) и основами компьютерной верстки (программа PageMaker). Новые компьютерные технологии постепенно внедрялись и в учебный процесс, и в научно-исследовательскую работу самих преподавателей.

Было, однако, понятно, что без конкретного сотрудничества с опытными специалистами в этой области, без их непосредственной «сопричастности» к нашему «периоду взросления» новую специализацию будет очень трудно «поднять на ноги». Именно поэтому мы обратились с предложением о чтении лекций и проведении занятий к ведущим специалистам нашей страны. В ноябре 2008 года доцент кафедры прикладной лингвистики Белорусского государственного университета Анастасия Ивановна Головня прочла для студентов-филологов нашего университета теоретический курс по компьютерной лингвистике, основное содержание которого было связано с изложением общей теории систем, автором которой является известный философ и системолог Ю. А. Урманцев. В мае 2009 года для студентов-«компьютерщиков» был прочитан спецкурс и проведены практические занятия заведующей кафедрой общего и славянского языкознания Гродненского государственного университета имени Я. Купалы доцентом Людмилой Васильевной Рычковой. Основное внимание Людмила Васильевна сосредоточила на содержании и формах работы с Национальным корпусом русского языка, что значительно активизировало работу в использовании этого ценного ресурса не только студентами, но и преподавателями, аспирантами и магистрантами кафедры. Наши представления о Национальном корпусе русского языка углубились и в результате участия ассистента кафедры Инны Геннадьевны Гомоновой в работе проведенного в Гродненском университете в апреле 2010 года Международного научно-практического семинара «Использование Национального

корпуса русского языка в лингвистических исследованиях и преподавании русского языка».

Немалое значение для успешного продвижения специализации «Компьютерная лингвистика» в плане создания и укрепления ее «материальной базы» стало открытие по инициативе кафедры в сентябре 2008 года компьютерного класса на 12 рабочих мест, предназначенного для проведения практических занятий по дисциплинам специализации. Весьма кстати оказалось и открытие при кафедре в феврале 2009 года научно-методического Центра русистики, которое произошло при финансовой поддержке Посольства Российской Федерации в Республике Беларусь. Центр располагает не только новейшей учебно-методической, справочной и научной литературой, изданной в России, но и современным оборудованием: двумя компьютерами, ноутбуком, мультимедийной установкой, множительной техникой – всем тем, что так необходимо для проведения занятий в рамках специализации «Компьютерная лингвистика».

В начале 2009 года кафедра выступила с инициативой о проведении в ноябре 2010 года Международной научно-практической конференции «Компьютерная лингвистика: научное направление и учебная дисциплина». Предполагалось, что в работе конференции примут участие ведущие специалисты в этой области из университетов Беларуси и Института языка и литературы Национальной академии наук, а также коллеги из России и Украины. Одна из целей конференции, по задумке ее организаторов, заключалась в своеобразной «сверке часов», т. е. в выяснении общих и отличительных мнений относительно самого понятия «компьютерная лингвистика», а также содержания нового понятия, закрепленного в Образовательном стандарте Министерства образования Республики Беларусь – «специалист по компьютерной филологии». По завершении конференции, в которой приняли участие более 50 специалистов, можно с полной уверенностью сказать, что она не только удалась с точки зрения реализации поставленных задач, но и стала очевидным важным этапом в становлении специализации «Компьютерная лингвистика» в Гомельском университете.

В продолжение работы конференции – на пленарном и секционных заседаниях – были обсуждены наиболее актуальные вопросы компьютерной лингвистики с точки зрения ее научной и учебно-методических составляющих: теория и практика машинного перевода; корпусная лингвистика; использование компьютерных технологий в научной и учебно-методической деятельности; прикладная лексикография, составление электронных учебных словарей и др.

Живой интерес участников конференции вызвали мастер-классы, проведенные известными специалистами в области компьютерной лингвистики. Так, московские ученые – старший научный сотрудник Института русского языка имени В. В. Виноградова РАН С. О. Савчук и научный сотрудник Всероссийского института научной и технической информации РАН Д. В. Сичинава – подробно рассказали о структуре и использовании

возможностей Национального корпуса русского языка – огромного виртуального массива художественных, публицистических, научных и диалектных текстов, открывающего большие перспективы в научно-исследовательской и учебно-методической работе филологов. Заведующая кафедрой общего и славянского языкознания Гродненского университета Л. В. Рычкова и ассистент этой же кафедры А. Ю. Станкевич поделились опытом преподавания компьютерной лингвистики в своем вузе, новыми наработками в этой сфере.

Основываясь на накопленных собственных наработках и богатом опыте коллег, специализация «Компьютерная лингвистика» в Гомельском университете будет и далее развиваться в русле основных направлений этой молодой отрасли языкознания: изучение и использование корпусов текстов, программное обеспечение лингвистических исследований, подготовка и использование в научной и учебно-методической деятельности электронных словарей.

#### СПИСОК ЛИТЕРАТУРЫ

1. Рычкова, Л. В. Язык. Образование. Компьютер: сб. науч. ст. / Л. В. Рычкова. – Гродно : ГрГУ, 2010. – 167 с.

Л. В. Рычкова

#### **«СПЕЦИАЛИСТ ПО КОМПЬЮТЕРНОЙ ФИЛОЛОГИИ»: ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ**

В 2009 году (всего через год после разработки и утверждения нового поколения образовательных стандартов) по филологическим специальностям, по которым головным вузом является БГУ, были предложены измененные стандарты, предполагающие подготовку специалистов по направлениям, в том числе по новому для республики направлению «Компьютерное обеспечение». Очевидно, что, изменяя стандарты и отходя от сложившегося подхода подготовки специалистов в рамках специальности – специализации, головной вуз имел целью повышение престижа и привлекательности филологических специальностей и расширение рынка труда для выпускников. Тем не менее, как весьма широкое и потому неопределенное название самого направления «Компьютерное обеспечение» (справедливости ради, следует отметить, что оно конкретизируется при определении объекта профессиональной деятельности специалиста), так и дополнительная квалификация, которую оно предполагает – «Специалист по компьютерной филологии», – реально создают трудности в идентификации сути предлагаемой квалификации и в трудоустройстве специалистов, которые ее в скором будущем получат.

Дело в том, что сочетание «компьютерная филология» далеко от терминологичности и, если и используется, то, как правило, не специалистами,

да и то с оговорками. Наиболее частые среди них – уточнение посредством достаточно широко известного и понятийно более или менее определенно-го термина *компьютерная лингвистика*, приводимого в скобках после сочетания «компьютерная филология» (это означает, что авторы ставят знак равенства между «компьютерной филологией» и компьютерной лингвистикой), либо путем употребления сочетания «компьютерная филология» в контекстуальном смысле (в тексте заключается в кавычки) как, например, в широко известной статье К. В. Вигурского и И. А. Пильщикова «Филология и современные информационные технологии (к постановке проблемы)» [1], где данное сочетание употребляется для обозначения собирательного понятия, отвечающего общему содержанию статьи, стимулировавшей, в определенной степени, появление в России новых магистерских программ, таких, например, как «Компьютерные технологии в филологии», «Инновационные технологии в филологии» и пр. Отметим, что подобных программ магистерской подготовки в Республике Беларусь пока нет.

Авторы вышеупомянутой статьи (тут и далее по тексту при цитировании первоисточников сохраняются орфографические, пунктуационные и иные особенности оригиналов) определяют шесть направлений, «осуществление которых представляется целесообразным с точки зрения использования современных информационных технологий в интересах филологической практики», а именно:

«1) Создание электронного фонда русской словесности, который должен обеспечивать сбор, обработку, надежное хранение и эффективное использование культурно значимых произведений русской литературы и фольклора <...>.

2) Создание национального корпуса русского языка, включающего грамматически размеченные тексты, которые отражают функционирование языка на разных этапах его истории.

3) Создание электронного словарно-справочного фонда, в котором должны накапливаться и предоставляться в свободный сетевой доступ словари, энциклопедии, справочники и указатели, необходимые для проведения филологических исследований <...>.

4) Создание средств автоматизации работ в сфере филологии, допускающих применение универсальных либо специальных алгоритмов. Средства автоматизации должны использоваться в составе электронных библиотек и других информационных систем для расширения их функциональных возможностей. Некоторые программные продукты могут применяться автономно на персональных компьютерах.

5) Организация цикла филологических исследований по подготовке изданий литературных произведений в электронной форме. Особое внимание должно быть уделено проблемам электронных научных изданий русской классики, а также изданиям произведений, не имеющих печатных аналогов, и мультимедийным текстам.

б) Учет и систематизация филологически значимых электронных фондов. Создание сводного электронного каталога сетевых информационных ресурсов филологического профиля».

Следует отметить, что обозначенные авторами цитируемой статьи направления перекликаются с целями и задачами получившего широкую известность мегапроекта создания Машинного фонда русского языка, который реализовывался в СССР в 80-е годы XX века, и к настоящему времени достигнуты определенные успехи в их практическом воплощении.

Рассмотрим «наполнение» направления «Компьютерное обеспечение» на примере ОСРБ 1-21 05 02-2009 – Образовательного стандарта Республики Беларусь по специальности 1-21 05 02 «Русская филология» [1], предусматривающего квалификацию специалиста «Филолог. Преподаватель русского языка и литературы. Специалист по компьютерной филологии».

В соответствии с ОСРБ 1-21 05 02-2009, объектом профессиональной деятельности, определяемым направлением подготовки «Компьютерное обеспечение», является «компьютерное обеспечение филологической деятельности». Предполагается также, что «выпускник должен быть компетентным в следующих видах деятельности: <...>

- компьютерно-филологическая» (?).

Среди требований к профессиональным компетенциям, которые могут быть отнесены к вышеуказанному виду деятельности, в ОСРБ 1-21 05 02-2009 отмечаются следующие способности выпускников:

*«в научно-исследовательской деятельности:<...> использовать в работе современные компьютерные методы сбора, обработки и хранения информации (?);*

*<...> в компьютерно-филологической деятельности (?):*

- работать с современными лингвистическими информационными ресурсами и экспертными компьютерными системами (?);
- применять навыки работы по компьютерному обеспечению филологической деятельности (?);
- использовать в профессиональной деятельности умения и навыки по автоматической обработке текстов (?);
- работать с прикладными системами обработки информации (?);
- применять современные компьютерные технологии при работе с разными корпусами текстов (?).

Поставленные в скобках вопросы означают неточность либо неполноту формулировок, что препятствует их однозначному пониманию и, следовательно, не снимает проблемы идентификации специалиста с квалификацией «Специалист по компьютерной филологии» посредством определения его компетенций.

Типовой учебный план рассматриваемого стандарта (таблица 2) определяет дисциплины направления «Компьютерное обеспечение» общим объемом 296 часов, из них аудиторного времени – 204 часа. Предусматриваются

также две недели производственной практики «Компьютерное обеспечение филологической деятельности», цель которой определена весьма нечетко и аморфно: «Цель данной практики – познакомить студентов с современными подходами к компьютерному обеспечению филологической деятельности, с основами лингвистической информационной базы, закрепить умения и навыки по автоматической обработке текстов, аннотированию и реферированию, при работе с экспертными компьютерными системами». Отметим, что нормой обеспечения специализации «Компьютерная лингвистика» в прежнем поколении образовательных стандартов считался объем аудиторного времени не ниже, чем 300 часов.

По направлению «Компьютерное обеспечение» ОСРБ 1-21 05 02-2009 предусматривает следующий стандартный набор дисциплин: Введение в компьютерную филологию (52/34); Инженерия знаний (48/34); Методы автоматической обработки текстов (102/68); Корпусная лингвистика (48/34); Формализация языка в экспертных системах (46/34); Вузовский компонент (50/34); Дисциплины по выбору студента (50/34). В скобках в числителе приведен общий объем часов, в знаменателе – объем аудиторного времени, отводимого на изучение конкретных дисциплин. За исключением первой дисциплины, все остальные являются традиционными для специализации «Компьютерная лингвистика», но представлены в сокращенном объеме, что изначально предопределяет их ознакомительный, а не познавательный характер.

Цикл естественнонаучных дисциплин учебного плана предусматривает изучение Основ информационных технологий в объеме 154 часов (из них 102 часа – аудиторное время). Этот значительный объем часов предусматривается на освоение того объема знаний, который сегодня получают выпускники средней школы, и никак не работает на подготовку специалистов в области «компьютерного обеспечения филологической деятельности».

Рассмотрим содержание дисциплины «Введение в компьютерную филологию», которая, по сути дела, должна являться введением в содержание направления подготовки специалиста, определять систему знаний и задавать траекторию дальнейшей подготовки специалистов: «Компьютерная филология как прикладная дисциплина. Когнитивный инструментарий компьютерной филологии (лингвистики). Направления компьютерной филологии. Моделирование общения. Моделирование структуры текста. Гипертекстовые технологии представления текста. Исторические замечания. Теоретические основания гипертекста. Компоненты гипертекста. Элементы типологии гипертекста. Некоторые гипертекстовые системы». Предполагается, что «выпускник должен

знать:

- место и роль компьютерной филологии в прикладной лингвистике;
- историю развития компьютерной филологии (лингвистики);
- основные направления компьютерной филологии (лингвистики);
- компоненты гипертекста



уметь:

- ориентироваться в проблемах прикладной лингвистике и компьютерной филологии;
- пользоваться когнитивным инструментарием компьютерной филологии;
- владеть моделированием общения;
- ориентироваться в гипертексте». Полагаем, что комментарии тут излишни, т. к. налицо интерференция «компьютерной филологии» и компьютерной лингвистики, не ясно также, по какому принципу отбирались и упорядочивались содержательные компоненты.

Рассмотрение содержания других дисциплин, изучение которых предполагается в рамках направления «Компьютерное обеспечение», а также знаний и умений, которые должны приобрести будущие «специалисты по компьютерной филологии», заставляет придти к весьма неутешительным выводам, которые кратко могут быть сформулированы следующим образом: отсутствие системности и преемственности в изучении дисциплин, концептуальная неопределенность в плане интерференции с компьютерной лингвистикой, смежной, но не тождественной, предметной областью, весьма слабая корреляция между декларируемыми компетенциями специалиста и содержанием подготовки.

Направление «Компьютерное обеспечение» требует существенной доработки и экспликации как для специалистов, так и для широких кругов населения. Только в этом случае можно ожидать повышения престижа филологических специальностей и, самое главное, обеспечить подготовку весьма необходимых в современных условиях специалистов в области компьютерного обеспечения филологической деятельности, то есть специалистов, способных распознавать, искать и использовать, а также создавать компьютерные (электронные) ресурсы филологической направленности. Следует также определиться, целесообразна ли дополнительная подготовка специалистов направления «Компьютерная лингвистика» (сегодня такая возможность отсутствует), и, наконец, трансформировать профессионально не ориентированный курс Основы информатики в дисциплину профессиональной направленности, например, в Основы лингвистической информатики.

#### СПИСОК ЛИТЕРАТУРЫ

1 Вигурский, К. В. Филология и современные информационные технологии (к постановке проблемы) / К. В. Вигурский, И. А. Пильщиков // Известия Российской академии наук. – Сер. Литературы и языка. – 2003. – Т. 62. – № 2. – С. 9–16. (Доступна также как электронный ресурс НПБ им. К.Д. Ушинского по адресу: <http://rd.feb-web.ru/vigursky-03.html>).

2 ОСРБ 1-21 05 02-2009 / Образовательный стандарт Республики Беларусь по сп-ти 1-21 05 02 Русская филология (по направлениям). – Мн., 2009. – 34 с.

## ИСПОЛЬЗОВАНИЕ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА В ПРЕПОДАВАНИИ ФИЛОЛОГИЧЕСКИХ ДИСЦИПЛИН<sup>1</sup>

Национальный корпус русского языка (НКРЯ) – мощный лингвистический ресурс, предназначенный для научных исследований и преподавания филологических дисциплин. Для того чтобы корпус стал, как того бы хотелось его создателям, необходимым и обязательным инструментом филолога и использовался максимально эффективно, пользователь должен представлять себе принципы его устройства и функциональные возможности.

В корпусе выделяется три составляющих: 1) коллекция текстов в электронной форме; 2) система лингвистической аннотации; 3) поисковая машина, обеспечивающая выполнение поисковых запросов. Каждая из этих составных частей может существовать либо как самостоятельный ресурс (текстовая коллекция как электронная библиотека, система лингвистической аннотации как особым образом представленное описание языка), либо в составе других информационных систем (поисковая машина как неотъемлемая часть любой информационной системы). Соединение этих компонентов в единое целое создает то, что называется электронным корпусом текстов в современном смысле этого слова. Рассмотрим роль каждого из компонентов в составе корпуса.

1. Коллекция текстов. Размер и состав коллекции зависит от целей и задач составителей корпуса. Национальный корпус современного языка по определению не может содержать всех текстов, созданных на этом языке, но он должен быть представительным, т. е. с максимальной полнотой отражать современное речевое употребление во всем разнообразии функциональных разновидностей. Таким образом, при составлении представительной коллекции текстов нельзя обойтись без знаний о функциональном устройстве современного языка и его истории, о соотношении литературного языка с нелитературными сферами (территориальными, социальными, профессиональными диалектами), об истории литературного языка и ключевых этапах литературного процесса и пр. В свою очередь, исследования, проводимые на основе корпусных данных, имеют обратное влияние на эти знания, расширяя и уточняя их.

2. Лингвистическая аннотация, или разметка состоит в том, что каждая словоформа текста сопровождается в корпусе тегами, содержащими лингвистические сведения об этой словоформе. Объем этих сведений зависит от типа аннотации и ее глубины. Большинство корпусов размечены морфологически – это означает, что в тегах указана лексема (словарная форма), которой принадлежит данная словоформа, информация о части речи и набор грамматических признаков. Например, в НКРЯ словоформа *книгою* сопровождается

---

<sup>1</sup> Работа выполнена при поддержке: РФФИ (грант 08-06-00371-а и 10-06-00151а), Программы ОИФН РАН «Генезис и взаимодействие социальных, культурных и языковых общностей».

следующим набором тегов: {книга=S,f,inan=ins,sg}, что в данном случае совпадает со стандартным морфологическим разбором существительного в курсе русского языка: начальная форма *книга*, существительное среднего рода, неодушевленное, в форме творительного падежа, единственного числа. Подробнее о морфологическом стандарте в НКРЯ см. [4].

Если в корпусе используется семантическая разметка, то теги, сопровождающие словоформу, содержат сведения о семантике слова. Это может быть таксономический класс, как в НКРЯ [3, с. 159]: для предметных имен – лица, животные, растения, вещества и материалы и др.; для прилагательных – размер, форма, цвет, вкус, запах, место, время, свойство человека и др.; для глаголов – движение, эмоция, речь, поведение человека и др.). Можно разметить лексико-грамматические разряды (конкретные, отвлеченные, собирательные, вещественные существительные; качественные, относительные, притяжательные прилагательные) или деривационные классы (уменьшительные, увеличительные, отглагольные, отадективные имена и пр.).

В корпусах устной речи используется аннотация, отмечающая место ударения в слове, сочетания согласных, позицию гласных и др., что существенно для постановки и решения фонетических и орфоэпических задач [1].

Метатекстовая аннотация – это информация, приписанная каждому тексту в целом. Она характеризует текст как с помощью «внешних» признаков (автор, его пол, возраст, дата создания текста, целевая аудитория, место и время издания), так и собственно текстовых категорий (сфера функционирования, жанровая и стилистическая принадлежность, тематика) [5]. Метатекстовая разметка отражает архитектуру текстового состава корпуса и служит для формирования пользовательских корпусов.

Следует иметь в виду, что правила (стандарты) аннотации в разных системах могут отличаться друг от друга, что объясняется возможностью множественных интерпретаций одного и того же языкового явления. Например, в грамматической теории не существует единого мнения о характере видовых противопоставлениях глаголов в русском языке, поэтому в одних системах видовая пара может рассматриваться как реализация одной лексемы (*убедить* и *убеждать* как формы *св* и *нсв* глагола *убеждать*), а в других системах – *убедить* и *убеждать* рассматриваются как разные глаголы.

3. Система поиска. Система лингвистической разметки превращает библиотеку текстов в подобие гигантской картотеки, в которой каждая словоформа расписана на «карточки», причем сразу на несколько – по числу лингвистических признаков, приписанных словоформе. Для извлечения информации из этой картотеки в корпусе используется *поисковая система*, которая позволяет сортировать «карточки» по любому из признаков. Чем более функциональна система поиска, тем большее число лингвистических задач можно решать с помощью корпуса. «Дружественный» интерфейс, понятная логика и удобство формирования запросов обеспечивают скорость решения поставленных задач. Например, при обычном контекстном

поиске результаты выдаются в виде предложений, которые в дальнейшем подлежат ручной обработке. При поиске по ключевым словам контексты обрабатываются автоматически и результаты выдаются в виде конкорданса, что в большинстве случаев значительно облегчает и ускоряет обработку данных.

Система лингвистической аннотации каждого типа отражает систему правил соответствующих разделов науки о языке. Поэтому корпус можно эффективно использовать при изучении лингвистических дисциплин, прежде всего тех, категории которых были использованы в аннотировании текстов. Покажем это на примере морфологической аннотации.

Изучение курса морфологии русского языка в школе и в вузе начинается с теоретического освоения системы правил (например, состава частей речи, системы категорий имени существительного) и завершается выработкой умения выполнять морфологический анализ текста, при котором каждой словоформе приписываются присущие ей грамматические признаки. Тот же путь проходят и создатели корпуса: сначала разрабатывают систему морфологической аннотации (правила), отвечающую целям и особенностям корпуса, а затем осуществляют ручную или автоматическую разметку текстов (морфологический анализ). Однако анализ в этом втором случае используется не для проверки знаний, а выступает как средство создания нового продукта – размеченного корпуса.

Для пользователей корпуса ситуация выглядит перевернутой. Размеченный (морфологически проанализированный) текст для них – не цель, а данность, так же как и набор морфологических категорий и набор инструментов поиска, с помощью которых можно строить запросы. Запросы могут быть трех типов: 1) относительно словоформ, слов и словосочетаний (напр., употреблялась ли форма *городы* в текстах XVIII в.); 2) относительно грамматических признаков (напр., актуально ли употребление форм 2-го родительного падежа для текстов современного периода); и 3) относительно различных комбинаций лексем и грамматических признаков (напр., проверить, продуктивна ли фразеологическая модель *нужный как кому/чему кто/что*). В ответ на запросы выдаются контексты, в которых представлены языковые единицы, найденные по запрашиваемым признакам.

Корпус можно использовать для решения учебных задач, формулируя простые запросы, например: привести примеры субстантивированных прилагательных; на основе анализа найденных в корпусе примеров определить, какие синтаксические функции могут выполнять слова *холодно*, *интересно* и установить частеречную принадлежность этих слов и т. д. На подобные запросы корпус выдает десятки, и даже сотни примеров – понятно, насколько это облегчает труд преподавателя, упрощая составление многовариантных заданий для самостоятельных, лабораторных и проверочных работ [2; 6].

Помимо обучающих целей корпус может быть использован для решения исследовательских задач. Благодаря корпусу сбор материала перестает

быть рутинной, поглощающей значительную часть времени, отпущенного на исследование, и превращается в увлекательное занятие. Исследователь имеет возможность сосредоточиться на творческих участках работы: оптимизировать поисковые запросы, чтобы извлечь из корпуса максимум необходимой информации при минимуме ненужной; рассматривать предмет исследования в разных аспектах; обрабатывать большой объем данных для подтверждения выводов статистическими результатами. Скорость, с которой подбираются примеры, приобщает к исследовательскому процессу и студентов, которые получают возможность за короткий срок провести небольшую творческую работу. О практике использования НКРЯ в учебном процессе см. [2, 8].

Практика показала, что корпус можно использовать не только при изучении морфологии, лексикологии, семантики, синтаксиса, фонетики и орфоэпии, орфографии, но и в курсах фразеологии, словообразования, стилистики, литературного языка и его истории. Покажем это на нескольких примерах.

1) Проверить продуктивность фразеологической модели. На запрос: «*нужный как на расстоянии 1 от сущ., дат. на расстоянии 1 от сущ. им.*» получаем следующий результат: наряду с зафиксированными в словарях (*нужно*), *как собаке пятая нога; как мертвому припарки*, в текстах встречаются обороты, построенные по данной модели, что свидетельствует о ее продуктивности: (*нужен*) *как игуменье шпоры, как архиерею гармонь, как рыбке (щуке) зонтик, как заднице гвоздь в диване*.

2) Недостаток изучения фонетики и орфоэпии состоит в том, что студенты, как правило, знакомятся с образцами произношения в виде транскрипции, а не в виде звучащей речи. Мультимедийный корпус в составе НКРЯ, снабженный орфоэпической разметкой, поможет провести наблюдения над звучащими образцами и познакомиться с такими особенностями «старшей нормы» литературного языка, как «эканье» ([в'э]нок, [б'эл'э]соватое, [м'э]тался); произношение твердых согласных перед [э]: *бу[γ]а'л[тэ]р, эф[фэ]кт*; произношение сочетания *щн* как [шн] словах *беспомощн[шн]о, в сущн[шн]ости, всено'щн[шн]ого*; произношение *ј* в формах местоимений (*им, их [јим], [јих]*) и др.

#### СПИСОК ЛИТЕРАТУРЫ

1 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.

2 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.

3 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.

4 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.

5 Савчук, С. О. Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции / С. О. Савчук // Национальный корпус русского языка: Новые результаты и перспективы. – М. : Нестор-История, 2009. – С. 62–88.

6 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.

7 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru>. – Дата доступа: 28.10.2010.

8 Национальный корпус русского языка и преподавание [Электронный ресурс]. – Режим доступа: [http://studiorum.ruscorpora.ru/index.php?option=com\\_tag&Itemid=75](http://studiorum.ruscorpora.ru/index.php?option=com_tag&Itemid=75). – Дата доступа: 28.10.2010.

Д. В. Сичинава, М. А. Шведова

## **ПАРАЛЛЕЛЬНЫЕ КОРПУСА В СОСТАВЕ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА: ТЕХНОЛОГИИ И РЕШАЕМЫЕ ЗАДАЧИ**

Параллельные корпуса в составе НКРЯ (поиск по ним доступен по адресу <http://ruscorpora.ru/search-para.html>) развиваются начиная с 2005 г., и им посвящено несколько публикаций, подготовленных коллективом разработчиков (см. [1], [2], [3], [4]). Параллельный корпус – это корпус, в котором тексту на одном языке сопоставлен перевод этого текста на другой язык. В состав НКРЯ включаются корпуса, где в качестве языка оригинала или перевода выступает русский язык.

Между единицами оригинального и переводного текста (обычно – между предложениями) с помощью специальной процедуры устанавливается соответствие; эта процедура называется выравниванием, а тексты, соответственно, выровненными. Возможен также вариант, при котором сопоставляются друг с другом различные переводы одного и того же текста (не только на разные языки, но и на один и тот же: сравнение между различными переводами на один язык много говорит как о внутриязыковой вариативности, так и о значении тех или иных выражений в языке оригинала). Однако обычно они включаются в корпус всё же наряду с оригиналом (но существенно здесь прежде всего то, что любая пара текстов в таком массиве может быть сопоставлена по принципу «каждый с каждым»). Параллельные корпуса НКРЯ пока не включают в себя таких троек (и больших массивов) выровненных текстов, однако такая возможность технически предусмотрена, реализована в крупных параллельных славянских корпусах (см. в конце статьи) и будет развиваться по мере разработки общего поискового механизма по этим корпусам.

В настоящее время на сайте НКРЯ размещены параллельные англо-русский, русско-английский, немецко-русский, украинско-русский и русско-украинский корпуса. Работа над разработкой украинско-русского и русско-украинского корпусов ведётся авторами настоящей статьи, над остальными – коллективами под руководством А. А. Кретьова (Воронеж) и Д. О. Добровольского (Москва – Вена). На сентябрь 2010 г. объём доступных для поиска корпусов следующий:

Таблица 1

	Текстов	Предложений	Слов
Англо-русский	57	436 431	6 229 619
Русско-английский	24	61 259	963 507
Немецко-русский	15	72 553	1 270 128
Украинско-русский	64	32 496	316 649
Русско-украинский	16	16 330	168 125

К размещению на сайте готовятся русско-белорусский и русско-белорусский параллельные корпуса (продолжение проекта, начатого и развивавшегося группой под руководством А. В. Зубова в Минском государственном лингвистическом университете, см. [5]) и польско-русский параллельный корпус (подготавливаемый Б. В. Ореховым (Уфа, Башкирский государственный педагогический университет) совместно с коллективом в Варшавском университете под руководством М. Лазиньского).

Параллельные корпуса НКРЯ представлены в формате XML и выровнены по предложениям. Корпуса получают автоматическую морфологическую разметку на основе морфологических анализаторов компании «Яндекс»; планируется размещение части текстов со снятой вручную омонимией. При поисковой выдаче помимо контекста на искомое слово (грамматическое значение, конструкцию) выдаётся также переводная (или, соответственно, оригинальная) параллель данного предложения.

Выровненный параллельный корпус представляет собой важный инструмент для научных исследований. Сюда входят, в частности, контрастивные исследования лексики и грамматики (имеющие большое значение не только в типологическом освещении, но и при сравнении близкородственных языков).

К примеру, анализ передачи украинского плюсквамперфекта (типа *прийшов був*) показывает, что в русском языке он лишь в части случаев соответствует генетически тождественной конструкции *пришел было*; в значительном количестве примеров в русском тексте незавершённость ситуации или аннулированный результат, передающиеся украинским «давнопрошедшим временем», передаётся исключительно контекстом (*Лягнула була з маху, так би мовити, довірчо поділилась цікавим спостереженням* (О. Забужко) – *Лягнула с пылу с жару, так сказать, доверительно поделилась интересным*

наблюдением). Обнаруживается также ряд примеров с семантикой 'ирреальная модальность', недостаточно выявленной в существующих описаниях украинской формы: *Ледве держалася на ногах, і нікого не було коло неї, о кого могла була опертися* (О. Кобылянская). – *Едва держалась на ногах, и никого не было около нее, на кого могла бы опереться.*

Для исследований по языковой картине мира в сопоставительном аспекте значительный материал предоставляет, в частности, перевод по данным Корпуса культурно-специфических лексем. Например, лексема *тоска* рассматривалась в известной работе А. Вежбицкой о соответствующих концептах в русском языке; это слово соответствует в английском языке (в тех или иных контекстах) словам *gloom, dismay, grief, agony, excruciating feelings, anguish, ache, despair, loneliness, yearning, longing, distress, frustration, wistfully, desperation, nostalgia, misery...* Ср. например, также передачу специфического употребления украинских притяжательных прилагательных со значением 'любимая': дословная (*Вона давно вже була Іванкова, ще з тринадцяти літ* (М. Коцюбинский) – *Она давно уже была Иванкова, еще с тринадцати лет*) и с передачей семантики (*А ти все не вірила, що вона була Михайликова!* (О. Кобылянская) – *А ты все не верила, что Михайлик ее любил!*). Наглядный пример соответствия двух слов-понятий в одном языке одному слову в другом даёт следующий пример: *То ж любити, а то – кохати... Любиш – батька, матір, людей; а кохаєш – милого* (П. Мирный). – *Любовь любви рознь. Одно – любить отца, мать, людей; а другое – милого.*

Большое значение имеет корпус для исследований по теории и практике перевода, например, по передаче заглавий, идиом, различного рода идеологем. Так, анализ переводов советского времени с украинского на русский язык показывает, что в них систематически различными переводчиками устранялась религиозная фразеология и «политически некорректная», испытывая современную терминологию, этническая лексика – вместо *неба* персонажи апеллировали к *справедливости, жид* достаточно регулярно представлялся, к примеру, *ростовщиком, а москаль солдатом*. Опускались иногда целые фразы (*Святий Михайло був його патроном* у Кобылянской) и эпизоды (например, с развратным монахом у Нечуя-Левицкого; «удовлетворительный» идеологически, он, по-видимому, был опущен как «неприличный»).

Параллельный корпус – существенный механизм для построения дву- и многоязычных переводных словарей, в том числе специальных. Например, англо-немецкий корпус парламентских слушаний (<http://corpus.leeds.ac.uk/paraquery.html>) позволяет вычислить при помощи статистического механизма межъязыковые аналоги тех или иных слов, важные для перевода терминологии, например, неоднозначно переводимый (и не имеющий полного аналога в континентальном праве) английский термин *rule of law* («верховенство права») соответствует немецким *Rechtstaat* («правовое государство»), *Rechtsstaatlichkeit* («правовая государственность»), *Demokratie* («демократия»).



Параллельный корпус может также использоваться при обучении русскому и иностранным языкам. Параллельные тексты («билингвы») давно используются в педагогике в рамках различных методов обучения иностранному языку; существуют специальные программы построения электронных параллельных текстов для этой задачи, например, бесплатная утилита LeoBilingua ([www.hot.ee/bclogic/descr-r.html](http://www.hot.ee/bclogic/descr-r.html)). Собственные параллельные корпуса (построенные полуавтоматическим методом) широко используются в коммерческих программах машинного статистического перевода и вообще обработки текста (language processing). Именно на больших параллельных корпусах текстов основаны такие проекты, как Google Translate. Например, пушкинская строчка *Паду ли я, стрелой пронзенный* в Google Translate при переводе на украинский язык передана как *Впаду я, вражений стрілою* – то есть не пословной передачей, а строчкой из перевода «Евгения Онегина», сделанного Максимом Рыльским (из этого примера непосредственно видно обращение перевода к корпусу).

В своей работе НКРЯ постоянно сотрудничает с построенными на основе схожей методологии (в том числе на принципе общедоступности материалов для научной работы) параллельными корпусами славянских языков, такими, как ParaSOL (ранее Regensburg Parallel Corpus, <http://www-korpus.uni-r.de/ParaSol/>, см. также [7]), ASPAC (Amsterdam Slavic Parallel Aligned Corpus, авторский проект А. Барентсена, <http://home.medewerker.uva.nl/a.a.barentsen/>), InterCorp в рамках Чешского национального корпуса (<http://www.korpus.cz/intercorp/>, см. также [8]). При поддержке НКРЯ Б. В. Ореховым разработан параллельный корпус переводов «Слова о полку Игореве» (<http://nevmenandr.net/slovo/>, см. также [6]).

В настоящее время сотрудниками всех этих корпусов обсуждается проект создания общей системы, объединяющей поиск по существующим параллельным корпусам.

#### СПИСОК ЛИТЕРАТУРЫ

1 Андреева, Е. Г. Грамматика и лексика (на материале англо-русского корпуса параллельных текстов) / Е. Г. Андреева, В. Б. Касевич // Национальный корпус русского языка: 2003–2005. – М. : Индрик, 2005. – С. 51–60.

2 Добровольский, Д. О. Корпус параллельных текстов в исследовании культурно-специфичной лексики / Д. О. Добровольский // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. – СПб. : Нестор-История, 2009. – С. 383–401.

3 Добровольский, Д. О. Корпус параллельных текстов / Д. О. Добровольский, А. А. Кретов, С. А. Шаров // Научная и техническая информация, сер. 2. Информационные процессы и системы. – 2005. – № 6. – С. 121–127.

4 Добровольский, Д. О. Корпус параллельных текстов: архитектура и возможности использования / Д. О. Добровольский, А. А. Кретов, С. А. Шаров // Национальный корпус русского языка: 2003–2005. – М. : Индрик, 2005. – С. 45–49.

5 Зубов, А. В. Лингво-методические возможности русско-белорусского параллельного корпуса текстов / А. В. Зубов // Русский язык: Исторические судьбы и современность. IV Международный конгресс исследователей русского языка. Труды и материалы. – М. : МГУ, 2010. – С. 516–517.

6 Орехов, Б. В. Параллельный корпус переводов «Слова о полку Игореве»: итоги и перспективы/ Б. В. Орехов // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. – СПб. : Нестор-История, 2009. – С. 462–473.

7 v. Waldenfels R. Compiling a parallel corpus of slavie languages. Text strategies, tools and the question of lemmatization in alignment. In: Brehmer, B., Zdanova, V., Zimny, R. (Hrsg.); Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV) 9. München, 123–138, 2006 (доступно по адресу: <http://www-nw.uni-regensburg.de/%7E.war05297.slavistik.sprachlit.uni-regensburg.de/pub/WaldenfelsParallelCorpora2006.pdf>)

8 Vavřín, M. Korpus InterCorp/ M. Vavřín, A. Rosen // [Электронный ресурс] – Режим доступа: <http://korpus.cz/intercorp-info.php> – Дата доступа: 28.10.2010.

Б. Э. Хакимов, Р. А. Гильмуллин

## **МОРФОЛОГИЧЕСКАЯ РАЗМЕТКА В КОРПУСЕ ТАТАРСКОГО ЯЗЫКА<sup>1</sup>**

В данной статье описаны основные принципы морфологической разметки для корпуса татарского языка, разрабатываемого в Институте прикладной семиотики Академии наук Республики Татарстан. Как известно, одним из эффективных средств решения многих лингвистических задач являются электронные корпуса текстов – унифицированные, структурированные, размеченные и репрезентативные массивы языковых данных. В течение последних десятилетий было разработано большое количество корпусов для разных языков, в первую очередь европейских. В то же время следует отметить, что для тюркских языков эта работа была начата значительно позже и, например, для татарского языка в данный момент вообще не существует каких-либо специализированных лингвистических корпусов. Этим и определяется актуальность подобных исследований и разработок не только для татарского языкознания, но и для тюркологии в целом.

В процессе разработки корпуса татарского языка было принято решение на первом этапе сосредоточиться на морфологической разметке. С одной стороны, это обусловлено особенностями самого татарского языка как представителя тюркской языковой семьи, принадлежащей к агглютинативному типу. В таких языках значительная часть информации о слове заключена в его морфемной структуре. С другой стороны, отсутствие эффективных синтаксических анализаторов и специализированных

---

<sup>1</sup> Работа выполнена при поддержке молодежного гранта Академии наук Республики Татарстан № 05-9/2010 (Г)

семантических словарей татарского языка затрудняет процесс автоматизации других видов лингвистической разметки.

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. По этой причине предлагается подход, при котором морфологические характеристики словоформ в корпусе являются отражением его морфемного состава, т. н. «аффиксальной цепочки».

Автоматизация процесса морфологической разметки текстов в корпусе татарского языка обеспечивается путем создания специализированных программных средств, использующих в своей работе лингво-процессоры. Мы проанализировали возможности автоматизации морфологической разметки татарского текста на примере морфологического анализатора на базе правил РС-КИММО [1; 2], использующего двухуровневую модель татарской морфологии. Данный морфоанализатор был создан нами ранее и получил использование в различных системах автоматической обработки текста. Информационная база РС-КИММО, с точки зрения разработчика морфологического анализатора, состоит из двух файлов, созданных пользователем. Первый файл – файл правил, который описывает алфавит и фонологические правила. Второй файл – лексикон, содержит описание морфотактических правил и словарь лексических единиц (корневых и аффиксальных морфем) с их грамматическими характеристиками.

Грамматические характеристики татарских частей речи достаточно полно описаны в многочисленных трудах по татарскому языкознанию [4; 6]. Данные характеристики имеют следующее формальное выражение с точки зрения их распознавания и автоматической разметки:

I. Лексико-грамматические (семантические) разряды не имеют явных и однозначных формальных признаков, некоторые из этих разрядов могут иметь непостоянные формальные признаки. Однако следует отметить, что подобные модели могут также образовывать слова с другим значением и подобная структура произвольной словоформы не указывает однозначно на соответствующий семантический разряд.

II. Собственно морфологические категории в татарском языке (например, у существительных – число, принадлежность, падеж, сказуемость; наклонения и времена у глаголов) имеют явные формальные признаки, которые в большинстве случаев однозначно характеризуют словоформу.

Помимо упомянутых грамматических форм, с точки зрения словоизменительных парадигм, существует еще значительное количество аффиксов, не учитываемых в традиционных описаниях. В первую очередь это

относится к так называемым «двухфункциональным» аффиксам, принадлежность которых к той или иной категории в татарском языкознании является предметом дискуссии. По нашему мнению, несмотря на то, что данные аффиксы окончательно не включены в татарском языкознании в систему грамматических категорий, их функции должны быть отражены в морфологической разметке корпуса.

В результате анализа был сделан вывод о том, что в татарском языке морфологические характеристики легко распознаются при автоматическом анализе аффиксального состава словоформы, а существующая версия морфоанализатора на базе двухуровневой модели позволяет в значительной степени автоматизировать процесс морфологической разметки текста. Задача же выявления и автоматической разметки семантических (лексико-грамматических разрядов) различных частей речи не может быть решена с учетом лишь данных морфологии. Исходя из этого, можно полагать, что, по крайней мере, на первом этапе создания корпуса морфологическая разметка должна содержать лишь информацию о грамматических категориях, явно выраженных аффиксами.

Морфологическая разметка в проектируемом корпусе татарского языка представляет собой систему условных обозначений – тэгов, приписываемых каждой словоформе текста в результате работы уже упомянутого модуля морфологического анализа. На данном этапе планируется использовать язык разметки HTML с добавлением специфических тэгов. В корпусе татарского языка каждое слово получает определенный набор специальных морфологических атрибутов: основа (lex), часть речи (POS), грамматические признаки (gr). Каждый атрибут принимает определенные значения, альтернативные варианты разбора перечисляются через символ «/». Так, слово «балаларыбызга» (“нашим детям”) получает следующую разметку: `<w lex='бала' POS=N gr=PLU, 1pPOSS, DAT>балаларыбызга</w>`.

Далее представлен фрагмент таблицы грамматических признаков для имен существительных и глагола в татарском языке.

В дальнейшем планируется включение в корпус татарского языка синтаксической и семантической разметки, а также средств разрешения многозначности.

Таблица 1

Аффиксальный класс морфоанализатора	Лексическая запись морфемы	Обозначение	Определение
Имя существительное			
N_Plural	+ЛАр	PLU	множественное число
N_Poss	+Ым	1sPOSS	принадлежность 1 л. ед. числа
	+Ыбыз	1pPOSS	принадлежность 1 л. мн. числа
	+Ың	2sPOSS	принадлежность 2 л. ед. Числа
	+Ығыз	2pPOSS	принадлежность 2 л. мн. Числа
	+СЫ	3sPOSS	принадлежность 3 л. ед. числа
N_Case	0	NOM	основной падеж (номинатив)
	+нЫң	GEN	притяжательный падеж
	+ГА	DAT	дательный падеж (датив)
	+нЫ	ACC	винительный падеж (аккузатив)
	+ДАН	ABL	исходный падеж (аблатив)
	+ДА	LOC	местно-временной падеж (локатив)
Глагол			
Reflexive	+Ыш	Ref	взаимно-совместный залог
	+Ын	Ref	залог
	+Ыл	Pass	возвратный залог
	+ДЫр	Cous	страдательный залог понудительный залог
Tenses	+Й	Present	настоящее время
	+ДЫ	PastDef	прошедшее результативное
	+ГАН	PastUnDef	прошедшее категорическое
	+Ыр	FutUnDef	будущее неопределенное
	+мАс	FutNeg	будущее неопределенное в отрицательной форме
	+АчАк	FutDef	будущее определенное
Condition	+сА	Cond	условное наклонение
Negative	+мА	Neg	Отрицание

## СПИСОК ЛИТЕРАТУРЫ

1 Suleymanov, D. Sh. Tatar phonological rules as a base of two-level morphological analyzer, in Proceedings of LP'2000, ed. B.Palek and O.Fujimura/ D. Sh. Suleymanov, R. A. Guilmouline, A. A. Guilmouline – Prague : The Karolinum Press. – P. 495–504.

2 Сулейманов, Д. Ш. Двухуровневое описание морфологии татарского языка / Д. Ш. Сулейманов, Р. А. Гильмуллин // Тезисы Международной научной конференции, посвященной 200-летию университета: "Языковая семантика и образ мира". 7–10 октября 1997. Книга 2. – Казань : Изд-во КГУ. – С. 65–67.

3 Сулейманов, Д. Ш. Регулярность морфологии татарского языка и типы нарушений в языке / Д. Ш. Сулейманов // Когнитивная и компьютерная лингвистика / Науч. ред. Р. Г. Бухараев, В. Д. Соловьев, Д. Ш. Сулейманов. – Казань : КГУ, 1994. – С. 77–106.

4 Татарская грамматика. Т. II. Морфология / под ред. М. З. Закиев, Ф. А. Ганиев, К. З. Зиннатуллина. – Казань : Татарское кн. изд-во, 1993. – 397 с.

5 Хакимов, Б. Э. К разработке системы параметров разметки документов в электронном корпусе татарских текстов / Б. Э. Хакимов, Р. А. Гильмуллин // Труды Казанской школы-семинара по компьютерной и когнитивной лингвистике TEL-2008. – Казань : Изд-во КГУ, 2009. – С. 24–29.

6 Хисамова, Ф. М. Татар теле морфологиясе / Ф. М. Хисамова. – Казань : «Магариф», 2006. – 335 с.

И. Н. Юдкин-Рипун

## ДЕСКРИПТОРЫ ИДИОМАТИКИ КОРПУСА ДРАМАТИЧЕСКИХ ТЕКСТОВ: МЕТОДИКА ТЭГГИРОВАННЫХ СПИСКОВ

Задачу использования тэга и так называемых тэггированных списков для потребностей корпусной лингвистики впервые поставил белорусский исследователь А. В. Зубов, предложив описывать «словоупотребление текста» [6, с. 37]. Очевидна сопоставимость тэгов с тезаурусами типа *in situ*, т.е. описаниями лексики по сочетаемости и распределению в тексте методами дистрибутивного анализа, где белорусской науке также принадлежит приоритет [4]. Понятие тэга предстает обобщением и распространением фрейма как перечня дескрипторов алгоритма повествования (слотов). Такая списочная структура пригодна для описания фабул: в частности, «глагольные фреймы представляют собой полные структуры будущих предложений с соответствующими глаголами», причем в отношении сказки, например, это – «последовательность сказуемых», тогда как «подлежащие, дополнения ... определяют ее сюжет» [5, с. 166, 171]. С подбором лексики в тэг связана «семантическая изотопия» как «наличие какого-либо общего семантического признака» [5, с. 200], что обобщает как принадлежность семантическому полю (синонимии, гипонимия), так и сочетаемость (валентность).

Списки типа семантически изотопных цепочек, распространяющие фреймы (тэги), пересекаются с по крайней мере тремя видами известных лексических таксонов. Это – изосемантические ряды, в которых описываются закономерные семантические переходы, осуществляемые в рамках

этимонов и представимые в тексте в этимологических фигурах (например, [камень - счет], лат. *calx* – *calculation* или рус. *муть* – *смута*) [9, с. 203, 206]; ряды партонимов или частичных характеристик целого, например, «наименований частей тела» [11, с. 122]; наконец, широко истолкованные В. Г. Адмони супплетивные отношения, согласно которому «самое общее содержание отношений между грамматическими значениями в синтагматическом ряду..., как и в ряду парадигматическом – это содержание супплементности, восполнения» [2, с. 21]. Общей их основой является интеграция неоднородностей в представляемом ими таксоне.

По самому принципу своего построения такого рода таксоны принадлежат к тому классу явлений, в которых, согласно В. Г. Адмони, «словосочетания ... выявляют парадигматическое строение», так же как, со своей стороны, элементы парадигматических таксонов не могут находиться «в состоянии синтаксического покоя» [1, с. 39–40], демонстрируя преодоление дуализма парадигматики и синтагматики. Иначе говоря, списочные структуры, строящиеся на основе описания текста его же материалом, представляются как превращение синтагматических «узлов» текста в таксономические «гнезда» парадигматики. Выявляется общая полевая структура парадигматики и синтагматики, основанная на противопоставлении центра и периферии, на «асимметрии в языке плана выражения и плана содержания», где «билатеральность» знака играет ведущую роль [1, с. 26].

Представляя лексический состав описываемого участка текста по сочетаемости, такого рода таксоны включают не отдельные лексические единицы как словарные элементы, а именно участки текста, то есть синтагмы, причем «синтагма ... может состоять из единого слова» [7, с. 95]. Эти синтагмы, названные А. М. Пешковским «словами с формой словосочетаний» [10, с. 70], дают основания рассматривать соответствующие списочные структуры как «множества множеств» ( $\{\{a\}, \{b\}, \{c\} \dots\}$ ), а не элементов  $\{a, b, c \dots\}$ , что отражает построение самого текста по принципу «диагонализации» таксонов и позволяет говорить о подобии языка континууму. Иначе говоря, таксоны – тэги, описывающие текст, строятся не из лексических единиц как элементов потенциального словаря, а из идиоматических словосочетаний (а том числе однословных), присутствующих в идиолекте исследуемого корпуса текстов.

Воссоздание синтагматических характеристик тут обеспечивает воспроизведение анизотропности (обобщения вышеотмеченной билатеральности знака) или, по А. М. Пешковскому, «необратимости» как ведущего отношения слов в словосочетании: «Неизбежным спутником необратимости отношений является еще сознание неравенства», а потому «необратимые отношения в каждой паре соотносящихся внешне выражаются лишь в одном из них», что «создает простое подчинение одного слова другому» [10, с. 80–81]. Отсюда выводится тезис о том, что «подчинение внутри предложения лежит в основе связей между словами, а сочинение лишь его дополняет» [10, с. 445], и даже более того, «сочинение внутри предложения – лишь

эпизод на фоне подчинения» [10, с. 83]. Это знаменитое рассуждение, устанавливающее приоритет неоднородности по отношению к однородности (в частности, перечислениям – однородным членам предложения, приложениям), гипотаксиса – к паратаксису, позволяет и саму парадигматику представлять как периферийную зону синтагматики – как списочные структуры, потенциально присутствующие в самом тексте. Сходную мысль развивает и В. В. Мартынов, обобщая понятие предикации в понятии центральной позиции предложения и утверждая, что «названия предметов тяготеют к маргинальным позициям, а названия их признаков – к центральной» [8, с. 17]. Наличие полевой структуры «центр – периферия» определяет приоритет синтагматики текста с его ведущей ролью гипотаксиса, подчинения по отношению к паратаксису, сочинению. Списочные структуры, описывающие корпус текстов, представимы тогда как его «оболочка», предназначенная для аппроксимации, приближения к его содержанию.

Такие идиоматические ориентированные структуры описывают текст на основе его автодескрипции, то есть использования его же материала в качестве метаязыка описания. Преимущества такого подхода – в том, что учитываются особенности мотивировки семантической деривации в идиоматическом рассматриваемого корпуса текстов. Особенно интересно для выявления мотивировки использовать содержащиеся в самом тексте списочные структуры сочинительного строения (однородные члены, паратаксисы). Минимальные списки такого рода представляют паросочетания, в частности, фигуры *гендиадес*: их рассматривают как идиоматические (оказиональные, ситуативные) синонимы, однако они демонстрируют признаки семантической изотопии и могут считаться компонентами тэгов. Например, в словесной дуэли королевы Маргариты со сторонниками Глостера в «Ричарде III» (акт 1, сцена 3) В.Шекспира в ее устах в гендиадесе [shame & charity] оба компонента представляются скорее как партонимы (“My charity is outrage, life my shame”). Очевидно, в таких условиях сочинительные отношения между членами гендиадеса находятся в состоянии неустойчивого равновесия и могут легко сместиться в сторону подчинения, преобразуя списочную структуру в синтагму. Существенным являются мотивы обращения к таким идиомам, определяющие ту или иную трактовку паросочетания. Обусловленность трактовки демонстрируют внутренняя рифма и, шире, явления паронимической аттракции как мотивировочные факторы плана выражения. Связанные аллитерацией слова, в частности, указывают на подразумеваемое *tertium comparationis*, определяющее сдвиг их смысла. Такое сопоставление позволяет, например, выделить скрытый гендиадес [horse & hostess] в реплике Бастарда из шекспировского «Короля Джона» (акт 2, сцена 1), мотивированного язвительностью персонажа (“Saint George ... sits on his horse back at mine hostess’ door”).

Особый интерес для задач идиоматической автодескрипции представляют случаи, в которых разворачивание фабулы в наррацию (как,



например, при упоминавшемся описании сказки глагольными фреймами) отсутствует и реконструируется косвенным образом. Такая ситуация имеет место в драме, где имплицитно присутствует скрытое повествование о событиях, восстанавливаемое по репликам персонажей как нарративная presupпозиция, а смысл идиоматики определяется дейктической сетью взаимных ссылок [12]. Драматический текст предстает как рефлексия над текстом эпическим и в свою очередь порождает рефлексии, определяющую смысл идиоматики. Для уяснения такой рефлексии автодескрипция драматического текста представляется весьма продуктивной. Существенно, что в драме локация высказываний то есть их отнесенность к соответствующим персонажам, выступает на первый план (относительно предикации и номинации, по Ю. С. Степанову), отражая роль намерений в определении смысла. «Понятие интенциональности представляет собой одно из проявлений связи значения и смысла» [3, с. 156], и именно в драме оно реализуется наиболее явно.

Перспективность автодескрипции драматической идиоматики демонстрирует изучение списочных структур в сопоставлении с гипотаксисами в шекспировском корпусе. Фигуры перечисления широко используются как средство аргументации. Одним из примеров может служить известный монолог Архиепископа Кентерберийского, содержащий сравнение государства с ульем из «Генриха V» (1.2), где списочная структура демонстрирует примеры партонимии как упоминания различных сословий общества («merchants» «soldiers» «singing masons» «mechanic porters» etc.). Противоположные примеры находим в развернутых определениях (в частности, гипотаксисах, сопоставимых с конструкциями так называемого изафета в традиционной поэтике исламской традиции). Так, например, в «Генрихе IV» (3.3) Нортумберленд завершает свой разговор с леди Перси такой идиомой: «'T is with my mind / As with the tide swelled up into the height, / That makes a still-stand, running neither way». Отсюда реконструируется определение [\*still-standing / neither way running / in the height swelled up / tide], компоненты которого составляют списочную структуру, составленную из взаимно дополняемых эпитетов-партонимов. Драматические тексты представляются благодарной почвой для разработки методов автодескрипции при помощи тэгов.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Адмони, В. Г. Грамматический строй как система построения и общая теория грамматики / В.Г. Адмони. – Л. : Наука, 1988.
- 2 Адмони, В. Г. Основы теории грамматики / В. Г. Адмони. – М.: УРСС, 2004 – 104 с.
- 3 Бондарко, А. В. Теория значения в системе функциональной грамматики / А. В. Бондарко. – М. : Языки славянской культуры, 2002. – 736 с.
- 4 Слоўнік тыпу INSITU беларускай мовы / А. І. Дабралюбаў [і інш.] – Мінск. : Беларуская навука, 2003. – 431 с.
- 5 Зубов, А. В. Основы искусственного интеллекта для лингвистов / А. В. Зубов, И. И. Зубова. – М. : Логос, 2007. – 320 с.
- 6 Зубов, А. В. О целях издания и принципах организации параллельного тэггированного русско-белорусского корпуса учебных текстов / Язык и дискурс в статике и

динамике. – Минск : Минский государственный лингвистический университет, 2008. – С. 37–38.

7 Кротевич, Е. В. К вопросу о синтагматическом членении речевого потока / Наукові записки Львівського державного університету ім. І. Франка. Сер. філол. – 1948. – Т. 3. Вип.. 2. – С. 85–107.

8 Мартынов, В. В. Категории языка / В. В. Мартынов. – М. : Наука, 1982 – 192 с.

9 Майзель, С. С. Пути развития корневого фонда семитских языков / С. С. Майзель. – М. : Наука, 1983. – 308 с.

10 Пешковский, А. М. Русский синтаксис в научном освещении / Пешковский А. М. – М. : Госучпедиздат, 1933. – 452 с.

11 Харитончик, З. А., Шавель А. А. Логические основания семантических отношений / Харитончик З. А. // Очерки о языке. – Минск : Минский гос. лингвистический университет, 2004. – С. 117–126.

12 Юдкин-Рипун, И. Н. Дейктическая мотивировка идиоматической лексики в драмах Н. В. Гоголя / Взаимодействие лексики и грамматики в русском языке. – Тамбов : Тамбовский гос. университет, 2009. – С. 217–229.

# КОМПЬЮТЕРНОЕ ОБЕСПЕЧЕНИЕ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ

С. Н. Андреев

## МНОГОМЕРНЫЙ АНАЛИЗ ЯЗЫКОВЫХ И РЕЧЕВЫХ ДАННЫХ

В число важнейших задач лингвистики входит исследование структурной организации системы языка, выявление связей и зависимостей между его элементами и признаками. Система с точки зрения ее организации характеризуется как некоторым набором элементов определенной субстанции (субстанциальный аспект), так и схемой отношений своих элементов (реляционный аспект).

Одним из основных направлений анализа структуры объекта является построение классификации его элементов (единиц и признаков). Осуществляя классификацию, мы создаем конструкт, моделирующий структуру исходных данных. Такой подход рассматривается как самый простой и естественный способ анализа информации, отражающий логику эмпирического исследования.

Понятие классификации и связанные с ней понятия таксономии, систематики и др. являлись предметом неоднократного рассмотрения в работах как общетеоретического плана, так и исследований, направленных на решение задач конкретных областей знаний (работы С. В. Мейна, Ю. А. Шрейдера, С. С. Розова, А. А. Любищива; Р. Р. Сокала П. Х. А. Снита и др.).

Классификационное упорядочение данных занимает важное место в любом лингвистическом исследовании, направленном на анализ конкретного языкового материала. В результате получены детальные описания и общие схемы, отражающие как отдельные участки языка, так и целые уровни и планы языковой системы. Вместе с тем следует признать, что в большинстве лингвистических классификаций проявляются следующие тенденции, реализующиеся в них в различной степени.

1. Во многих классификациях не выдерживается *принцип системности и последовательности* анализа. Это проявляется в том, что классификационная схема признаков не является единой для сопоставляемых классов и не соотносится со всеми объектами, включенными в исследования, т.е. не применяются ко всему исследуемому материалу. Иными словами, имеет место такое положение, когда к одной группе единиц применяются признаки одного вида, а к другой группе – признаки, выделенные на иных принципах. Вполне естественно, что в таких исследованиях полных списков признаков не приводится, а сам анализ приобретает вид комментария.

Кроме того, следует отметить, что часто не учитывается общая представленность признака на всей совокупности материала, из-за чего трудно

судить о диагностической силе этого признака (не ясно, имеется ли он в других классах и насколько часто он там встречается).

2. Большинство лингвистических классификаций не учитывает *реляционного* аспекта систематики – соотношений и связей, существующих между признаками и элементами языковой системы.

По мнению Б. Г. Миркина, проанализировавшего большое количество классификационных задач в биологии, социологии и ряде др. таксономических наук, все они могут быть сведены к двум основным типам, соответствующим двум основным средствам обогащения теоретических представлений о системе (конструирование новых категорий, отражающих существенные стороны системы, и установление новых закономерностей, связывающих различные категории): конструирование признаков и описание одних признаков значениями других [8].

Значение реляционного аспекта в систематике отражается уже в самом определении «естественности» классификационной схемы (ее научной значимости, адекватности описываемых явлений). Считается, что классификация является естественной в том случае, если она позволяет выводить или предсказывать наличие у элемента признаков, исходя из его других базовых характеристик, определяющих место этого элемента в системе [4; 10].

Мерой естественности классификационной системы в этом случае признается ее прогностическая (импликационная) сила, количество существующих связей между элементами, которые она отражает.

3. В абсолютном большинстве случаев систематизация материала основывается на фактах, выделяемых на каком-либо *одном* лингвистическом уровне.

Признание уровней структуры в качестве имманентной сущности языка сделало правомерным и оправданным изучение изолированных подсистем без учета их внешних, межуровневых связей. Однако по мере накопления информации об отдельных уровнях естественно поставить вопрос о переходе от одноуровневого анализа к многоуровневой классификации единиц. Такой подход помогает не только выяснить характер взаимодействия единиц различных подсистем, но и взаимосвязей между подсистемами в целом. С другой стороны, учет межуровневых соотношений позволяет в целом ряде случаев по-новому построить схему отношений и в рамках одной, отдельно взятой подсистемы.

4. Многим лингвистическим классификациям в большой степени свойственен *описательный* метод упорядочения материала. Он заключается в отсутствии математической обработки данных и, как следствие этого, проверке степени достоверности результатов, справедливости получаемых выводов для широкого лингвистического материала.

5. В тех случаях, когда исследователь обосновывает свои выводы не отдельными примерами, а обращается к систематическому анализу материала, в его распоряжении, как правило, оказывается большое количество данных. Обилие статистической информации делает невозможным ее

использование непосредственно как базы классификационного анализа единиц – группировки признаков и объектов. Закономерности структурной организации, отражаемые в эмпирических таблицах, *распыляются на множество фиксируемых связей*. Огромное число индивидуальных наблюдений не позволяет выделить наиболее существенные тенденции. Причем любая попытка углубить исследование вызывает еще большее возрастание признакового пространства, что может привести к тому, что общая картина станет вообще необозримой. Поэтому возникает необходимость обратного сведения этой многообразной информации к небольшому количеству обобщающих выводов [2].

Указанная цель по «сжатию» представленной информации для большого количества разноуровневых признаков, определяющих пространство исследования, может быть достигнуто путем проведения многомерного анализа данных.

В область многомерных задач исследование вступает в том случае, если начинает учитываться совместная вариация как минимум двух признаков для нескольких объектов. Многомерные данные, таким образом, задают значения нескольких переменных для нескольких объектов, то есть  $X_{ij}$ , где  $i$  – соответствует объекту, а  $j$  – переменной.

В рамках коллективной темы, проводимой в Смоленском государственном университете по многомерному анализу языковых и речевых данных, используется подход, с привлечением большого набора (от десятков до сотен) разноуровневых признаков, включающих фонетические, морфемные и деривационные, морфологические, синтаксические, семантические характеристики, а также (при исследовании стиля авторов на стихотворном речевом материале) ритмо-метрические, строфические, рифменные и др. Они выявляются у языковых и речевых элементов на больших массивах текстов или репрезентативных языковых выборках с последующим многомерным анализом полученных баз данных. В качестве основных математических методов анализа используются многомерный дискриминантный анализ, многомерный регрессионный анализ, различные виды кластерного анализа, факторный анализ [1; 3; 5; 11; 12 и др.].

Следует отметить, что применяемый нами подход в целом соответствует методике анализа стилеметрических исследований [6; 7; 9; 14; 15].

Вместе с тем предстоит, по-видимому, еще более углубленно изучать как специфику применения различных конкретных процедур многомерного анализа на языковом материале (выбор исходного списка признаков в качественном и количественном аспектах, алгоритмы сокращения признакового пространства, соотношение различных видов группировок), так и способы интерпретации получаемых разбиений, а также саму стратегию многомерного исследования в целом.

#### СПИСОК ЛИТЕРАТУРЫ

1 Андреев, В. С. Динамика стиля Э. По (На материале лирики) / В. С. Андреев // Известия Российского государственного педагогического университета им. А. И. Герцена. – СПб, 2008. – № 11 (72). – С. 168–174.

- 2 Дюран, Б. Кластерный анализ / Б. Дюран, П. Одел. – М. : Статистика, 1977.
- Баевский, В. С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы/ В.С. Баевский. – М. : Языки славянской культуры, 2001.
- 3 Забродин, В. Ю. О критериях естественности классификаций / В. Ю. Забродин // Научно-техническая информация. ВИНТИ. Сер.2. № 8. – 1981. – С. 22–24.
- 4 Кучер, И. Н. Структурно-семантические признаки именных образных моделей в английском языке (на материале образной системы А. Теннисона). Автореф. дис. ... канд. филол. наук/ И. Н. Кучер; Смоленский государственный университет. – Смоленск, 2006.
- 5 Мартыненко, Г. Я. Основы стилеметрии / Г. Я. Мартыненко. – Л. : Изд-во Ленинградского ун-та, 1988.
- 6 Марусенко, М. А. Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов / М. А. Марусенко. – Л. : Издательство Ленинградского ун-та, 1990.
- 7 Миркин, Б. Г. Анализ качественных признаков и структур / Б. Г. Миркин. – М. : Статистика, 1980.
- 8 Тулдава, Ю. Опыт квантитативного анализа художественного стиля / Ю. Тулдава // *Studia metrica et poetica*. Tartu, 1976. – С. 122–141.
- 9 Шрейдер, Ю. А. Типология как основа классификации/ Ю. А. Шрейдер // НТИ ВИНТИ. – Сер.2. М., 1981. – № 11. – С.1–5.
- 10 Andreev, S. Evolution of style. Young and versatile periods of F. I. Tutchev/ S. Andreev // *Glottology. International Journal of Theoretical Linguistics*. – 2009. V. 2, N 1. – P. 1–11.
- 11 Andreev, S. Lermontov: Dynamics of style/ S. Andreev // *Studies in Quantitative Linguistics 5: Issues in Quantitative Linguistics* / Ed. Reinhard Köhler. – RAM-Verlag, 2009. – P. 1–9
- 12 Andreev, V. A Multivariate Study of Style Differences in Poetry/ V. Andreev // *The State of Stylistics* / Ed. Greg Watson. – Amsterdam-New York: Rodopi, 2008. – P. 177–190.
- 13 Holmes, D. The Evolution of stylometry in humanities scholarship/ D. Holmes // *Literary and Linguistic Computing*. 1998. – Vol.13, N.3. – P. 111–117.
- 14 Hoover, D. L. Frequent collocations and authorial style/ D.L. Hoover // *Literary and Linguistic Computing*. 2003. Vol.18. – P. 261–286.

В. Ю. Бойка

## **СТЫЛІСТЫЧНЫЯ ФУНКЦЫЯ АКАЗІЯНАЛІЗМАЎ У ТЭКСТАХ БЕЛАРУСКАМОЎНЫХ БЛОГАЎ**

Аказіянальнае словаўтварэнне ўяўляе сабой надзвычай цікавы лінгвістычны феномен. Яно адлюстроўвае словаўтваральны патэнцыял мовы, спрыяе выражэнню разнастайных стылістычных адценняў. Аднак з пункту гледжання шараговага носбіта мовы (актыўнага або пасіўнага) аказіянальнае словаўтварэнне – гэта, перш за ўсё, спосаб індывідуалізацыі маўлення, а таксама рэалізацыі крэатыўнага пачатку. Таму цалкам заканамерна, што ў стылістыцы беларускай мовы аказіяналізмы, або індывідуальна-аўтарскія неалагізмы, звычайна разглядаюцца як элемент

мастацкага маўлення. Але рэальная моўная сітуацыя гаворыць аб тым, што шырокае выкарыстанне аказіяналізмаў магчыма не толькі ў тэкстах мастацкага стылю. Так, аўтары манаграфіі “Русская разговорная речь” яшчэ ў канцы 70-х гг. XX ст. адзначалі характэрнасць “неузуальнага словаўтварэння” для гутарковага маўлення [1, с. 27]. У беларускай мове на сённяшні дзень таксама назіраецца досыць актыўнае выкарыстанне аказіянальнай лексікі ў сферах зносін, не звязаных з літаратурнай творчасцю. Яскравым прыкладам таму з’яўляецца беларускае інтэрнэт-маўленне, прадстаўленае ў тэкстах блогаў (інтэрнэт-дзённікаў). Варта адзначыць, што аказіянальныя адзінкі ў тэкстах блогаў характарызуюцца неаднароднасцю. Яны могуць узнікаць як у выніку рэалізацыі словаўтваральных магчымасцяў мовы, так і з парушэннем патрабаванняў моўнай сістэмы. Таму стварэнне класіфікацыі, якая ахапіла б усе аказіяналізмы, што ўзнікаюць ў межах інтэрнэт-жанраў, пакуль уяўляецца малаверагодным. Аднак аналіз падобных адзінак паказаў, што мы можам выявіць агульныя словаўтваральныя мадэлі, уласцівыя маўленню розных блогераў, але пры гэтым не заўсёды ўласцівыя беларускай літаратурнай мове.

Адной з найбольш выразных тэндэнцый з’яўляецца выкарыстанне ў тэкстах блогаў адносных і прыналежных прыметнікаў, утвораных ад абрэвіятур:

*Таму перад Захадам зараз стаіць даволі прагматычная задача: выбраць, як зінтэрпрэтаваць справаздачи ад **абсеісных** назіральнікаў і які мэссадж паслаць афіцыйнаму Мінску; Падазраю, што многім мясцовым нацыяналістам вельмі ўжо хочацца адкінуць нікому не цікавыя **бнраўскія** казачкі і легітымізаваць Вайсрутэнію; Прывезла ў Менск куртку й красоўкі й ўжо пасьпела сьвяточна адкрыць сезон бегу на **бнтушным** стадыёне; Учора з працы усіх пагналі на нейкую **СНДоўскую** канферэнцыю у прэзідыум НАН.*

Пры дапамозе аналагічных суфіксаў падобныя прыметнікі ўтвараюцца і ад іншамоўных неалагізмаў, а таксама ад некаторых уласных назоўнікаў:

*У **айк’юшнай** групіроўкі 12 сьнежня касьмічная паці; На біфрах было супер. мора ўражаньняў. З трафэяў – **біфрышная** маечка, львоўскі брэлок, дзьве магнітныя налепкі на лядоўню; Фоткі зь **вікендаўскае** вандроўкі ў горы; Відавочна, што кожнае **анімэішная** дзяўчо ня лічыць сэкс нечым брудным ці ненатуральным, бо самая гэтая дэкламацыя сэксуальнасці, як нармалёвасці, як неабходнасці ці адпаведнасці сусветнаму стану, намалавана ў іхніх светлых вочках; Натуральна, я не магла прайсці моўчкі міма. падыходжу да цёценькі, якая прадае нашую **белітаўскую** касметыку, і кажусь, маўляў, так і так, у вас надпіс з памылкай.*

Ад абрэвіятур і ўласных назоўнікаў са значэннем арганізацыі, установы ўтвараюцца і назоўнікавыя аказіяналізмы са значэннем асобы:

**Фбравец:** “Быў тут у нас адзін Карл, які нам вельмі не падабаўся... Прозвішча Маркс у яго было” (галоўнага героя таксама клічуць Карл);

Гаспада знатакі – прыкольныя чэлы. Я колісь хацеў напісаць пра іх (**шідкашнікаў**) крэатыў-разьвянчаньне, але раптам зразумеў, што некаторыя з тых прэтэнзіяў, што я да іх маю, я б мог мець і да сябе любімага; у калідорах “мінскага форуму” бачыла як мінімум 4 журналістаў + за спіной сядзяць **бтшнікі**; Так вось і пляскаюць у аладкі “**віцязеўцы**” седзячы на унітазах; Як пішуць самі **бі-бі-сішнікі**, Этом человек открыл пакет, ознакомился с его содержимым и передал находку Би-би-си.

Змяняецца і набор словаўтваральных афіксаў. Так, некаторыя карані (або часткі каранёў) выкарыстоўваюцца ў якасці прыставак: бел- (у значэнні ‘беларускі’), **быдла-** і інш.

Толькі сёння зьвярнуў увагу на тое, што з 12-ці маіх аднаклясьнікаў толькі адзін (!!!) служыў у **белвойску**, астатнія так ці інакш “адкасілі”... ці то плакаць, ці то сьмяяцца, нават і ня ведаю...; Сіджу тут, каля кампа, літаратуру сабе шукаю. Усё адно прыйдзецца кнігі набываць, бо ініцэрнэт, што да **белтэорыі**, не вельмі такі карысны; Купіў таўсценную кнігу “карона Вітаўта вялікага”. Ціна зборнік **белдэтэктываў**; Для азнаямленьня – пару матэр’ялаў: пазіцыя беларускамоўных і тых, хто не прымае **белмову** як сродак зносінаў; Мяне хапіла толькі на фоты на пл. Перамогі, далей ісьці на **быдламаскарад** жаданьня не было ніякага; **Быдла-аматар** пасля фінальнага свістка некалькі хвілін трымаў выцягнуты ўверх сярэдні палец у бок галандскіх футбалістаў...таксама напэўна прыкмета...; Па танных гучнагаварыцелях, прыкручах іржавымі цвікамі да тэлеграфных слупоў ужо цэлы дзень перашкаджае патрыятычная **быдла-папса**; Потныя працаўнікі наспех збіраюць металічныя шкідеты немудрагелістых канструкцый, рыхтуючыся да чарговага **быдла-свята**; Жахлівае спалучэньне неінтэлегэнтнай манеры “гаварыць” гучна з выкарыстаньнем расейскіх **быдла-славечак** “ціпа чо”.

Даволі актыўна ўтвараюцца прыметнікавыя і назоўнікавыя аказіяналізмы пры дапамозе прыставак іншамоўнага паходжаньня. Некаторыя прэфіксы ўжываюцца і ў літаратурнай мове, а новыя лексічныя адзінкі ўтвараюцца па агульнамоўных словаўтваральных узорах. У прыватнасці, гаворка ідзе аб прыстаўцы **анты-**:

Эстэт, беларучка і **антымачо**. Жэню цяжка ўявіць шахцёрам ці сантэхнікам; Пазыкі, праблемы, ад якіх людзям станавілася жудасна, бязьмежныя ахвяры, толькі каб моладзь не згуртоўвалі алкаголь, наркатыкі і **антыбеларускасьць**; Я люблю боль ва ўсіх формах і праяўленьнях: страх, пакуты і **антынастрой**; 10 год узмоцненага рэжыму на падставе “паказаньняў” людзей, са словаў каторых перад кожнай масавай акцыяй усе раптам пачынаюць лаяцца матам, сцацца на ўсё і на ўсіх ды махаць рукамі, выкрыкваючы **антыфігведаеякія** лозунгі!

Калі ў літаратурнай мове выкарыстанне іншамоўных прыставак звычайна мае абмежаваны характар, то ў тэкстах блогаў некаторыя запазычаныя прэфіксы становяцца актыўным словаўтваральным сродкам. Яскравым прыкладам з’яўляюцца прыстаўкі **мега-**, **супер-**, **гіпер-**:



Дарэчы, калі ўжо на тое пайшло – трымайце **мегаэратычную** фотку маёй персоны; Піва праставім і **мега-удзячныя** будзем!; Я проста **мега-мега-мега шчаслівая**, хоць мы й ледзь знаёмыя зь ім; Я, відаць, буду надзвычай, **гіпер-, супер-, мега-** (патрэбнае падкрэсліць) **багатая**; Нягледзячы на некаторыя непрыемнасці майго жыцця апошніх дзён, у ім трапляюцца некаторыя **мегастаноўчыя** падзеі; Працягну чытаньне **суперлітаратуры**, бяз якой я сапраўды стала занадта добрай апошнім часам; Да новых **супербоцікаў** праўда торбачкі нідзе ня бачыла – але штосьці прыдумаю; Як жа я так – сёння ўранку слухала афрыканскі хіп хоп з нігерыі з новымі калонкамі – **суперэкзатычная** музыка і **супергучныя** калонкі – дзень не такі ўжо і ніякі; Прыкольныя гэтыя італьянцы... Нават люблю іх за такую эмацыйную **гіпер-шчырасць**...

У тэкстах беларускамоўных блогаў таксама шырока ўжываюцца аказіянальныя назоўнікі і прыметнікі, утвораныя спосабам слова- і асноваскладання:

Бедныя студэнты, яны відаць ніколі не траплялі на такіх **выкладчыкаў-раздалбаеў** як я; Першы раз у жыцці мне прысвяцілі верш! (Першы калі не лічыць былых вольгіных школьных **вершаў-сцёбаў** дзе я прадстаўлена не ў самых лепшых ролях); Фамільныя надгробкі баронаў ператварылі ў забаву для **гультаёў-турыстаў**; Рацыянальна цікавей быў бы Маккейн, бо мадэль свету пры ім была б нашмат болей **амерыканска-прадказальная**; Мой скрыпучы, **брыдкавадкасны** глузд зараз гэную інфу перарабляе ў сны, я гэта буду сніць, шляхам сну; Фішка тут нават не ў самім факце, а ў песнях а-ля пачатак дзевяностых і нешта з **брутальна-дваравога** (“а ты опята сегодняя не прішла...”); Насамрэч гэта дзіка цяжкія тыдні. І што будзе на наступным (пасля **гістарычнабілетных фізычназаконных геамэтрычнатэарэмных геаграфічнагенацыдных і шчэ чорт ведае якіх ядзернабяссоннаскладаных выходных**) – страшна ўяўляць.

Калі ў літаратурнай мове звычайна адбываецца складанне дзвюх асноў, то ў інтэрнэт-маўленні іх можа быць тры і болей: ...палова беларускіх “патрыётаў” трымае мэйлбоксы на mail.ru, а тут вэрхал узняўся з нагоды таго, што **руско-мядзьвез’я-азіяцкая** кантора перакупіла жж.

Дзеяслоўнае аказіянальнае словаўтварэнне адрозніваецца большай складанасцю і разнастайнасцю [2, с. 4]. Але аналіз інтэрнэт-маўлення паказвае, што сярод неузואльных дзеясловаў найбольш пашыраным з’яўляецца прэфіксальны спосаб утварэння:

Я тут збіраюся ў бліжэйшыя часіны вас весяліць. А пакуль **адзлююся**... (карцінка доўга чакала на пару, а тут факт блізкі па “тэме” зьявіўся.); Ну а паколькі форма будынку крыху змянілася ў параўнаньні з першапачатковай, ды й рэжымку офісаў даваць ужо хутка трэба, мне было даручана **адмадэліць** будынак; А яшчэ за гэты час я патраціла “хренову кучу бабок”, пабіла фоцік і купіла новы, пару разоў закахалася і **выкахалася**; Прапаноўваю **заняўмець**: што для вас ёсьць здрадай? і ці існуе яна ўвогуле? ці можна яе дараваць?; Калі мы няверныя голасу сумленьня, наша душа становіцца “крывой”, яна

*дэфармуецца і занячышчваецца; Яна знацца наасацыіравала нешта пра мяне – а мне аддувайся, што гэтая фігня значыць; Цікава, хто здагадаецца, што гэта я **напраектаваў**, а мой сантэхнік гераічна намантаваў?*

Аказіянальнае словаўтварэнне, такім чынам, з’яўляецца адной з характэрных прымет жанру блога. Аказіяналізмы ў інтэрнэт-маўленні адрозніваюцца сваёй падкрэсленай “немастацкасцю”, яны выконваюць іншыя стылістычныя функцыі: 1) спрыяюць эканоміі моўных сродкаў (*вікендаўская вандроўка – вандроўка, якая адбылася ў вік-энд*); 2) індывідуалізуюць аўтарскае маўленне; 3) запаўняюць “лакуны” ў словаўтваральнай сістэме (*закахацца – выкахацца*); 4) выражаюць эмацыянальнае стаўленне аўтара да прадмета гаворкі.

#### СПІС ЛІТАРАТУРЫ

1 Земская, Е. А. Русская разговорная речь: лингвистический анализ и проблемы обучения / Е. А. Земская. – М. : Рус. язык, 1979. – 240 с.

2 Барковіч, А. А. Лексічны патэнцыял беларускай мовы ў кантэксце корпуснай лінгвістыкі (на матэрыяле новай дзеяслоўнай лексікі): аўтарэф. дыс. ... канд. філал. навук: 10.02.01 / А. А. Барковіч; Беларус. дзярж. ун-т. – Мінск, 2008. – 17 с.

Т. В. Бусел

### **СТРУКТУРА СИСТЕМЫ АВТОМАТИЧЕСКОГО ПОРОЖДЕНИЯ АНГЛОЯЗЫЧНЫХ ДЕЛОВЫХ ЭЛЕКТРОННЫХ ПИСЕМ**

В эпоху развития информационных и коммуникационных технологий и их внедрения во все сферы человеческой деятельности проблема автоматического порождения текстов приобрела особую актуальность [1; 2; 3], в связи со значительным увеличением количества информации на английском языке, представляемой в виде электронных документов.

Обращение к англоязычным текстам делового характера диктуется потребностью современного общества в создании интеллектуальных компьютерных систем, призванных помочь специалистам, работающим в сфере международных экономических и внешнеторговых отношений, значительно сократить время, необходимое на их создание, снизить вероятность появления ошибок, вызванных недостаточно высоким уровнем владения английским языком и техническими погрешностями.

Актуальность современных исследований в данной области также связана с активизацией в Республике Беларусь предпринимательской деятельности, расширением и углублением деловых контактов как внутри страны, так и на международном уровне, формированием новых форм деловой коммуникации, вызванных развитием глобальной сети Интернет, необходимостью быстрого ввода, обновления и обмена текстовой информацией в режиме реального времени.

Таким образом, проблема создания эффективных методов автоматического порождения, современного алгоритмического и лингвистического

обеспечения, на базе которых может быть разработана компьютерная система, позволяющая автоматически создавать различные типы англоязычных деловых электронных писем по заданной пользователем теме, имеет важное практическое значение и является одной из актуальных задач компьютерной лингвистики.

В Минском государственном лингвистическом университете на кафедре информатики и прикладной лингвистики была разработана *система автоматического порождения англоязычных деловых электронных писем*, общая архитектура которой включает две части:

- *ресурсы порождения*, которые описывают знания необходимые для синтеза текстов, которые представлены в виде лингвистической базы знаний;
- *обрабатывающий компонент* применяет эти знания к входным данным, он реализуется как интерпретатор ресурсов порождения и организатор всего процесса работы системы.

Лингвистическая *база знаний* системы порождения деловых электронных писем включает:

- *электронный словарь*, в котором в закодированном виде содержится подробное описание грамматических и семантических характеристик каждой лексической единицы;
- *логико-семантические формулы* позволяют моделировать процесс развертывания содержания порождаемых системой текстов англоязычных деловых электронных писем;
- *семантико-синтаксические формулы* текстов на формальном языке СЕМСИНТ содержат сведения об абзацах с определенным предметно-логическим содержанием, которые входят в их состав. На основании семантико-синтаксических формул происходит построение синтаксической структуры предложений, а также структурирование всего текста;
- *базы данных согласования лексических единиц*, в которых зафиксированы характерные для официально-делового стиля нормы лексической сочетаемости.

В системе автоматического порождения все уровни лингвистического описания (семантический, лексический и грамматический) тесно связаны друг с другом, поэтому отношения такой сложности тяжело поддаются алгоритмизации, так как очень трудно вскрыть все зависимости и обеспечить их координацию в системе. Для облегчения задачи создания системы порождения текстов, процесс синтеза текста упрощается до последовательного применения модулей.

Описываемая система порождения деловых электронных писем включает три основных модуля: *модуль планирования содержания*, *модуль языкового оформления* и *модуль редактирования переменных данных в тексте*.

*Модуль планирования содержания* – это комплекс алгоритмов по определению и структурированию содержания будущего текста, который состоит из двух подмодулей, работающих в тесном взаимодействии.

Первый подмодуль определяет, какая информация будет участвовать в синтезируемом тексте, а второй – порядок следования информации в документе. Результатом работы первого модуля должно стать логико-семантическое представление будущего текста.

*Модуль языкового оформления* происходит выбор лексических и грамматических средств, представленных в базе знаний системы, которые необходимо использовать, чтобы выразить на английском языке содержание, сформированное в первом модуле. На данном этапе порождения осуществляется выбор лексических единиц, а также их грамматических форм, выполняется морфологическое согласование между членами грамматических групп, а также соответствие выбранных слов структурно-семантическим и лексико-семантическим правилам сочетаемости и т. д. Таким образом, данный модуль переводит абстрактное представление, построенное предыдущим модулем, в текст на английском языке.

В *модуле редактирования переменных данных* происходит окончательное оформление содержания текста, подстановка в текст делового письма таких данных, таких как вступительное обращение, дата, сумма, количество и т. д.

На рисунке 1 представлена *структурно-функциональная схема*, которая отражает общую картину преобразований в созданной системе порождения.



Рисунок 1 – Структурно-функциональная схема порождения

Описываемая в данном докладе архитектура системы автоматического порождения англоязычных деловых электронных писем обеспечивает

координацию всех уровней лингвистического описания английского языка, которые тесно связаны друг с другом и с трудом поддаются алгоритмизации, и, как показала практика, является весьма эффективной.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Жинкин, Н. И. О кодовых переходах во внутренней речи / Н. И. Жинкин // Жинкин Н. И. Язык – речь – творчество. – М.: Лабиринт, 1998. – С. 146–162.
- 2 Кацнельсон, С.Д. Речемыслительные процессы / С. Д. Кацнельсон // Вопросы языкознания. – 1984. – № 4. – С. 1–12.
- 3 Человеческий фактор в языке: Язык и порождение речи / Отв. ред. Е. С. Кубрякова. – М.: Наука, 1991. – 239 с.

И. И. Бутякова

### ИСПОЛЬЗОВАНИЕ КОМПЬЮТЕРНОЙ ПРОГРАММЫ «WORDSTAT» ПРИ ПОЛЕВОМ ИССЛЕДОВАНИИ ПОЭТИЧЕСКОГО ТЕКСТА

Информационные технологии последнее время приобрели важное значение не только для точных наук, но и для гуманитарных. В частности, современные лингвистические исследования зачастую используют компьютерные обеспечение.

В качестве примера остановимся на характеристике компьютерной программы, которая используется для подсчета частотности слов в тексте – WordStat. Ценность данной программы в том, что в автоматическом режиме можно вывести количественные показатели употребления той или иной лексемы в тексте, а это крайне важно при работе с поэтическим текстом, так как фактический материал довольно большой, а подсчет указанных показателей по сборникам стихотворений крайне сложен.

Современные лингвистические исследования, связанные с полевым подходом к изучению поэтического текста, базируются на частоте употребления той или иной лексемы в тексте. При построении поля необходимо определить его ядро (лексемы, количество употреблений которых достаточно велико) и периферию (лексемы, количество употреблений которых ограничено). Для того чтобы поле было представлено достаточно объективно, необходимо определить частотность употребления лексем. Под частотностью в данном случае понимается «показатель, характеризующий частоту употреблений чего-нибудь» [1, с. 878]. Для этого лучше всего использовать программу WordStat.

При работе с указанной программой необходимо поэтический текст перевести в текстовый редактор «Блокнот», который используется для просмотра и редактирования текстовых (ТХТ) файлов. Затем в окне программы можно выбрать файл для обработки, настройку подсчетов (накапливать сумму результатов, объединять похожие слова, считать только больше одного) и сортировать результаты по алфавиту и частоте, также можно показать сводку и автоматически открыть файл результатов.

С такой программой определение частотности той или иной лексемы занимает считанные секунды, при этом мы получаем точный расчет с указанием количественного показателя употребления той или иной лексемы. Это позволяет минимализировать работу с текстом и получить объективные результаты со ссылкой на компьютерные данные.

При помощи данной программы можно по частотности употребления лексемы определить центр (ядро) и периферию того или иного лексико-семантического поля.

Примером такого инновационного подхода к изучению творчества поэтов может служить анализ ЛСП «горе» в творчестве поэтов начала 20 века. При определении ядра и периферии поля следует отметить, что «большинство лингвистов признает в структуре концепта неизменную (ядерную) часть и периферийную «социопсихокультурную» часть – зыбкую, размытую, включающую оценки, интерпретации, эмоции коннотации, связанные с пониманием концепта, ассоциации, образы, символы, возникшие на базе концепта, т. е. часть, в определенной степени содержащую потенциальную субъективность» [2, с. 36]. Ядро исследуемого ЛСП состоит из следующих лексем:

1) *горе*: *Перед этим горем гнутся горы,/ Не течет великая река,/ Но крепки тюремные затворы,/ А за ними «каторжные норы»* (Ахматова); *Здесь на всем оттенок давний, Горе всюду прилегло* (Цветаева); *И совсем настоящее горе Застилало слезами глаза* (Гумилев).

Ядерная лексема *горе* в поэтических текстах А. Ахматовой, М. Цветаевой, Н. Гумилева и О. Мандельштама ассоциируется с одушевленными и неодушевленными объектами:

– с любимым человеком: *В небо, чистое как горе,/ Глаз твоих,/ В пещу сказочного моря/ Рук твоих?* (Гумилев); *И от красавиц тогдашних – от тех европейок нежных –/ Сколько я принял смущенья, насады и горя!* (Мандельштам), что свидетельствует о прямой оппозиции к лексеме *счастье*;

– с близким или родным человеком: *Перед этим горем гнутся горы,/ Не течет великая река,/ Но крепки тюремные затворы,/ А за ними «каторжные норы»* (Ахматова), что свидетельствует о взаимосвязи судеб людей и страны, в которой они живут;

– со смертью: *Было горе, будет горе,/ Горю нет конца,/ Да хранит святой Егорий/ Твоего отца* (Ахматова), что вполне соответствует традиционному представлению о горе;

– с войной, поскольку в судьбах поэтов начала 20 века революция и две войны сыграли свою значимую роль, отодвинув на второй план все личные переживания: *Во мне еще, как песня или горе,/ Последняя зима перед войной* (Ахматова); *Слова: «Встает великий Рим,/ Берите ружья, дети горя...»* (Гумилев).

Обращает на себя внимание и тот факт, что в поэтических текстах начала 20 века в одном контексте сосуществуют обе оппозиционные лексемы (*счастье/ горе*): *Не пойму, от счастья или горя/ Плачешь ты, к моим ногам припав// Я друзьям моим сказала:/ «Горя много, счастья мало»* (Ахматова);

В декабре на заре было счастье,/ Длилось – миг./ Настоящее, первое/счастье/ Не из книг!/ В январе на заре было горе,/ Длилось – час./ Настоящее, горькое горе/ В первый раз! (Цветаева). Приведенные примеры позволяют сделать вывод о том, что граница между счастьем и горем настолько незначительная, что порой их невозможно разделить и даже определить значение каждого в отдельности порой затруднительно.

2) беда ‘то же, что несчастье’. Данная лексема в поэтических текстах начала 20 века ассоциируется, как и ядерная лексема горе, с одушевленными и неодушевленными объектами:

– с любимым человеком: *Я спросила: «Чего ты хочешь?»/ Он сказал: «Быть с тобой в аду»./ Я смеялась: «Ах, напророчишь/ Нам обоим, пожалуй беду»* (Ахматова); *Это не я с тобою, –/ Это беда с бедою/ Каторжная – сошлась* (Цветаева);

– с близким или родным человеком: *Брат по песенной беде –/ Я завидую тебе* (Цветаева);

– со смертью: *Затем и насылает беды/ Бог, что живой меня на небо/ Взять замышляет за труды* (Цветаева);

– с войной: *Еще на всем печать лежала/ Великих бед, недавних гроз, –/ И я свой город увидала/ Сквозь радугу последних слез* (Ахматова); *Как в страшный год, возвышены Бедою,/ Ты – маленькой была, я – молодую* (Цветаева).

Обращает на себя внимание тот факт, что в некоторых текстах объединяются обе ядерные лексем, что подтверждает данные компьютерных исследований и что позволяет сделать вывод о равнозначном делении ядра на две равноправные лексем: *Была над нами, как звезда над морем,/ Ища лучом девятый смертный вал,/ Ты называл ее бедой и горем,/ А радостью ни разу не назвал// Вот и берег северного моря,/ Вот граница наших бед и слав, –/ Не пойму, от счастья или горя/ Плачешь ты, к моим ногам припав* (Ахматова).

С помощью программы Wordstat мы получили следующие данные относительно частотности употребления лексем, относящихся к ядру и периферии ЛСП «горе»:

Таблица 1

Лексемы Авторы	Горе	Беда	Грусть	Печаль	Несчастье	Мўка
А. Ахматова	78	46	44	38	2	18
М. Цветаева	52	48	96	37	16	100
Н. Гумилев	58	16	10	37	6	12
О. Мандельштам	4	25	12	16	5	10

Компьютерная программа WordStat позволяет определить и периферийные лексем, которые сохраняют проявляющееся в контексте значение, сближающее их с ядром анализируемого ЛСП:

1) *несчастье* ‘горестное событие’: *Настанет и вашим несчастьям черед!// Над вами собираются тучи!* (Гумилев); *Сохрани мою речь навсегда за привкус несчастья и дыма* (Мандельштам). Эта лексема представлена в ограниченном количестве контекстов, причем данное чувство связано не с любимым человеком, а с третьим лицом или неодушевленным объектом;

2) *грусть* ‘чувство печали, уныния’: *С хлебом ем, с водой глотаю/ Горечь – горе, горечь – грусть* (Цветаева); *Он длится без конца – янтарный, тяжкий день!// Как невозможна грусть, как тщетно ожиданье!* (Ахматова), причем это чувство имеет определенный временной промежуток, зачастую не очень длительный;

3) *мука* ‘сильное физическое или нравственное страдание’: *Если за горькое счастье и сладкую муку свою/ Принят он в сине-хрустальном высоком господнем раю* (Гумилев); *Кто скажет, что гитане гибкой/ Все муки ада суждены?* (Мандельштам). Употребление данной лексемы связано с усилением более «яркого» чувства – горя, счастья, ненависти и др.;

4) *печаль* ‘чувство грусти, скорби, состояние душевной горечи’: *Будь вечно с ним: пусть верности научат/ Тебя печаль его и нежный взор* (Цветаева); *Невыразимая печаль/ Открыла два огромных глаза* (Мандельштам), что вносит дополнительный оттенок в тяжелые душевные переживания поэтов, ведь еще Н. Н. Болдырев заметил, что «если ядерная зона включает общий, универсальный опыт познания, то периферия затрагивает личный, субъективный опыт и отражает то, как мы непосредственно познаем окружающий мир и как мы с ним взаимодействуем» [3, с. 36].

В результате можно сделать вывод о том, что компьютерное обеспечение лингвистических исследований (в данном случае – программа статистической обработки текста WordStat) позволяет значительно повышать степень достоверности получаемых выводов.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Ожегов, С. И. Толковый словарь русского языка / С. И. Ожегов, Н. Ю. Шведова. – Москва, 2003. – 944 с.
- 2 Болдырев, Н. Н. Курс лекций по английской филологии / Н. Н. Болдырев. – Тамбов : Изд-во Тамб. ун-та, 2001. – 123 с.
- 3 Маслова, В. А. Когнитивная лингвистика: учебное пособие/ В. А. Маслова. – Минск. : ТетраСистемс, 2004. – 256 с.

С. А. Важнік

#### ВЕТЛІВЫЯ ЗВАРОТЫ Ў БЕЛАРУСКІМ СЕЦІВЕ

1. *Інтэрнэт*, або *сеціва* (тэрмін С. Шупы) як аб’ект лінгвістычнага даследавання абраны не выпадкова. Па-першае, сеціва – люстэрка таго, што адбываецца ў нашым грамадстве і з нашай мовай. Па-другое, беларускі сегмент сусветнага павуціння – гэта тое месца, дзе беларуская мова не стрымліваецца жорсткімі нарматыўнымі рэгламентацыямі, дзе ніхто і нішто не замінае яе натуральнаму развіццю. Іншымі словамі, у



байнэце беларуская мова з поспехам выконвае функцыі не толькі *мовы культуры*, але і *мовы камунікацыі*. Спецыфічнасць жа моўнай сітуацыі на Беларусі ў тым, што шырокія масы ўжо амаль страцілі сваю *мову культуры* з усёй ідыяматыкай і адметным сінтаксісам і ствараюць культуру на аснове другой дзяржаўнай мовы.

Сімптомы “страчанай мовы культуры” ўсё часцей і часцей праяўляюцца сёння ва ўсіх сферах. Гэта, на жаль, ужо рэальны факт: беларусы не ведаюць сваёй фразеалогіі, ці больш шырока, сваёй нацыянальнай ідыяматыкі. Адзначанае вышэй непасрэдна тычыцца і этыкетных формул. У такім “становішчы рэчаў” мы змаглі пераканацца на ўласным вопыце падчас правядзення ўступнай рэспубліканскай алімпіяды па беларускай мове і літаратуры “*Абітурыент Ліцэя БДУ-2010*”. Абітурыентам было прапанавана заданне прывесці адпаведнікі да рускіх этыкетных формул: *Как дела?; Приятного аппетита!; Прошай!; Спокойной ночи!; Вам идёт это платье!; Пожалуйста!*. Бездакорна справіліся з заданнем толькі 3 % абітурыентаў, толькі 8 асоб з 250 (!) прывялі да першай этыкемы правільныя адпаведнікі: *Як маешся?; Як жывяце?; Як жыццё?*. У асноўным жа прыводзілася калька *\*Як справы?* або няправільныя варыянты тыпу *\*Як дзяла?; \*Як дзялы?* і пад. Губляючы сваю нацыянальную сінтаксічную ідыяматыку, беларусы губляюць сваю нацыянальную моўную ідэнтычнасць.

2. Ветлівыя звароты – гэта своеасаблівая візітоўка пэўнага інтэрнаўта (ці карыстальніка сеціва) пры ўваходзе ў чат ці на форум. Станаўленне і дынаміка сістэмы ветлівых зваротаў адбывалася ў беларускай літаратурнай мове ў пачатку – першай палове ХХ ст. Таму нам цікава было параўнаць тыя ветлівыя звароты, што выкарыстоўваюцца сучаснымі карыстальнікамі сеціва, з этыкемамі, што выкарыстоўваліся ў эпістальярнай практыцы беларусаў перыяду фарміравання новай беларускай літаратурнай мовы і Першага нацыянальнага адраджэння.

2.1. У 10–20-я гады ХХ ст. у беларускай мове склалася ўнікальная сістэма базавых этыкетных мадэляў. Як паказвае матэрыял, у канцы ХІХ ст. – пачатку ХХ ст. некалькі разоў змяняліся акцэнтны ў афіцыйных эпістальярных паводзінах беларусаў: на змену інтэлігенцкім польскім *пане / паночку* (канец ХІХ ст. – 10-я гады ХХ ст.) прыходзяць уласна беларускія народныя *дзядзька / цётка* (20-я гады ХХ ст.), якія ў сваю чаргу ў 30-я гады ХХ ст. выцясняюцца стылістычна-маркіраванымі канструкцыямі з вакатывам *таварыш / таварышы* (гл. табліцу) і рускай інавацыяй з імем па бацьку:

Пры гэтым заўважаюцца наступныя тэндэнцыі: (1) выразны польскі ўплыў, а таму адпаведна ўжыванне польскіх этыкетных мадэляў (інтэлігенцкая лістоўная традыцыя); (2) уплыў рускай мовы (пачатак савецкай лістоўнай традыцыі); (3) тэндэнцыя да нацыяналізацыі, што звязана з “адштурхоўваннем” як ад польскіх, так і ад рускіх мадэляў (уласна беларуская народная традыцыя). Традыцыі (1), (3) складаюць нашаніўскую лістоўную традыцыю.

Табліца 1

<i>пане / паночку / паночки</i> (канец XIX ст.–10-я гг. XX ст.)	<i>дзядзька / цётка</i> (20-я гг. XX ст.)	<i>таварыш / таварышы</i> (30-я гг. XX ст.)
1) вакатыў <i>пане</i> ; вакатыў <i>паночку / паночки</i>	1) вакатыў <i>дзядзька / цётка / дзядзькі</i>	1) вакатыў <i>таварыш / таварышы</i>
2) фатычны маркёр + <i>пане</i> ; фатычны маркёр + <i>паночку</i>	2) фатычны маркёр + <i>дзядзька / цётка / дзедка</i>	2) фатычны маркёр + <i>таварыш</i>
3) <i>пане</i> + імя / прозвішча	3) <i>дзядзька / цётка</i> + імя / прозвішча	3) <i>таварыш</i> + імя / прозвішча
4) <i>пане</i> + пасада / званне / род заняткаў	–	4) <i>таварыш</i> + пасада / званне / род заняткаў
5) фатычны маркёр + <i>пане / пані / панна</i> + імя / прозвішча	4) фатычны маркёр + <i>дзядзька / цётка</i> + імя / прозвішча	5) фатычны маркёр + <i>таварыш</i> + імя / прозвішча
6) фатычны маркёр + <i>пане</i> + пасада / званне / род заняткаў	–	6) фатычны маркёр <i>таварыш</i> + пасада / званне / род заняткаў
7) фатычны маркёр + <i>пане</i> + званне + прозвішча	–	–
8) фатычны маркёр + <i>пане / пані</i> + <i>мосці</i> + <i>дабрадзею</i>	–	–

2.2. Ветлівыя звароты сучасных карыстальнікаў сеціва выяўляліся на электронным рэсурсе *by\_tova* / “Беларуская мова ў Інтэрнэце” на *LJ (Live Journal)*: праведзены зрэз за год (з мая 2009 г. – па май 2010 г.), у выніку чаго зафіксаваны 82 ужыванні ветлівых зваротаў.

Разгледзім найбольш прадуктыўныя формулы, сярод якіх мэтазгодна вылучаць традыцыйныя і арыгінальныя этыкемы.

2.2.1. Традыцыйныя ветлівыя звароты ў маўленні карыстальнікаў рэсурса *by\_tova*:

Табліца 2

<b>Ветлівы зварот</b>	<b>Колькасць ужыванняў</b>
1. <i>Спадарства! / Шаноўнае спадарства! / Паважанае спадарства! / Прывітанне, шаноўнае спадарства! / Добры дзень, спадарства! / Вітаю спадарства!</i>	35
2. <i>Сябры! / Шаноўныя сябры!</i>	13
3. <i>Панове!</i>	4
4. <i>Сяброве!</i>	2
5. <i>Шаноўная грамада!</i>	2
6. <i>Пані і панове!</i>	1
7. <i>Паважаныя пані і паненкі!</i>	1
8. <i>Шаноўнае панства і спадарства!</i>	1
9. <i>Браткі беларусы!</i>	1
10. <i>Шаноўныя!</i>	1
11. <i>Паважаныя ўдзельнікі суполкі!</i>	1
12. <i>Таварышчы!</i>	1
13. <i>Прывітанне, народ!</i>	1
14. <i>Šanoŭnaje państwo!</i>	1
15. <i>Vitaju shanounaje spadarstva!</i>	1

Сабраны матэрыял пацвярджае статус этыкемы *Спадарства!* (ізаляванае ўжыванне; ужыванне з маркёрамі “шаноўнае”, “паважанае” або з вітальнай этыкемай) як найбольш запатрабаванай у сучаснай маўленчай практыцы беларусаў. Замацаваная ў 90-я гады ХХ ст., гэтая этыкетная формула, на нашу думку, у найбольшай ступені выяўляе нацыянальную спецыфіку беларускага моўнага этыкету разам з этыкемамі *Шаноўная грамада!* і *Браткі беларусы!*.

Як бачым, дастаткова шырокае “хаджэнне” сярод карыстальнікаў маюць і відавочныя польскія этыкемы: *Панове!*; *Пані і панове!*; *Паважаныя пані і паненкі!*; *Šanojnaje państwo!*, якія хоць і былі традыцыйнымі і адметнымі ў ліставанні беларускай інтэлігенцыі пачатку ХХ ст., сёння павінны кваліфікавацца як іншасістэмныя, а таму непрымальныя. У сувязі з адзначаным вышэй цалкам недарэчнай нам падаецца кантамінаваная формула *Шаноўнае панства і спадарства!*, дзе сутыкаюцца ўласна польскі і ўласна беларускі пачаткі. Паводле польскай мадэлі *Панове!* утворана і формула *Сяброве!*.

Відавочны рускі ўплыў мае месца ў выпадку *Шаноўныя!*. Этыкема *Таварышчы!*, ужытая карыстальнікам *Smatrusenka*, гэта хутчэй не даніна павагі гадам мінулым, а сцёб з рэалій сучасных, перадусім – з актуальнай моўнай сітуацыі і трасянкі.

Прэрагатыву звяртацца да ўсяго беларускага народа ў нас маюць адзінкі, пачынаючы з К. Каліноўскага і класікаў новай беларускай літаратуры, а таму формула *Прывітанне, народ!* у нашым матэрыяле прадстаўлена толькі адзінкавым выпадкам.

2.2.2. Арыгінальныя ветлівыя звароты ў маўленні карыстальнікаў рэсурса *by\_tova*:

Табліца 3

Ветлівы зварот	Колькасць ужыванняў
1. <i>Шанавенства!</i>	1
2. <i>Браты і сёстры грамадзі-пісьменцы!</i>	1
3. <i>Спадары фрэндзы!</i>	1
4. <i>Шаноўныя бульбазаўры!</i>	1
5. <i>Спадарасы!</i>	1
6. <i>Прывітанне, мае любыя шкаляры!</i>	1
7. <i>Алё! Лінгвісты і перакладчыкі!</i>	1

Як бачым, народная словатворчасць, якая з’яўляецца найбольш тыповай праявай інтэрнэт-маўлення, закранула і такі пласт этыкем, як ветлівыя звароты. Індывідуальна-аўтарскія ўтварэнні-неалагізмы, як правіла, вызначаюцца сваёй аказіянальнасцю, непаўторнасцю, непаўтаральнасцю і вобразнасцю, а таму ілюструюць адмысловы спосаб бачання свету моўцамі. Так, напрыклад, этыкема *Спадарасы!* (націск на перадапошні склад) прыдуманая, па словах “вынаходніка”, як “новы зварот

да свядомага ліцвінскага грамадства”. Некаторыя карыстальнікі форума ўспрынялі прыведзены неалагізм як абразлівы. А іншыя, наадварот, падтрымалі “моўную гульню”: *спадарайтэ* (незам.) і *спадаренэ* (зам.). Мы ж бачым тут нішто іншае, як улік асаблівасцей літоўскай мовы: гл. таксама па аналогіі *мадэратарас* замест *мадэратар* у гэтым жа аўтарскім допісе.

3. Высновы. Толькі ў пачатку 90-х гадоў ХХ ст. у моўным этыкеце беларусаў зноў адзначаецца зварот да ўласна беларускіх формул: другая этыкетная нацыяналізацыя звязана з заменай “чужых” (пераважна рускіх і польскіх) мадэляў на “свае” (маюцца на ўвазе мадэлі з вакатывамі *спадар / спадарыня / спадарства*, а таксама “вяртанне” мадэляў з вакатывамі *дзядзька / цётка*).

Сістэма ветлівых зваротаў беларусаў у сеціве не разбураецца: тыя этыкемы, што пачынаюць складвацца ў беларускай літаратурнай мове ў пачатку ХХ ст., актыўна выкарыстоўваюцца і зараз, але з большай дамінацыяй уласна беларускіх этыкетных формул.

Даследаванне дазволіла выявіць шэраг тэндэнцый. Па-першае, гэта вяртанне тых этыкетных формул, што выкарыстоўваліся ў ліставанні беларускай інтэлігенцыі пачатку ХХ ст. Па-другое, актывізацыя народнай словатворчасці ў беларускім сеціве, што з’яўляецца праявай другой этыкетнай нацыяналізацыі.

А. В. Валодзіна

### **ВЫКАРЫСТАННЕ КАМП’ЮТАРНАГА РЭСУРСУ “PATROLOGIA LATINA” ДЛЯ ДАСЛЕДАВАННЯ КАНЦЭПТУ “PAUPERES CATHOLICI”**

У ХХІ стагоддзі ўсё часцей абмяркоўваюцца наступствы і вынікі працэсу глабалізацыі. У большасці выпадкаў акцэнтуюцца менавіта палітычныя, эканамічныя ці культурныя яе аспекты, глабалізацыя ж у галіне навукі звычайна застаецца па-за ўвагай сродкаў масавай інфармацыі. Між тым дадзены працэс у гэтай сферы даволі значны, можна канстатаваць яго разгортванне ў двух асноўных кірунках: геаграфічным і дысцыплінарным. Пад геаграфічным кірункам разумеецца ўсё большая інтэрнацыяналізацыя навукі, калі навукоўцы з розных куткоў зямлі могуць свабодна карыстацца напрацоўкамі адзін аднаго, абменьвацца вопытам, не здзяйсняючы для гэтага вялікіх перамяшчэнняў у прасторы. Пад дысцыплінарным жа я маю на ўвазе ўсё большае ўзаемапранікненне розных галінаў навукі. Яскравым прыкладам гэтых тэндэнцый можа паслужыць даследаванне тых аспектаў заходнееўрапейскай гісторыі, якія абапіраюцца на выкарыстанне кампутарных тэхналогій (Patrologia Latina Database) і лінгвістычных метадаў (канцэптэуальны аналіз).

Patrologia Latina – зборнік вялікай колькасці твораў Айцоў Царквы і іншых царкоўных аўтараў на лацінскай мове, які быў выдадзены

французскім абатам Жанам-Полям Мінем у 1844–1855 гг. Збор уключае ў сябе 221 том з паказальнікамі і ахоплівае вельмі шырокае кола крыніц, напісаных на працягу больш за 1000 гадоў: пачынаючы ад Цертуліяна (прыкладна 200 г. н. э.) і да папы Інакенція III (памёр у 1216 г.). Мінь хацеў уключыць у збор усе тэксты, створаныя да часоў Рэфармацыі, але гэтая задача не магла быць ажыццёўлена з прычыны літаральна выбуху колькасці пісьмовых твораў, які адбыўся ў першай чвэрці XIII стагоддзя. Аднак Мінь зрабіў устаўкі больш позніх тэкстаў, якія традыцыйна далучаюцца да ранніх, у выглядзе прадмоваў або каментароў да іх [3].

Patrologia Latina Database уяўляе сабой поўную электронную версію першага выдання Жана-Поля Міня, у тым ліку ўсе ўступныя артыкулы, арыгінальныя тэксты, крытычныя апарат і індэксы. Нумары слупкоў, пададзеныя Мінем, таксама захаваны, што прадстаўляе вялікую каштоўнасць для спасылак навукоўцаў. Інтэрфейс праграмы выкананы на італьянскай мове. Распрацавана шматвектарная сістэма пошуку:

- па ключавому слову;
- па грэцкаму слову (для абедзвух опцый можна ўключыць пошук па даведачным матэрыяле);
- па аўтару (сярэднявечнаму, сучаснаму, усім аўтарам);
- па назве;
- па тому (тамы разбіты па аўтарах і пераважна вытрыманы ў храналагічным парадку).

Магчымы хуткі прагляд вынікаў. Таксама выданне дазваляе карыстацца зместам або пераходзіць адразу да нумару калонкі ў томе.

Такім чынам, для даследавання канцэпту “*pauperes catholici*” (літаральны пераклад – бедныя каталікі) першапачаткова трэба скарыстацца сістэмай пошуку, ужываючы гэта словазлучэнне ва ўсіх родавых формах. Атрыманы наступны вынік: знойдзена 8 выпадкаў выкарыстання дадзенага словазлучэння ў 6 дакументах. Але ўсе гэтыя дакументы з’яўляюцца аднатыпнымі – гэты лісты/булы Папы Рымскага Інакенція III, том № 216. У трох з гэтых выпадкаў словазлучэнне стаіць у Nom. і з’яўляецца дзейнікам, у двух іншых – прамым дапаўненнем (у Dat. і Acc.), у апошнім – ускосным дапаўненнем у Abl., прычым назіраецца адваротны парадак словаў. Таксама ў першым дакуменце маюцца два ідэнтычныя выпадкі ўжывання ў Gen.

Найперш неабходна высветліць этымалогію дадзеных словаў – *pauperes* і *catholici*, а таксама іх семантычнае поле. Прыметнік *pauperes* узыходзіць да індаеўр. \**pauro-pars* ‘мала здабываецца, мала атрымліваецца’. Асноўныя значэнні: ‘бедны, бедненькі, сціплы, скудны’, пазн. таксама ‘хворы’. Блізкія: *pauperia*, *pauperium* ‘беднасць, убоства’; *pauperies* ‘шкода, страта’; *pauperticulus* (з Вульгаты) ‘убогі, бедненькі’; *paupero* ‘рабіць бедным’; *pauperesco* ‘быць бедным’ [5, с.267–268]. Слова *catholici* прыйшло ў лацінскую мову з грэцкай ад *καθολικός* – прыметніка I–II скланенняў “усеагульны”. У грэцкай мове роднаснымі

з'яўляецца *καθολικίη* – “увогуле, у агульным выглядзе” [1]. У лацінскай мове: *catholicus* як прыметнік ‘усеагульны, усеахопны, каталіцкі, прававерны’ і вытворныя ад яго *catholica* ‘агульныя і пастаянныя прызнакі, сукупнасць, сусвет’ і *catholice* – у агульным выглядзе ‘каталіцкі, прававерна’ [2, с.127]. Але ў дадзеным выпадку нас больш цікавіць прымяненне гэтага тэрміна ў адносінах да царквы, што адбылося каля 110 г. у лісце св. Ігнація да жыхароў Смірны. Замацаваны дадзены тэрмін у Нікейскім Сімвале веры.

Прагматычнае поле значэнняў канцэпту:

- так называюць Дурандэ дэ Оска і яго паслядоўнікаў (у двух выпадках);
- пішуць лісты да Папы Рымскага;
- просяць заступніцтва ў Святога Прастола;
- трымаюцца артадаксальнай веры і набожных паводзінаў;
- ізноў далучаны да каталіцкага адзінства;
- навучаюць і кантралююць вернікаў.

З гэтага можна патлумачыць канцэпт “*pauperes catholici*” наступным чынам: убогія прававерныя, якія пасля перапынку ізноў уваходзяць у каталіцкую царкву, на чале з Дурандам дэ Оска і яго паслядоўнікамі звяртаюцца асабіста і праз лісты да Папы Рымскага з просьбай аб апякунстве, а таксама навучаюць і кантралююць іншых вернікаў.

Таксама *Patrologia Latina Database* ўтрымлівае і даты напісання булаў: 1208, 1209 і 1212. З дадзеных гэтага рэсурсу можна вывесці прыкладны арэал распаўсюджання “бедных каталікоў”: Паўночна-Заходняя Італія (Мілан, Генуя), Паўднёва-Заходняя Францыя (Марсэль, Нарбона), Паўночна-Усходняя Іспанія (гарады Уэска, Утыка, Барселона, Таракона, Арагонскае каралеўства ў цэлым). Гэта вынікае з адрасатаў папы: царкоўных іерархаў, гарадскіх уладаў і нават караля. Такі шырокі спіс атрымліваецца за кошт таго, што ў тэксце прыведзены спасылкі на падобныя тэксты, што дазваляе адсачыць амаль аналагічныя лісты іншым адрасатам. Пэўныя звесткі можна атрымаць і з імёнаў людзей, названых “беднымі каталікамі”: Дуранд з Оскі (*Durandus de Osca*) – Оскес у Рурку (*Osques, Rouergue*) (Францыя) і Уэска ў Арагоне (*Huesca, Aragon*) (Іспанія), а таксама яго паплечнікаў: Дуранда з Найака (*Durandus de Naiaco*), Гуйлема з Сант-Антонія (*Guillelmus de Sancto Antonio*), Іаана з Нарбоны (*Joannus'a de Narbona*), Эрменгаўда (*Ermengaudus*), Бернарда Бітэрэнсіка (*Bernardus Bitterrenses*), Раймунда з Сан-Паўла (*Raymundus de Sancto Paulo*) і Эрбіна (*Ebrinus*) з секты вальдэнсаў.

Пра апошніх варта сказаць больш падрабязна. Вальдэнсы – ерэтычны рух, які узнік як частка свецкага рэлігійнага руху ва ўлонні каталіцызму ў апошняй чвэрці XII стагоддзя. Заснавальнік – Пётр Вальдо – былы багаты гандляр з французскага горада Ліёна, які распачаў прапаведванне Слова Божага і жыцця па прыкладу апосталаў. З шэрагу розных прычынаў ён і

яго папличнікі былі ў рэшце рэшт адлучаны ад царквы, хоць некаторымі і рабіліся спробы вярнуцца назад. Менавіта такой, як вынікае з тэксту лістоў, і была група “бедных католікаў”.

Такім чынам, на падставе атрыманай у выніку канцэптуальнага аналізу дэфініцыі і супастаўлення яе з іншымі дадзенымі можна зрабіць выснову, што *pauperes catholici* – група вальдэнсаў, ачоленых Дурандам з Оскі і яго вышэйпералічанымі папличнікамі, якая пасля адлучэння ад каталіцкай царквы ўсёй секты пажадала вярнуцца ў яе лона, просячы ў Папы Рымскага Інакенція III у 1208, 1209 і 1212 асабіста і праз лісты заступніцтва, матывуючы сваю просьбу прававернасцю і набожным жыццём, а таксама тым, што яны займаюцца навучаннем і кантролем над іншымі вернікамі.

#### СПІС ЛІТАРАТУРЫ

- 1 Дворецкий, И. Х. Древнегреческо-русский словарь: в 2 т. / под ред. С. И. Соболевского. – М. : Изд-во иностранных и национальных словарей, 1958. – 2 т.
- 2 Дворецкий, И. Х. Латинско-русский словарь. – М., 2008.
- 3 Patrologia Latina Database. – 1 электр. апт. дыск (DVD-ROM).
- 4 Patrologia Latina. The Full Text Database [Electronic resource]. – Mode of access: <http://pld.chadwyck.co.uk/> – Date of access: 12.09.2010.
- 5 Walde, A. Lateinisches Etymologisches Wörterbuch/ A. Walde, J. B. Hofmann. – Heidelberg, 1938.

Н. А. Говорушкина

### **ИСПОЛЬЗОВАНИЕ ГРАФИЧЕСКОЙ ПРОГРАММЫ COREL DRAW ПРИ ВЕРСТКЕ ТЕКСТОВЫХ БЛОКОВ**

Бурное развитие компьютерных программ для работы с текстовыми блоками требует подготовки специалистов-филологов, владеющих навыками работы в графических программах и умеющих грамотно преподнести текст с точки зрения дизайна и графики. Текст является едва ли ни единственным средством передачи информации во Всемирной сети, широко используется в рекламе и полиграфии при изготовлении буклетов, каталогов, журналов, газет.

Современный рынок труда предоставляет много возможностей трудоустройства, где часто можно встретить вакансии совместителей: веб-редактор, копирайтер, переводчик порталов, верстальщик. Считается, что наиболее успешными для данных вакансий являются журналисты и филологи. Без всяких сомнений, любой лингвист должен уметь использовать возможности какой-нибудь графической программы по верстке и редактированию текстовых блоков. Однако, помимо технических навыков, необходимо также иметь представление о некоторых особенностях и основах самого дизайна: подбор шрифтов и начертаний, сочетаемость цветов, раз-

мер шрифтов и композиция.

Большие возможности для работы с текстом предоставляет графическая программа Corel Draw. Главными особенностями данной программы являются:

- автоматическая проверка орфографии, которая выполняется далеко не всеми графическими программами;
- различные способы выравнивания текстовых блоков, включая обложки различной формы и размера;
- создание цепочек связанных текстовых блоков;
- гибкая система динамического редактирования текстовых блоков относительно занимаемого пространства.

Для создания текста в программе Corel Draw используется инструмент Type Tool (Текст).

Для верстки документов с большими массивами текста (например, брошюр) следует использовать "текст в рамке", что расширяет круг возможных операций по оформлению документов. Преимущества такого текста следующие:

- автоматически при достижении правой границы текстового блока выполняется перенос слова на новую строку;
- при изменении размеров текстового блока текст автоматически переверстывается;
- имеется возможность размещения текста в блоке в несколько колонок;
- в текстовый блок возможно импортирование текстовых файлов из текстовых редакторов.

В блоке имеется управление различными специальными параметрами верстки текста – расстояние между абзацами, отступы, красная строка и т. д. На текст в рамке распространяются все возможности форматирования символов и абзацев.

*Цепочками связанных рамок* простого текста пользуются для того, чтобы разместить один текст в разных местах одной страницы или на разных страницах документа. Речь идет не о размещении копий, а о продолжении текста при переходе из одной рамки в другую. Чтобы добавить в цепочку новую рамку, остается только перетащить указатель по ее диагонали.

Рамки, входящие в состав одной цепочки, соединяются цветными стрелками, которые отображаются только на экране, а на печать не выводятся. По форме верхних и нижних индикаторов можно судить о местоположении рамки в цепочке и наличии в ней неразмещенного текста. У первой рамки цепочки верхний индикатор пустой. Если рамка в цепочке не последняя, ее нижний индикатор выглядит как миниатюрная страница с текстом; если рамка не первая в цепочке, ее верхний индикатор выглядит как миниатюрная страница с текстом. Наконец, если рамка последняя, ее нижний индикатор пуст или содержит в себе стрелку, направленную вниз (если в рамке имеется неразмещенный текст).



Если необходимо, чтобы следующая рамка в цепочке имела форму, отличную от прямоугольника, следует вначале построить замкнутый криволинейный объект желаемой формы, а затем, выбрав инструмент текст, щелкнуть на нижнем индикаторе последней рамки цепочки. После этого при перемещении указателя внутрь замкнутого криволинейного объекта он примет форму широкой стрелки, направленной вправо. После щелчка на объекте таким указателем создается рамка, совпадающая с объектом по форме, – текст заключается внутрь объекта.

В результате описанной выше процедуры создается составной объект, состоящий из замкнутой фигуры (управляющий объект) и прямоугольной рамки, заключенной в оболочку, по форме совпадающую с управляющим объектом. После построения рамки в оболочке управляющий объект больше не нужен, поэтому составной объект можно разъединить командой Arrange > Break Apart (Монтаж > Разъединить) и удалить замкнутую кривую. При этом рамка в оболочке остается в составе цепочки рамок простого текста.

Чтобы удалить рамку, входящую в состав цепочки, достаточно выделить ее инструментом Pick (Указатель) и нажать клавишу Delete. Если удаляется последняя рамка цепочки, содержащийся в ней текст будет вытеснен в предпоследнюю рамку, сохраняясь в ней как неразмещенный. При удалении первой рамки цепочки текст вытесняется во вторую и размещается в ней, из второй – в третью и т. д. Избыток текста в цепочке всегда находится в последней рамке. При удалении промежуточной рамки автоматически создается новая связь, соединяющая предыдущую и следующую рамки.

Рамки, входящие в состав цепочки, могут размещаться на различных страницах документа Corel Draw. Так можно размещать весьма значительные объемы простого текста в связанной цепочке рамок, расположенных в пределах одной или нескольких страниц документа.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Буленица, Т. Дизайн на компьютере: Самоучитель / Т. Буленица. – СПб. : Питер, 2003. – 320 с.
- 2 Бэйн, С. Эффективная работа с CorelDRAW11 / С. Бэйн. – СПб.: Питер, 2003. – 768 с.
- 3 Петров, М. И. CorelDRAW11: Самоучитель / М. И. Петров. – СПб. : Питер, 2002. – 624 с.
- 4 Хант, Ш. CorelDRAW9 для профессионалов / Ш. Хант. – СПб. : Питер, 2000. – 512 с.

Н. В. Ерашко

### **ЛЕКСИКО-СЕМАНТИЧЕСКОЕ ПОЛЕ «СОВРЕМЕННАЯ ЖЕНЩИНА» (по данным интернет-источников)**

В настоящее время большое внимание уделяется развитию и внедрению компьютерных технологий во все сферы жизни общества, что позволяет очень быстро и эффективно овладевать разнообразной информацией для удовлетворения различных потребностей.

Внедрение в практику курса «Компьютерная лингвистика» является очень важным и перспективным для современной жизни. Компьютерная лингвистика, изучающая взаимодействие естественного и искусственного языков, способствует активному развитию высоких технологий, в том числе – интеллектуальных, позволяет лучше понять систему и структуру самого языка, сориентированного на алгоритмы формального анализа и синтеза.

Мы пока не знаем, какие возможности может дать человечеству надвигающаяся новая компьютерная революция. Однако можно надеяться, что компьютерная лингвистика перейдет на совершенно новую технологическую базу, основа которой закладывается в наше время. Одно из направлений практического использования названной дисциплины – освоение Всемирной паутины, иначе говоря, Интернета. Под этим понимается правильная систематизация данных и эффективное использование интернет-источников. Всё это помогает активно исследовать различные лексико-семантические поля в современном русском языке, в том числе – ЛСП «современная женщина».

Актуальность исследования определяется, с одной стороны, возросшим в современной лингвистике вниманием к когнитивной и этнокультурной составляющим картины мира, с другой стороны – бурным развитием гендерологии и ее ответвления – гендерной лингвистики. Исследование проведено на стыке нескольких лингвистических дисциплин: когнитивной и гендерной лингвистики, лингвистической семантики, лексикологии, психолингвистики и лингвокультурологии. Именно интегрированный подход обеспечивает наиболее адекватный ракурс исследования такого сложного по своей природе объекта, как концептуальное пространство «современная женщина».

В ходе исследования было выявлено, что для обозначения понятия «современное лицо женского пола» используются многочисленные лексические, фразеологические, афористические и текстемные наименования. Объемность номинативного поля «женщина», вербализация богатой и разнообразной ономасиологической парадигмой, широкая лексическая объективация свидетельствуют о его коммуникативной релевантности для русского национального сознания.

Прежде всего, в зеркале фразеологии отражаются традиционные черты семейных отношений, характерные для обыденного русского менталитета. Веками установленные религиозные каноны поведения женщины в семье определяют ее как *честную, порядочную женщину*. В мужскую картину мира не укладывается тот факт, что женщина может меняться, в ее характере со временем проявляются какие-то ее негативные черты: *несамостоятельная* ‘зависящая от кого-либо’, *меркантильная* ‘излишне рас-

четливая', *курящая* (качество, изначально несвойственное женщинам) 'с сигаретой в зубах', *матерящаяся как сапожник* 'употребляющая нецензурную лексику', *развращенная* 'демонстрирующая развратный образ мыслей, поведение', *пошлая* 'низкая в нравственном отношении; безвкусно-грубая', *безманерная* 'лишённая пристойного поведения', *вульгарная* 'пошлая и грубая; непристойная', *циничная* 'полные цинизма, бесстыдные', *эгоистка* 'она активно требует от мужчин денег, хочет, чтобы ее содержали уже на ранних стадиях знакомства, и приходит в негодование, получая отказ'.

Большую часть номинаций, вербализующих концепт «современная женщина», представляют собой слова, соотносящие женщину с животным миром. Сексуально привлекательная женщина соотносится с *кошкой, тигрицей, львицей, пантерой*. Наряду с физиологической привлекательностью женщины обращается внимание и на уровень развития ее интеллекта, хотя в целом женский ум всерьез не принимается. Об этом свидетельствует такой современный стереотип женщины, как *блондинка*. Так называют женщину красивую, сексуально привлекательную, но глупую, с низким уровнем развития интеллекта. Согласно обыденным представлениям, такая женщина много времени тратит на себя, скрупулезно следит за своей внешностью, не занимается своим образованием и существует как игрушка для мужчин.

Интегральная сема «отношение к одежде» наиболее ярко проявляется в семантике существительных *шмоточница, тряпичница, куколка, фифа*, в которых актуализируется особое повышенное внимание к самым разным новинкам модной одежды и приобретению её в большом количестве для своего гардероба. Интегральный признак «привлечение внимания» в русском языке реализуется преимущественно в дифференциальных признаках «способ привлечения внимания» (манеры, одежда, поведение) и «внутренние качества субъекта действия» (легкомыслие, непостоянство): *кокетка* 'женщина, которая использует лёгкий флирт в поведении', *вертихвостка* 'женщина, которая использует мужчин ради своей выгоды'.

Несомненно, неоспоримым идеалом для русского мужчины является женщина не просто привлекательная, а *женщина-загадка*, которая является одновременно и мужественной, и опасной. Слово *стерва* 'подлая, негодная женщина' употребляется для обозначения женщины, физически недоступной, но вместе с тем сексуально привлекательной. Ее трудно завоевать, но победа от этого становится еще слаще.

При характеристике женской одежды релевантными для положительной оценки выступают следующие признаки: *красота, чистота, аккуратность, новизна, индивидуальность*. Оценивается также соответствие одежды моде и социальной обстановке: *стильная, гламурная* (поведение + стиль одежды) 'главная забота – самые последние «писки» модных вещей, симпатичные парни и вечеринки', *эффектная* 'производящая эффект, впечатляющая', *шикарная* 'роскошная', *зефирная* 'лёгкая, воздушная в пере-

носном смысле'. Особенно активно (применительно к понятию «современная женщина») употребляется определение *гламурная*, т. е. склонная к роскошному, «шикарному» стилю жизни, культивируемому «глянцевыми» журналами (напр., образ гламурной «светской львицы» Ксении Собчак).

В русском сознании красота определяется преимущественно предписанным набором атрибутов, которые составляют один из существующих типов красоты. Таким образом, одним из главных критериев оценки женщины в языке является внешняя привлекательность. Для русских красота – это то, что предписано другими (некий национальный эталон).

И. И. Зубова

### **ФОРМАЛЬНОЕ ОПРЕДЕЛЕНИЕ СРЕДСТВ ВЫРАЖЕНИЯ НЕИСКРЕННОСТИ В АНГЛОЯЗЫЧНЫХ ВЫСКАЗЫВАНИЯХ**

Термин «неискренность» был введен в лингвистический обиход Дж. Остином во второй половине 20-го века. Ученый понимал под ним такие условия осуществления речевого акта, которые делают его «пустым». Согласно теории Дж. Остина «неискренние действия есть злоупотребления, при выполнении которых отсутствуют определенные мысли, чувства, намерения» [1, с. 17]. Последующее развитие теории речевых актов осуществлялось по линии исследования высказываний, произведенных в условиях искренности. Это – один из законов, которому говорящие должны следовать в процессе общения. Поэтому в отличие от понятия «искренность», понятие «неискренность» не получило детальной разработки в теории речевых актов.

Для обозначения осознанного выражения ложных смыслов помимо термина «неискренность» широко используются и другие термины, например, ложь, обман, ложная информация, манипулирование истиной, тенденциозное представление события, неправда, дезинформация и т. д. Такое терминологическое разнообразие свидетельствует о том, что интерес ученых привлекают различные приемы манипулирования истиной и управления пониманием. С. Н. Плотникова доказывает, что все приведенные выше термины соответствуют единому концепту, который можно сформулировать как «намеренное, осознанное выражение ложных пропозиций в процессе речевого общения» [2, с. 13], и считает, что термин «неискренность», с точки зрения отражения указанного концепта, является наиболее адекватным. Во-первых, он может заменить собой все другие термины, что позволяет включить в рассмотрение весь комплекс языковых средств воплощения концепта. Во-вторых, этот термин позволяет избежать ненужного параллелизма с такими понятиями логико-философского анализа, как ложность и ложь. В-третьих, он является наиболее нейтральным и

однозначным и не несет в себе сниженных или разговорно-бытовых коннотаций. Автор считает, что концепт «неискренность» необходимым образом включает в себя ложность, так как быть неискренним – значит, говорить ложь.

Концепт «ложь» присущ всем людям, заложен в сознании всех народов. Он является универсальной человеческой категорией. Поэтому появление такой компьютерной системы, которая могла бы объективно выявлять проявление обмана и лжи в письменной речи и решать некоторые проблемы в области судебной лингвистической экспертизы, объективного определения типа личности, особенно в тех профессиональных сферах, где этот фактор очень важен, весьма актуально.

Под формальной моделью какого-либо лингвистического явления понимается некоторая система правил, имитирующая структуру и/или поведение некоторых лингвистических объектов и позволяющая хотя бы частично воспроизвести эти объекты либо с помощью человека, либо с помощью компьютера. Для воспроизведения с помощью компьютера какого-либо лингвистического явления необходимо составить базу формализованных данных, описывающих это лингвистическое явление, и построить с опорой на нее алгоритм функционирования данного явления.

В процессе создания лингвистической базы данных системы определения средств выражения неискренности в англоязычных высказываниях были использованы ряд словарей английского языка [3; 4; 5; 6], а также фрагменты из 22 рассказов таких современных английских и американских авторов как A. Beard, S. Cronk, J. Haynes, T. Horton, T. Lewis и др. На первом этапе с опорой на словари английского языка из текстовых фрагментов были выявлены наиболее типичные лексические средства (слова и словосочетания) выражения неискренности. На втором этапе все лексические средства были разделены на группы с различной степенью выражения неискренности: 1) ложь; 2) обман / фальшь / неправда; 3) мошенничество / жульничество; 4) подделка; 5) хитрость / уловка / ловушка; 6) лицемерие / притворство / симуляция; 7) предательство / интриги; 8) уклонение от ответа / умалчивание; 9) преувеличение; 10) искажение / фантазия / введение в заблуждение. Десять списков лексических единиц в полном объеме вошли в базу данных.

В ходе исследования было определено, что формы глагола *lie* в значении 'лежать' и в значении 'лгать' в настоящем времени, а также их формы с окончанием *-ing* являются омонимами. Омонимичны формы глаголов в 3-ем лице единственном числе (*lie – lies*) и имени существительного во множественном числе (*lie – lies*). Для автоматического разрешения омонимии был составлен список наиболее употребительных словосочетаний с глаголом *lie* в значении 'лежать', например: *lie still, lie awake, lie hidden, lie in state, lie in wait for, lie heavy on* и т. п. Диагностирующими признаками для глагола *lie* в значении, отличном от значения 'лгать' также являются предлоги и послелоги: *ahead, along, alongside, at, back, before, behind, beside, be-*

*tween, beyond, by, doggo, down, down on, down under, fallow, low, off, on, open, out, over, up, with, within.* Все перечисленные выше словосочетания с ядерным глаголом *lie* и его временные формы также представлены в лингвистической базе данных.

В виде отдельного списка были выделены слова и словосочетания, выражающие искренность, поскольку при наличии перед такими лексическими единицами компонента с отрицательной коннотацией они трансформируются в средства выражения неискренности. В ходе выполнения исследования были выявлены случаи типа *I won't let you lie in my house / He was lying, I saw that.* Использование глагола *lie* в этих высказываниях неоднозначно и требует включения в процедуру анализа не микро-, а макроконтекста.

С опорой на лингвистическую базу данных были построены две формальные модели, позволяющие определить средства выражения неискренности в англоязычных высказываниях. Первая формальная модель имитирует диалог пользователя с компьютером, в ходе которого пользователь (П) использует средства выражения неискренности, а компьютер (К) должен это распознать и вербально правильно отреагировать. Например:

**П:** *I cheated at the exam.*

**К:** *It's not very good I think.*

**П:** *Liars have need of good memories.*

**К:** *Anyway, one should be punished for telling the untruth.*

**П:** *When I asked him where he had been he didn't give me the direct answer.*

**К:** *Probably, there are reasons for that.*

В данном случае, в зависимости от лексического значения слов и словосочетаний, описывающих неискренность, сообщения подразделяются на слова-стимулы и слова-реакции. В зависимости от группы, к которой относится та или иная лексическая единица, выражающая неискренность, выбирается определенная реплика-реакция. Например, если в реплике-стимуле есть слово, относящееся к группе слов, выражающих 'предательство / интриги', то и реплика-реакция будет соответствующей: *It's pure treachery!* Что касается эмотивной стороны анализируемого объекта, то положительные эмоции реплики-реакции будут вызывать только тогда, когда в репликах-стимулах неискренность будет отрицаться. Например:

**П:** *I couldn't lie to him .*

**К:** *Oh, that's good.*

Исходя из вышесказанного, в лингвистическую базу данных были включены возможные реплики-реакции компьютера на реплики-стимулы, выражающие различные виды проявления лжи. Следует отметить, что практически все реплики-реакции компьютера равноценны, и компьютер выбирает одну из них случайным образом, как в вышеприведенных примерах. Реплики-реакции на неискренность могут иметь референцию практически к любому речевому акту лжи. В то же время такие выражения как *That's a good thing. = That's nice! = Great! = Fine! = Good for you!* являются репликами-реакциями на выражение искренности в высказывании пользователя.

В первой модели представлен и следующий случай. Реплика-стимул пользователя может быть вопросительной. Это может быть обращение к компьютеру за советом, например, *What should I do If she lie to me?* Или попытка узнать его мнение, например, *Do you think it was just exaggeration?* Возможные реплики-реакции компьютера могут быть, например, следующими: *That's a very difficult question; Lying is bad anyway; It's up to you to decide* и т. д. Если в реплике-стимуле нет средств выражения неискренности, то для поддержания диалога на экран дисплея выводятся нейтральные реплики-реакции типа *I see, I know, I understand* и т. п.

Вторая формальная модель предназначена для определения средств выражения неискренности в англоязычных художественных текстах. Ее основной особенностью является то, что она позволяет не только обнаружить подобные лексические средства, но и определить форму проявления неискренности (обман / фальшь / неправда, подделка, мошенничество / жульничество и т. д.). Поскольку в текстовом фрагменте могут использоваться несколько разных форм проявления лжи, в модель системы заложена возможность подсчета числа слов и словосочетаний, выражающих конкретную форму проявления неискренности. В ней также учитываются лексические единицы, выражающие искренность, которые при отрицании будут причислены к неискренним элементам. Результаты работы формальной модели могут быть представлены в следующем виде: на экран дисплея (принтер) выводятся проанализированный компьютером текст, название лексической группы, к которой относятся слова и словосочетания, выражающие неискренность, список этих лексических единиц и частота употребления каждой лексической единицы.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Остин, Дж. Как производить действия при помощи слов / Дж. Остин // Избранное. – М. : Идея-Пресс, 1999. – С. 17–25.
- 2 Плотникова, С. Н. Лингвистические аспекты выражения неискренности в английском языке: автореф. дис. ...канд. филол. наук: 10.02.04 / С. Н. Плотникова; Рос. акад. наук. – М., 2000. – 42 с.
- 3 Oxford Phrasal Verbs. – Oxford, 1990. – 486 p.
- 4 English Thesaurus. The New Choice. – England, 1999. – 384 p.
- 5 Longman Dictionary of Contemporary English. – UK, 2005. – 1949 p.
- 6 Литвинов, П. П. Англо-русский фразеологический словарь с тематической классификацией / П. П. Литвинов. – М. : Яхонт, 2000. – 448 с.

Л. Ф. Крупа

**УСТОЙЧИВАЯ НОМИНАТИВНАЯ ЕДИНИЦА  
МОБИЛЬНЫЙ ТЕЛЕФОН И ЕЁ СИНОНИМИЧЕСКИЙ РЯД  
В СОВРЕМЕННОЙ РУССКОЙ РЕЧИ  
(по данным интернет-источников)**

Появление новых технологий непременно сопровождается возникновением большого пласта новой и соответственно неизученной лексики, включающей в себя не только терминологические единицы, но и жаргонную лексику. Особенностью неcodифицированного языка является его подвижность, проникновение в разговорную речь людей. Т. В. Китаева определяет жаргон как «языковые средства определенной профессии или социальной группы, не ограниченные намеренно рамками этой группы» [2]. «Его функциональное назначение – быть средством непринужденного, неофициального общения группы людей» [3]. Жаргонизированная речь слышна практически повсюду: на улице в беседе с друзьями и взрослыми, в школах, в высших учебных заведениях и т. д. Появление и быстрое распространение мобильной телефонии открыло новые возможности для словотворчества. Свобода выбора языковых средств в процессе самовыражения стала причиной появления многочисленных синонимов устойчивой номинативной единицы *мобильный телефон*.

В статье рассматриваются структурно-семантические особенности 24 синонимов данной единицы, преимущественно жаргонного характера, выявленных в Интернете, на соответствующих форумах, сайтах мобильных операторов и т.д.; см: <http://www.m-unlock.com>; <http://www.mobile-files.ru/forum>; <http://www.imho.ws/forumdisplay>; <http://www.lifemob.net>; <http://www.mobi.ru>; <http://www.forum.sotovik.ru>; <http://www.mobiletechnics.tabaru/forum>; <http://www.bcetyt.ru/computers/notebooks/mobil-nyj-sleng-chast-1.html>, [http://soft.mail.ru/pressrl\\_page.php?id](http://soft.mail.ru/pressrl_page.php?id)

Условно выявленные нами синонимы можно разделить на несколько семантических групп, в основу которых положены следующие признаки:

- говорение (*блаблафон, трепальник, тринделка*) – 12,5 %;
- звуковой сигнал (*звонилка, дрындозвон, жузжик*) – 12,5 %;
- технические характеристики (*трубка, труба, мобила, мобилла, мобилка, мобильник, могильник, мобильный, сотовый, сотик, сотка, соточка, телефоша, тело, тел, телефон*) – 67 %;
- указание на части тела, задействованные в процессе звонка (*хэндик, ухомор*) – 8 %.

Лексические единицы с нейтральной стилистической окраской составляют 12,5 % от общей выборки (*телефон, мобильный, сотовый*). Преобладают в рассматриваемом синонимическом ряду лексемы с ярко выраженной коннотацией (чаще – положительной).

В результате структурно-семантического анализа были выявлены следующие способы образования номинаций.

1. Сужение лексического значения. Лексема *телефон* имеет значения: «м.1. Система связи, созданная для передачи на расстояние речи при помощи электрических сигналов. 2. Аппарат для такой связи» [1]. В настоящее время наблюдается сужение лексического значения: лексема *телефон* всё чаще используется для обозначения непосредственно мобильного телефона.



Наряду с лексической единицей *телефон* употребляются номинации *тело, тел, телефоша*. Образуется парадигма: *тел – тело – телефон – телефоша*. Также имеют место парадигмы лексем *труба, сотовый, мобильный*:

- *труба – трубка*;
- *мобильный – мобила* (структурный вариант *мобилла*) – *мобилка – мобильник*;
- *сотовый – сотик – сотка – соточка*.

2. Конверсия. Номинации *мобильный телефон, сотовый телефон* упрощены, лексемы *мобильный, сотовый* употребляются самостоятельно в качестве отадъективных существительных и служат для обозначения телефона.

3. Фонетическая мимикрия. Данный способ образования фиксируется нечасто. Из 24 изучаемых единиц лишь номинация *Могильник* может быть отнесена к этой группе. Лексема *могильник* обладает пренебрежительной коннотацией.

4. Сложение. Среди подобных новообразований можно выделить лексемы со следующей структурой: первый элемент представлен звукоподражанием и вносит пренебрежительный оттенок; второй представлен именем существительным (*блаблафон, дрындозвон*). Лексическая единица *ухомор* образована путем сложения существительного *ухо* с основой глагола *морить* – «*приводить в истощение, изнувать*» [1]. Номинация *ухомор* приобретает, таким образом, шутливо-ироничную коннотацию.

5. Суффиксация. Окказиональные отглагольные субстантивы образованы суффиксальным способом при помощи суффиксов *-ик-* (*жужжик* от глагола *жужжать*), *-илк-* (*звонилка* от глагола *звонить*), *-елк-* (*тринделка* от жаргонного глагола *триндеть*, имеющего значение «*разговаривать*»), *-льник-* (*трепальник* от жаргонного глагола *трепаться* – (разговорно-сниженное) «*разбалтывать что-либо*») [1].

Отадъективные субстантивы образованы при помощи суффиксов *-к-* (*мобилка, сотка*), *-ик-* (*сотик*), *-ник-* (*мобильник*), *-очк-* (*соточка*). Среди единиц данной группы наблюдается усечение основы: *мобила* и структурный вариант *мобилла*.

Отсубстантивные субстантивы немногочисленны. Они составляют 12,5 % от выборки синонимов лексемы *мобильный телефон*. Наряду с суффиксацией *-к-* (*трубка*), *-ош-* (*телефоша*) используется усечение: *тел*.

6. Заимствование. Среди русскоязычного населения, например, в Германии и Англии, используется лексема *хэндик* (англ. разг. *handy* – *мобильный телефон*). Данное слово образовано посредством заимствования основы англицизма и дальнейшего прибавления русского суффикса *-ик-*: *хэндик*.

Синонимы слова *телефон* различаются по принадлежности к категории рода. Существительные женского рода представлены 10 единицами (*тринделка, звонилка, трубка, труба, мобила, мобилла, мобилка, сотка, соточка, телефоша*), существительные мужского рода представлены

также 10 единицами (*мобильный, сотовый, блаблафон, трепальник, дрын-дозвон, жужик, мобильник, хендик, ухомор, тел*). Что касается синонимов среднего рода, то выявлена лишь одна лексема *тело*. Таким образом, в процессе создания синонимов к слову *телефон* отмечается тенденция выбора существительных мужского и женского рода, нежели среднего рода.

Продуктивными суффиксами для существительных мужского рода являются *-ник-, -ик-*; женского рода *-к-*.

Проанализировав частоту употребления синонимов номинативной единицы *мобильный телефон* на материале интернет-источников, мы выявили, что наиболее используемыми лексическими единицами являются, *труба, мобильник, сотик*. Уникальными являются номинации *блаблафон, трепальник, тринделка, звонилка, дрындозвон, жужик, хэндик, ухомор, могильник*.

Таким образом, чаще всего в качестве синонимов номинативной единицы *мобильный телефон* используются лексемы с положительной коннотацией. Анализ исследуемого материала показал, что наиболее продуктивным способом образования новых слов в рассмотренном синонимическом ряду является суффиксальный – 59 %.

#### СПИСОК ЛИТЕРАТУРЫ

1 Англо-русский словарь общей лексики (The Universal English-Russian Dictionary) // ABBY Lingvo 12 [Электронный ресурс]. – Электрон. текстовые дан. (913 416 192). – М. : ABBY Software House, 2006. – 1 CD-ROM.

2 Китаева, Т. В. Сленг как фактор формирования речевой компетенции студентов / Т. В. Китаева // Высшая школа, 2008. – № 6. – 2008. – С. 75–78.

3 Борисова, Е. Г. О некоторых особенностях современного жаргона молодежи / Е. Г. Борисова // Русский язык в школе. – 1981. – № 3. – С. 83–87.

А. И. Ничипорчик, Е. В. Ничипорчик

### **ИСПОЛЬЗОВАНИЕ МЕТОДА СЕМАНТИЧЕСКОЙ РАЗМЕТКИ В ОРГАНИЗАЦИИ ПОИСКА ПО ТЕКСТОВЫМ ДОКУМЕНТАМ**

В то время как потребность в использовании поисковых систем продолжает расти, степень удовлетворённости пользователей данными системами не становится выше. Это объясняется, с одной стороны, тем, что поиск – это достаточно трудная задача, с другой стороны, тем, что пользователи становятся всё требовательнее и рассматривают поиск как нечто гораздо большее, чем традиционные методы получения информации.

В настоящее время существует 3 ключевых проблемы поиска: неточные результаты, необходимость уточнения запроса и необходимость поддержки сложных задач и решений. Семантические технологии могут помочь в решении этих проблем путём улучшения основных поисковых результатов, а также через введение расширенных методов взаимодействия с пользователями, таких, как многогранная навигация и конкретизация задач [1], поиск

по сущности [2], перефразирование запросов [3], систематизация ответов на вопросы на основе семантических графов и списка фактов [4] и др.

Понимание пользовательских запросов и содержимого ресурсов в существующих поисковых системах до сих пор достигается в большей степени за счёт простых, основанных на синтаксических признаках описаний содержимого ресурсов и искомой информации. Большинство поисковых систем до сих пор использует модель с преобладанием ключевых слов, в которой запрос из ключевых слов сопоставляется с набором слов из документа. Главным преимуществом семантического поиска является то, что семантический поиск принимает во внимание смысл пользовательских запросов и содержимого ресурсов.

В качестве инструмента для проведения семантического анализа может быть использован семантический компонент системы ДИАЛИНГ [5]. Эта система автоматической обработки текста построена помодульно. В её работе используются следующие процессоры: графематический, морфологический, синтаксический и семантический.

В основе семантического анализа системы ДИАЛИНГ лежит информационно-ролевой подход, который подразумевает, что в сферу семантического анализа входит:

1. Построение семантической интерпретации слов и конструкций;
2. Установление «содержательных» семантических отношений между элементами текста.

В центре теории, на базе которой был разработан семантический процессор, находится Русский общесемантический словарь, или РОСС, который включает семантическое описание по следующему шаблону:

1. Семантический класс лексемы (набор семантических характеристик);
2. Грамматический класс лексемы;
3. Валентная структура лексемы (в терминах семантических отношений);
4. Семантические и грамматические ограничения.

К примеру, словарная статья глагола винить в РОСС имеет следующее представление:

ЗГЛ = винить 1  
КАТ= 1 ЭТК.СИТ  
ГХ = 1 ГЛ:ГГ  
СХ = 1 ИНТЕЛ  
                  2 КОММУНИК  
ВАЛ= АГЕНТ , А1 , С  
                  ОБ , А2 , С  
                  СОДЕРЖ , А3  
ГХ1 = 1 подл : И  
СХ1 = 1 ОДУШ  
ГХ2 = 1 п\_доп : В

СХ2 = 1 ОДУШ

2 АБСТР

ГХ3 = 1 к\_доп : в+П

2 к\_доп : за+В

ИЛЛ = Я сам себя винил во всех неудачах.

Винить за это некого.

Не отрицаем, что осуществление семантического анализа также является достаточно сложной задачей. Главная трудность в отсутствии формальных показателей, на основании которых можно выстраивать семантические типологии. В семантическом анализаторе системы ДИАЛИНГ семантические классы выделяются с опорой на семантические отношения и семантические характеристики. Например:

Семантическое отношение

$R(A, B) \Leftrightarrow$  «А является R для В»

Таблица 1 – Перечень отношений

Название (R)	Примеры
Автор	Роман Толстого
Агент	Мы сократили отставание
Адресат	Я отдал стул отцу
Время	Это произошло вчера

Семантическая характеристика

$SX(A) = SX(B) \Leftrightarrow$  "А имеет смысловое сходство с В"

Пример:

$SX(\text{повар}) = SX(\text{генерал}) = \text{ДОЛЖ}$

Таблица 2 – Перечень семантических характеристик

Название	Примеры
Вещество	бензин, ядохимикат
Восприятие	слушать, видеть

Результатом работы семантического процессора является семантический граф, в узлах которого стоят леммы или связка *быть*, а отношения графа являются семантическими отношениями. Смотри рисунок 1.

*Он считал этого человека великим ученым.*



Рисунок 1 – Семантический граф

Использование семантического анализа для осуществления поиска предполагает следующий алгоритм. На первом этапе необходимо построить семантическое представление доступных ресурсов. На втором этапе, оперативном, анализируется поисковый запрос, на его основе строится семантический граф и производится поиск этого подграфа в семантически размеченном содержимом ресурсов. Таким образом, сам поиск происходит достаточно быстро и не требует больших вычислительных мощностей, в отличие от первого, подготовительного этапа. Поиск подграфа осуществляется путём строгого сравнения с добавлением других усовершенствований, таких, как подключение словаря синонимов, словаря эквивалентных структур русского языка, выделение главных семантических отношений в поисковом запросе и, как следствие, ослабление некоторых других семантических отношений. Возможны и другие оптимизации.

В итоге, к примеру, на запрос информации «Швеция была атакована» с использованием метода семантической разметки получаем:

- Швеция была атакована с запада.
- Был отдан приказ атаковать Швецию.
- Выиграл битву, атаковав Швецию внезапно.
- Сборная США атаковала уже потерявшую бдительность Швецию.

Сравните с результатами поиска на тот же запрос информации в Google:

- По данным военного архива Швеции, ... была атакована финская канонерская лодка.
- Швеция усиленно готовилась к войне против России. ... был атакован шведской гребной флотилией.
- «Атаковать должна Россия» / Сборную Швеции вполне устроит ...

Итак, какие преимущества имеет предложенный подход перед обычным, несемантическим поиском? Во-первых, использование такой реализации семантического поиска позволит повысить релевантность поисковых результатов. Этот метод по эффективности будет сравним с обычным поиском на простых запросах, состоящих из одного слова, либо из нескольких, не связанных между собой по смыслу. Однако на сложных запросах, которые содержат хотя бы одно семантическое отношение, семантический поиск работает гораздо эффективнее. Повышенная релевантность запросов достигается за счёт учёта семантических отношений, разрешения омонимии и вычленения семантически значимых частей запроса. Во-вторых, такой подход является очень гибким и удобным в использовании совместно с интерфейсом уточнения запроса, который включает в себя возможность строгого совпадения слов-отношений, времени глаголов и прочие характеристики.

## СПИСОК ЛИТЕРАТУРЫ

- 1 Yusuke Fukazawa, Jun Ota. Semantic SemSearch2010 – Semantic Search 2010 Workshop: [Electronic resource]. – Mode of access: <http://km.aifb.kit.edu/ws/semsearch10/Files/modelling.pdf>. – Date of access: 27.10.2010.
- 2 Krisztian Balog, Edgar Meij, Maarten de Rijke. Entity Search: Building Bridges between Two Worlds: [Electronic resource]. – Mode of access: <http://km.aifb.kit.edu/ws/semsearch10/Files/entity.pdf>. – Date of access: 27.10.2010.
- 3 Tomasz Imielinski, Jinyun Yan. Paraphrasing Invariance Coefficient: Measuring Para-Query Invariance of Search Engines: [Electronic resource]. – Mode of access: <http://km.aifb.kit.edu/ws/semsearch10/Files/para.pdf>. – Date of access: 27.10.2010.
- 4 Question Answering Based on Semantic Graphs / Lorand Dali, Delia Rusu, Blaž Fortuna, Dunja Mladenić and Marko Grobelni: [Electronic resource]. – Mode of access: [http://km.aifb.uni-karlsruhe.de/ws/semsearch09/semse2009\\_22.pdf](http://km.aifb.uni-karlsruhe.de/ws/semsearch09/semse2009_22.pdf). – Date of access: 27.10.2010.
- 5 Исходные коды системы ДИАЛИНГ: [Электронный. ресурс]. – Режим доступа: <http://www.aot.ru/>. – Дата доступа: 27.10.2010.

Е. В. Паршикова

### **ЛЕКСИЧЕСКИЕ ОСОБЕННОСТИ ФЕМИННО ОРИЕНТИРОВАННЫХ РЕКЛАМНЫХ ТЕКСТОВ (на материале интернет-источников)**

Интернет – всемирная система объединенных компьютерных сетей, образующая глобальное информационное пространство, служащая физической основой для Всемирной паутины и множества других систем передачи данных. Сейчас наиболее популярные услуги Интернета – это: Всемирная паутина, веб-форумы, блоги, интернет-магазины, интернет-аукционы и многое другое. Интернет также предоставляет широкие возможности для рекламы.

В настоящее время реклама стала неотъемлемой частью жизни общества. Она справедливо определяется как часть общечеловеческой культуры, которая развивается по своим внутренним законам. Реклама формирует определенное мнение, необходимые предпочтения и оценки, а также создает запланированную иерархию ценностей. Реклама оказывает разное влияние на представителей мужской и женской половин человечества, поэтому данное исследование направлено на изучение гендерных характеристик печатных рекламных текстов [4, с. 44].

В последние годы заметные изменения в рекламе происходят по отношению к поло-ролевой идентификации: все чаще печатный рекламный текст приобретает гендерную направленность. Под гендером понимается социальный пол, определяющий поведение человека в обществе и то, как это поведение воспринимается.

Гендерные стереотипы особо ярко проявляются в рекламе косметики. Отличие мужской рекламы от женской в данном случае проявляется прежде всего в отборе лексики, представляющей мир адресата и его гендерные характеристики. При этом в женской рекламе сделан акцент на визуальной характеристике адресата, а в мужской – на психологической, что также свидетельствует о зависимости языковых особенностей рекламного текста от гендерных характеристик адресата [3, с. 20].

Большой материал для выявления лингвистической сущности рекламных текстов представляют собой такие наиболее популярные сайты, как <http://www.lady.mail.ru/>; <http://www.nissan.ru/>, <http://www.cosmo.ru/>. На этих сайтах можно не только прочесть статьи, связанные с различными сторонами жизни, но и найти множество рекламных текстов, посвященных всему, что необходимо женщине.

Феминно-ориентированные рекламные тексты можно разделить на несколько групп, в связи с использованием в них:

1. Научных подтверждений, названий компонентов, входящих в состав продукта. Это связано с тем, что женщина более проницательна, для нее не существует мелочей в рекламе, и очень часто детали заслуживают серьезного внимания, поэтому рекламные тексты, ориентированные на женскую аудиторию, всегда убедительны и искренны. Например: *Обновляющий пилинг для кожи лица ARTISTRY Intensive Skincare – первый продукт, в основе формулы которого лежит компонент натурального происхождения – энзим гриба *Mucor miehei*, по своему белковому составу близкого к человеческой коже. Кроме отшелушивания и выравнивания кожи, пилинг ARTISTRY Intensive Skincare обладает питательным и омолаживающим эффектом. Этому способствуют входящие в его состав **витамины (С и Е), пантенол, лецитин и питательные масла, богатые незаменимыми жирными кислотами (масло вечерней примулы, масло макадамии, масло подсолнечника, масло оливы, фруктовые масла).** Клинические исследования показали, что уже после первого применения в 99 % случаев участницы эксперимента отмечали повышение эластичности кожи и визуальное улучшение ее структуры.*

2. Особых форм речевых выражений: риторические вопросы, определители (слова, фразы, вводные слова, вводные предложения, интонации речи, междометия, смягчающие смысл категоричного утверждения); экспрессивные выражения, междометия, фразы с оттенками просьб и обращений: *Но что если у вас нет возможности посещать косметолога или вы опасаетесь попасть в руки мастера, который «подарит» вам кислотные ожоги вместо желанной молодости и красоты? Может ли уход за кожей в домашних условиях быть таким же эффективным, как салонная процедура? Лаборатория Artistry утверждает: «да», и представляет первый продукт новой линии – обновляющий пилинг для кожи лица.*

3. Лексем, направленных на изменение эмоционального состояния реципиентов. Использование этих лексических средств объясняется тем,

что феминности стереотипно приписываются такие характеристики, как эмоциональность, пассивность и неустойчивость [1, с. 21–24]. Это можно проиллюстрировать следующим рекламным текстом: *Компания Garnier выпустила **уникальный** первый роликовый корректор темных кругов под глазами. Состав **волшебный**: кофеин для **улучшения** оттока жидкости и экстракт лимона, **осветляющий** кожу. Ролик действует сразу в нескольких направлениях: **мгновенно маскирует** темные круги благодаря стойкому корректирующему крему, а также **стимулирует** кровообращение, что позволяет **избавиться** от припухлостей и отеков. К тому же роликовый аппликатор оказывает **приятное** охлаждающее и массажное действие. Так что **веселитесь, сколько хотите, – утро будет добрым!***

4. Стандартов, зачастую недостижимых для большинства женщин. В рекламных сообщениях, объектом оценки которых является женщина, типичный женский образ включает ряд черт, связанных со слабостью, беспомощностью, зависимостью, меньшими притязаниями. Внимание реципиента рекламного объявления приковывается не к социальной личности, а к женской привлекательности [2, с. 81–92]. Реклама концентрируется в основном на таких женских качествах, как молодость, стройность, красота: ... *Хочется, чтобы появился такой чудо-крем для рук, который бы **одним махом удовлетворил широкий спектр наших запросов**. Новая линия кремов «Природная забота» от Garnier содержит растительные компоненты, **увлажняет и питает** кожу в течение дня (наносить нужно один раз в день), **укрепляет ногти, смягчает кутикулы** (а это увеличивает интервал между маникюром), **моментально впитывается и не оставляет жирных следов**. Кроме того, кремы обладают приятным ароматом, в зависимости от компонента в составе.*

5. Цветообозначений, связанных с образом красоты и здоровья (белый, бледно-сиреневый, персиковый, нежно-голубой, розовый, бежевый, зеленый – естественные краски природы, символизирующие красоту и свежесть весны): *Add a few drops of natural food coloring to your bath water – it's a great way to boost your mood, and it won't dye your skin. Try **red** for energy, **blue** to calm you, or **orange** for happiness.* – красный цвет здесь ассоциируется с энергичностью, голубой – со спокойствием, оранжевый – со счастьем.

6. Образных средств выражения оценки: эпитетов, метафор, сравнений, игры слов и т. д.: ***I like men but I couldn't eat a whole one** (Я люблю мужчин, но не смогла бы съесть целого). 17 Super shiny lip colour. 17 – it's not make-up. **It's ammunition** (Это – амуниция).*

Таким образом, феминно ориентированные рекламные тексты характеризуются как использованием определенных стереотипов, так и специфическими лингвистическими средствами.

#### СПИСОК ЛИТЕРАТУРЫ

1 Гольдман, И. А. Практика рекламы / И. А. Гольдман, Н. С. Добробабенко. – Новосибирск, 1991. – С. 3–33.



2 Гусейнова, И. А. Гендерный аспект в текстах современной рекламы (на материале журнальной прессы ФРГ) / И. А. Гусейнова, М. В. Томская // Филологические науки. – 2000. – № 3. – С. 81–92.

3 Гермогенова, Л. Ю. Эффективная реклама в России. Практика и рекомендации / Л.Ю. Гермогенова. – М.: РусПартнер Лтд, 1994. – С. 20.

4 Пронина, Е. Е. Психологическая экспертиза рекламы: теория и методика психотехнического анализа рекламы / Е. Е. Пронина. – М. : РИП-холдинг, 2002. – С. 44.

Л. В. Савченко

## **ДОСЛІДЖЕННЯ ЕТНОКУЛЬТУРНОЇ ЕТИМОЛОГІЗАЦІЇ ФРАЗЕОЛОГІЗМІВ ЧЕРЕЗ ІНТЕРНЕТ-РЕСУРСИ**

Стрімкий розвиток у сучасному світі інформаційних технологій і Інтернету одержав процес комп'ютерно-опосередкованого дослідження наукових проблем через засоби електронних текстів. Електронний текст – це письмове, текстове утворення, підготовлене для зберігання в електронній формі, призначене для сприйняття з екрана монітора комп'ютера, розташоване на електронному носії і яке може бути подане в комп'ютерну мережу (за С. Данилюком) [2, с. 15].

Інтернет-ресурси стали невід'ємною частиною в роботі з вивчення лінгвістики, зокрема етимології. В Україні поки що, на жаль, не створено етимологічний словник української фразеології, і ця проблема для українських фразеологів є однією з першочергових, про що акцентував увагу відомий фразеолог В. Мокієнко на конференції в Белгороді (вересень 2010 р.).

Етнокультурну основу етимології українських фразеологізмів досить всесторонньо можна досліджувати через Інтернет-ресурси. У сучасному мовознавстві відзначається тенденція до вивчення національно-культурної значущості фразеологізмів. Посилений інтерес викликає відображення в мові світоглядних уявлень, вірувань, традицій етносу, в контексті взаємодії культури і мови.

Теоретичні основи сучасних досліджень фразеології та пареміології в аспекті взаємозв'язку понять «мова – культура – етнос» заклали праці російських учених М. Толстого, В. Мокієнка, І. Подюкова, В. Телії. Питання історичного розвитку етнофразеології і, зокрема, етимології фразеологічних одиниць (ФО), на жаль, ще недостатньо досліджені в науковій літературі. На цьому неодноразово наголошували провідні фразеологи В. Мокієнко, Ю. Прадід, В. Коваль, Л. Скрипник, В. Ужченко та ін. Формування історичної фразеології української мови, на думку М. Алефіренка, «відбувається в трьох напрямках: фраземотворення, етимології і походження окремих фразеологізмів [1, с. 775]. До наших днів дійшло багато ФО, походження яких для фразеологів досі є загадковим. Оскільки-

ки ФО досить виразно відбивають позамовну дійсність і базуються на образній основі, корені їх етимології необхідно шукати в національній культурі, а саме традиціях і обрядах. Із цього приводу слушно зауважує В. Телія: «Відтворюваність фразеологічних знаків сприяє трансляції культурно значущих установок від покоління до покоління і цим забезпечує формування – у процесах оволодіння мовою і користування нею – культурного самопізнання як окремої особистості, так і культурно-національної ідентичності народу – носія мови» [6, с. 16]. На думку В. Мокієнка, «за відгадкою справжньої внутрішньої логіки вислову необхідно в першу чергу звертатися до мовних, а лише потім до історичних, культурних, етнографічних та подібних фактів» [5, с. 45]. Поділяємо погляд білоруського етнологвіста В. Ковалю, який стверджує, що «при реконструкції внутрішньої форми етнофразем значно зростає роль відомостей етнокультурного характеру, що на практиці виявляється в домінуванні позамовних даних над власне мовними» [3, с. 117].

Особливий інтерес викликають ті ФО, які кодують культурні цінності народу (наприклад, *палю забивати, причепити квітку, заговорити зуби, вилами по воді писано, з долоні вичитати, дати відкоша*). Культурну інформацію можна отримати з внутрішньої форми фразеологізму, в якій наявні «сліди» культури – міфи, архетипи, звичаї і традиції, відображені історичні події й елементи матеріальної культури. Ця інформація потім «ніби «воскресає» в конотаціях, які відображають зв'язок асоціативнообразної основи з культурою (архетипами, еталонами, символами, стереотипами)» [4, с. 271].

Оскільки у фразеології простежується постійна динаміка фразеологічного складу, необхідно досліджувати ФО саме в діахронному зрізі, щоб остаточно не втратити першоджерела походження окремих ФО, які виникли й функціонували ще кілька століть тому. Для вивчення історії походження української фразеології в нашому розпорядженні є багатий і цінний фактичний матеріал, зафіксований у фольклорних та етнографічних збірках, лексикографічних працях, які відіграли важливу роль у становленні та розвитку української фразеології і вплинули на формування фразеологічного фонду української мови. Вимагають детального опрацювання праці К. Зіновієва, О. Павловського, П. Білецького-Носенка, М. Закревського, М. Номиса, К. Шейковського, П. Чубинського, Ф. Піскунова, М. Комарова, І. Франка, Б. Грінченка та ін., які репрезентують ФО, уживані кількома поколіннями українців.

Дослідження етнокультурної етимологізації фразеологізмів через Інтернет-ресурси здійснюється в кількох напрямках: *по-перше*, дослідження сайтів лексикографічних праць; *по-друге*, дослідження сайтів фольклорних збірок; *по-третє*, вивчення сайтів архівних матеріалів і стародруків; *по-четверте*, опрацювання сайтів наукових праць; *по-п'яте*, аналіз сайтів, де розміщено культурну та етнографічну інформацію; *по-шосте*, аналіз сайтів, де міститься інформація щодо лексичного складу мов з метою етимологічно-

го визначення тощо. Матеріали з Інтернету значно полегшують роботу над вивченням словникових статей низки вищезазначених праць. Створення сайтів, на яких містяться не тільки словники та лексикографічні праці, а й копії архівних матеріалів, а саме етнографічних та фольклорних збірників (XIX – поч. XX ст.), періодичної преси (XVIII–XIX ст.), вважається справжнім проривом у комп'ютерній лінгвістиці, оскільки доступ до інформації є відкритим. Наприклад, фольклорна збірка І. Франка «Галицько-руські приповідки» (1901–1910 рр.) та лексикографічна праця Б. Грінченка «Словник української мови» (1903–1909 рр.) в електронному варіанті подані у форматі pdf, що дає змогу зручно переглядати матеріал, проте, на жаль, відсутня досконала система копіювання. Фольклорна збірка М. Номиса (1864 р.) створена у форматі djvu, що дає можливість через пошукову систему швидко знайти те чи інше стійке сполучення слів за ключовим словом чи коренем ключового слова і відтворити необхідну копію.

Для того щоб з'ясувати наявність досліджуваних одиниць у складі сучасної фразеосистеми, використовуємо «Фразеологічний словник української мови» (1993 р.), «Словник фразеологізмів української мови» (2003 р.) та ін. До аналізу фрагментарно залучаються фразеологізми та стійкі вислови слов'янських мов, що зафіксовані в електронних словниках, із метою виявлення більшої їх контрастності або паралелізму. Як відомо, практично більшість сучасних лексикографічних праць доступні в комп'ютерному варіанті в режимі on-line, який максимально спрощує алгоритм роботи зі словниковими статтями. Систему on-line, застосовувану у векторі лінгвістики, на нашу думку, можна вважати важливим проривом у досягненнях сучасних інноваційних технологій. Варто зауважити, що російська фразеологія та фразеографія досліджена значно глибше та фундаментальніше. Видано низку словників та довідників, у яких подана історико-етимологічна інформація щодо походження фразеологізмів, які розміщені на сайтах Інтернету. На жаль, відсутня електронна версія білоруського фразеологічного словника, що не дає можливості досліджувати фразеологію в порівняльному аспекті двох мов. Як відомо, через ФО відображається наше історичне минуле, у якому український народ заснував і узвичаїв надзвичайно велику кількість традицій та обрядів, чимало з яких своїм корінням сягають часів язичництва.

В останні роки на Інтернет-сайтах можна знайти етнографічні матеріали різноманітних видань XIX–поч. XX ст. (електронні копії книг подані у форматі pdf): «Живая старина» (1890–1913 рр.), «Русский филологический вестник» (1879–1901 рр.), «Киевская старина (1882–1907 рр.)», «Записки Императорскаго русскаго географическаго общества (по отделению этнографіи)» (1869–1913 рр.), «Филологическіе записки» (1868–1907 рр.), «Сборникъ отделения русскаго языка и словесности Императорской Академіи Наукъ» (1877–1902 рр.), «Записки Українського Наукового Товариства в Києві» (1909 р.). Інформація цих видань у багатьох випадках проливає світло на походження низки фразеологічних одиниць, оскільки містить етнографічні та

фольклорні розвідки про давні традиції та обряди, вірування та уявлення людини. Цікаво, що на сайтах можна знайти наукові статті, культурні та етимологічні словарі, а також матеріали, в яких фіксується лінгвістичний аналіз деяких стійких словосполучень слів.

Важливим кроком до вивчення генези фразеологізмів є опрацювання сайтів наукових праць із фразеології та етимології, аналіз культурної та етнографічної інформації, лексичного складу мов із метою етимологічного визначення тощо. Інтернет-ресурс дає можливість за допомогою пошукової системи (наприклад, Google) одержати інформацію про будь-яку лексему, її семантику, переклад слова, дані про походження слів, фразеологізмів тощо. Крім того, в ній подано джерела, що характеризують поняття культури, етнографії, фольклору, психології, філософії.

Отже, для дослідження історії та етимології українських фразеологізмів необхідно звертатися до комп'ютерних інформаційних технологій – Інтернет-ресурсів. Завдяки вищеописаним сайтам Інтернету сучасні лінгвісти мають доступ до такої інформації, з якою раніше можна було ознайомитися лише в архівах, бібліотеках тощо.

#### СПИС ЛІТЕРАТУРИ

1 Алефіренко, М. Ф. Фразеологія / М. Ф. Алефіренко // Українська мова: Енциклопедія. – К. : Наук. думка, 2004. – С. 775.

2 Данилюк, С. С. Структурні та національні особливості англomовних електронних тестів / С. С. Данилюк. – К. : АР, 2006.

3 Коваль, В. И. Славянские этнофраземы как языковой феномен / В. И. Коваль // Славянская фразеология в ареальном, историческом и этнокультурном аспектах: Материалы международной научной конференции. – Гомель : ГГУ им. Ф. Скорины, 2001. – С. 115–120.

4 Маслова, В. А. Homo lingualis в культуре / В. А. Маслова. – М., 2007. – 320 с.

5 Мокиенко, В. М. Образы русской речи. Историко-этимологические очерки фразеологии / В. М. Мокиенко. – Спб.: Фолио-Пресс, 1999. – 464 с.

6 Телия, В. Н. Русская фразеология. Семантический, прагматический и лингвокультурологический аспекты / В. Н. Телия. – М., 1996. – 215 с.

А. А. Станкевіч

### **ПРАГРАМНАЯ ПАДТРЫМКА ЛІНГВІСТЫЧНАГА ДАСЛЕДАВАННЯ ЛЕКСІКІ ІНШАМОЎНАГА ПАХОДЖАННЯ Ў БЕЛАРУСКІХ НАРОДНЫХ ГАВОРКАХ**

Адным з напрамкаў выкарыстання вылічальнай тэхнікі для лінгвістычнага даследавання ў наш час з'яўляецца інвентарызацыя тэкставага матэрыялу і яго статыстычная апрацоўка. Мэта нашага паведамлення – паказаць на матэрыяле вывучэння іншамоўнай лексікі ў народных гаворках магчымасці праграмнай падтрымкі лінгвістычнага даследавання.

Міжмоўныя кантакты з'яўляюцца адной з істотных крыніц папаўнення слоўніковага саставу любой мовы. Яны адлюстроўваюць разнастайныя зносіны паміж народамі. Працэс узаемадзеяння моў, які распаўсюджваецца на розныя ўзроўні моўнай сістэмы (фаналагічны, марфалагічны, лексічны, семантычны і сінтаксічны), найбольш эфектыўны ў галіне лексікі. Такая ўласцівасць лексічнай сістэмы абумоўлена рэфэрэнтнай функцыяй слова як асноўнай намінацыйнай моўнай адзінкі.

Міжмоўныя кантакты адыгралі важную ролю ў развіцці лексічнай сістэмы народна-дыялектнай мовы. Вынікам іх з'явіліся шматлікія запазычанні, найбольш значныя ў колькасных адносінах і разнастайныя ў семантычным плане групы якіх складаюць словы польскага, германскага, балтыйскага і лацінскага паходжання. Інтэнсіўнасць запазычвання іншамоўнай лексікі ў значнай ступені вызначалі ўмовы білінгвізму, а ў некаторых выпадках і полілінгвізму, у якіх развівалася беларуская мова, найбольш актыўным з якіх было польска-беларускае моўнае ўзаемадзеянне. Гэтым тлумачыцца тое, што польская мова з'явілася крыніцай запазычвання самай значнай па колькасці групы (36 %) лексем, а таксама была пасрэднай пры засваенні большасці слоў з германскіх, раманскіх і класічных моў.

Запычванне іншамоўнай лексікі – заканамерны вынік разнастайных эканамічных, сацыяльна-палітычных і культурна-гістарычных сувязей беларусаў з іншымі народамі. Засваенне запазычанняў народна-дыялектнай мовай у большасці выпадкаў было абумоўлена камунікатыўнымі задачамі – патрэбай у намінацыі новых з'яў і паняццяў у жыцці грамадства на розных этапах яго развіцця, а таксама імкненнем да разнастайнасці намінацыі і экспрэсівізацыі выражэння.

Намі было прааналізавана звыш трох тысяч лексічных адзінак іншамоўнага паходжання народна-дыялектнага ўжытку.

Мэтай даследавання было паказаць асноўныя заканамернасці запазычвання і функцыянавання лексікі іншамоўнага паходжання ў народна-дыялектнай мове. Умовай для дасягнення гэтай мэты было рашэнне наступных задач: 1) сістэматызаваць народна-дыялектныя запазычанні на лексіка-тэматычнай аснове; 2) даць іх агульную гісторыка-этымалагічную і храналагічную характарыстыку; 3) паказаць структурна-тыпалагічныя асаблівасці дыялектных запазычанняў; 4) апісаць іх арэальныя тыпы; 5) выявіць формы і ступень адаптацыі дыялектнай лексікі іншамоўнага паходжання, вызначыць спецыфіку і адметнасць яе засваення ў народнай мове, фактары, якія абумоўліваюць гэты працэс; 6) правесці квантытатывы аналіз дыялектных запазычанняў, які адлюстроўвае асноўныя статыстычныя заканамернасці іх функцыянавання і засваення.

Аб'ектам аналізу дадзенага даследавання з'явілася апеляцыйная лексіка іншамоўнага паходжання народна-дыялектнага ўжытку. Асноўным крытэрыем пры адборы фактычнага матэрыялу паслужыў рэгіянальны характар запазычанняў – у рабоце аналізаваліся ў першую чаргу іншамоўныя

словы, якія адсутнічаюць у літаратурнай мове або маюць пэўныя фармальныя ці семантыка-стылістычныя адрозненні ад сваіх нарматыўных адпаведнікаў, а таксама асобныя агульнанародныя запазычанні, якія вызначаюцца сталай традыцыяй выкарыстання ў дыялектнай мове.

Пры даследаванні іншамоўнай лексікі асноўная ўвага была сканцэнтравана на праблеме яе тыпалагізацыі, выяўленні заканамернасцей функцыянавання і адаптацыі запазычанняў у народных гаворках на тэрыторыі Беларусі. У рабоце выкарыстоўваўся пераважна апісальны метада даследавання, як дапаможныя прымяняліся статыстычны метада – пры вывучэнні квантытатывнай (колькаснай) характарыстыкі запазычанняў, кампанентны – пры вызначэнні іх семантычнай структуры, лінгвагеаграфічны – пры выяўленні іх арэальных асаблівасцей.

Матэрыялам даследавання паслужылі апелятыўныя лексічныя адзінкі іншамоўнага паходжання, крыніцай збору якіх з’явіліся асноўныя выданні беларускай дыялектнай лексікаграфіі, асобныя фальклорныя і этнаграфічныя запісы XIX–XX ст.ст., уласныя назіранні над лексічнымі асаблівасцямі гаворак Навагрудчыны. Пры напісанні работы выкарыстоўваўся таксама матэрыял гістарычнага, этымалагічнага і тлумачальнага слоўнікаў беларускай мовы.

У працэсе функцыянавання ў дыялектнай лексічнай сістэме іншамоўная лексіка падвяргаецца паступоваму засваенню ў адпаведнасці з асноўнымі заканамернасцямі фанетычнай, граматычнай і лексіка-семантычнай сістэмы мовы-рэцэптара. Працэс адаптацыі запазычанняў у народна-дыялектнай мове носіць спецыфічны характар. Ён значна адрозніваецца ад засваення іншамоўнай лексікі літаратурнай мовай больш поўнай асіміляцыяй, імкненнем максімальна наблізіцца да ўнутраных заканамернасцей мовы-рэцэптара.

Адаптацыя запазычанняў можа быць фармальнай (фанетычнай і марфалагічнай) і семантычнай. Пры фармальным засваенні змяняецца знешняя форма запазычанняў – іх фанемны склад, марфалагічная будова, граматычная характарыстыка, словаўтваральная структура. Пры адаптацыі знешняй формы запазычанняў у гаворках праяўляецца, з аднаго боку, тэндэнцыя да поўнай асіміляцыі іншамоўнага лексічнага матэрыялу ў адпаведнасці з фанетычнай сістэмай і граматычным ладам народна-дыялектнай мовы. З другога боку, лакальны характар, вусная форма існавання дыялектнай мовы абумоўліваюць узнікненне разнастайных варыянтаў – лексіка-фанетычных і марфалагічных, якія парушаюць аднастайнасць адаптацыі запазычанняў. Лексіка-фанетычныя і марфалагічныя варыянты запазычанняў адлюстроўваюць, як правіла, розныя этапы і формы адаптацыі – адны з іх па сваёй форме набліжаны да этымонаў, іншыя – з’явіліся ў выніку мадыфікацыі іх гучання і граматычнай формы ў адпаведнасці з фанетычнай сістэмай і граматычным ладам пэўнай гаворкі. У працэсе адаптацыі не ўсе запазычанні вар’іравалі сваю знешнюю форму, а толькі 37 %, больш чым трэць з іх – даўнія запазычанні. Найбольшай

актыўнасцю пры вар'іраванні формы вызначаліся канкрэтна-прадметныя намінацыі.

Вар'іраванне знешняй формы і значэння запазычанняў, якое з'яўляецца вынікам іх засваення ў народных гаворках, адлюстроўвае розныя этапы і формы адаптацыі. Вышэйшай ступенню адаптацыі запазычанняў з'яўляецца іх семантычная эвалюцыя, якая мае ў гаворках адметны характар і вызначаецца разнастайнасцю тыпаў полісемантызацыі, іх камбінаваннем, і ў многіх выпадках прыводзіць да дыфузнасці сэнсавага аб'ёму, няпэўнасці яго семасіялагічных межаў.

Важнейшым паказчыкам засвоенасці запазычанняў з'яўляецца рэалізацыя іх дэрывацыйных здольнасцей ў выніку ўключэння ў сістэму словаўтварэння мовы-рэцэптара ў працэсе функцыянавання і ўзаемадзеяння з уласнымі лексічнымі сродкамі. Найбольш актыўна ў словаўтварэнні ўдзельнічае канкрэтна-прадметная лексіка, аднак самы высокі словаўтваральны патэнцыял маюць назвы з працэсуальнай семантыкай, для якіх характэрна дэрывацыйная гнуткасць і разнастайнасць спосабаў структурнай матывацыі.

Ступень фармальнай і семантычнай адаптацыі дыялектных запазычанняў вызначаецца рознымі фактарамі – храналагічным (часам запазычвання), арэальным (тэрыторыяй распаўсюджвання), семіятычным (важнасцю і актуальнасцю называемых паняццяў у жыцці грамадства), частотнай характарыстыкай (прыналежнасцю да актыўнага або пасіўнага лексічнага фонду).

Характар адаптацыі запазычанняў залежыць таксама ад шляхоў і спосабаў іх засваення (непасрэднага ў працэсе прамых кантактаў або апасродкаванага, у многіх выпадках шматступянёвага, пры якім адсутнічае непасрэднае міжэтнічнае моўнае ўзаемадзеянне), абумоўлены генетычнай тоеснасцю і тыпалагічным падабенствам кантактуючых моў (запазычанні з неславянскіх моў падвяргаюцца больш істотнай фармальнай і семантычнай трансфармацыі ў параўнанні са словамі, засвоенымі з іншых славянскіх моў).

Для сістэматызацыі фактычнага матэрыялу і зручнасці яго апрацоўкі была распрацавана база даных, якая ўключае 3112 даследаваных іншамоўных лексічных адзінак з іх асноўнымі параметрамі (прыметамі) структура-семантычнай, функцыянальна-стылістычнай і арэальнай характарыстыкі: *паходжанне* (паводле крыніцы запазычвання і зыходнага этымона); *сувязь са значэннем этымона* (супадзенне, несупадзенне і тып семантычнага развіцця – звужэнне, пашырэнне, метафарычны або метанімічны перанос); *тып запазычання* (фанетычны, марфалагічны, словаўтваральны, лексічны, семантычны); *шлях запазычвання* (непасрэдны, апасродкаваны); *даўнасць запазычвання* (наяўнасць / адсутнасць у старабеларускай мове); *часцінамоўная прыналежнасць* (назоўнік, дзеяслоў, прыметнік, прыслоўе, лічэбнік, займеннік і г. д.); *тэматычная група* (назвы прадметаў, асобы, прыродных з'яў, адцягненых паняццяў і інш.); *асаблівасці семантыкі* (адназначнае / мнагазначнае); *знешняя форма* (аднастайнае / варыянтнае),

*словаўтваральны патэнцыял* (наяўнасць / адсутнасць вытворных); *стылістычная афарбоўка* (нейтральнае / экспрэсіўнае); *сфера ўжывання* (агульнаўжывальнае, спецыяльнае); *актуальнасць ужывання* (актуальнае, пасіўнае, устарэлае); *тэрыторыя распаўсюджвання* (з указаннем населеных пунктаў, у якіх зафіксавана выкарыстанне дадзенай моўнай адзінкі).

Аб'ём базы даных – больш за 50 кілабайт. На аснове распрацаванай базы даных напісана камп'ютарная праграма (мова Fox Pro 20) якая дазваляе: а) карэкціраваць усю інфармацыю базы даных; б) устаўляць новыя і выводзіць старыя запісы; в) ствараць любыя запыты для выбаркі з базы даных і даследавання інфармацыі; г) дае магчымасць захаваць запыты, друкаваць іх вынікі.

Асноўная ідэя праграмы – выявіць пэўныя заканамернасці ў характарыстыках слоў і на аснове гэтага прасачыць асаблівасці запазычвання, функцыянавання і адаптацыі лексікі іншамоўнага паходжання ў народна-дзялектнай мове. Праграма мае лёгкі аконны карыстальніцкі інтэрфейс, простая і эфектыўная ў рабоце.

Без камп'ютарнай праграмы было б чыста фізічна немагчыма выявіць асноўныя статыстычныя заканамернасці ў характарыстыках іншамоўных слоў, правесці аналіз і падагульненне гэтай інфармацыі.

Такім чынам, упершыню пры аналізе запазычанняў з мэтай сістэматызацыі фактычнага матэрыялу была распрацавана база даных па вывучэнні асноўных асаблівасцей іх функцыянавання і засваення народна-дзялектнай мовай, складзена камп'ютарная праграма па апрацоўцы базы даных, у якой рэалізаваны алгарытм па выяўленні асноўных статыстычных заканамернасцей выкарыстання і адаптацыі лексікі іншамоўнага паходжання ў беларускіх гаворках.

Камп'ютарная праграма, выкарыстаная ў рабоце, дае ўзор універсальнага даследавання статыстычных заканамернасцей канкрэтнага моўнага матэрыялу. Яна можа быць запоўнена любой лінгвістычнай інфармацыяй і выкарыстана як пры навуковай інтэрпрэтацыі сістэмных уласцівасцей моўных адзінак, так і ў якасці практычнага інструмента ў навучальным працэсе пры засваенні пэўных лінгвістычных навыкаў і ўменняў.

А. Л. Стрижак

## **ИСПОЛЬЗОВАНИЕ РЕСУРСОВ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА ДЛЯ ИЗУЧЕНИЯ ДЕТЕРМИНОЛОГИЗАЦИИ В СОВРЕМЕННЫХ ПУБЛИЦИСТИЧЕСКИХ ТЕКСТАХ**

Характерной особенностью публицистических текстов является одновременная реализация ими двух функций языка – сообщения и воздействия. Яркий, образный язык журналиста делает произведение более эмоциональным, «заряжает» его определённой оценочной коннотацией – положительной или отрицательной, является одним из важнейших условий адек-



ватного восприятия авторских интенций адресатом. Одним из наиболее эффективных средств создания экспрессивного высказывания является метафора, лежащая в основе явления детерминологизации, т. е. проникновения научной терминологии в общепотребительную лексику. «Попадая в общелитературный язык, термин приобретает лексическое значение и становится словом, отмеченным в стилистическом отношении, его преследует особый «технический ореол», почему он и используется нередко как определённое стилистическое средство. При этом экспрессивная окраска термина возникает из сопоставления, которое появляется в нашем сознании, между «строгим» научным термином и обычными словами» [1, с. 80].

Язык сегодняшней прессы повышенно метафоричен – как правило, метафоризации подвергаются объекты и явления, вызывающие повышенный общественный интерес. Нами были выделены две большие тематические группы, одна из которых отражает авторские оценки важнейших событий политической жизни, другая – реакцию журналистов на определённые явления социальной сферы. Проведённый анализ публицистических текстов показал, что наряду с использованием терминов, новое, переносное значение которых уже зафиксировано в современных словарях, в периодике последних двух десятилетий активно используются термины из различных сфер научного знания, ещё только осваивающиеся общелитературным языком, не получившие в нём достаточно широкого распространения. В процессе работы нами активно использовались ресурсы Национального корпуса русского языка (НКРЯ), в частности – Корпуса современной русской прессы, который составляют газетные тексты и сообщения новостных Интернет-агентств за период 2000–2008 гг. объёмом около 100 млн словоупотреблений. Важно отметить, что фактическое отсутствие ограничений на объём анализируемого материала и скорость поиска информации открывает перед исследователем совершенно новые возможности. Обращение к ресурсам НКРЯ, сопоставление найденных примеров с самостоятельно собранным фактическим материалом (выписками из русскоязычных белорусских газет за период 1990-х годов) позволило нам проследить частотность использования журналистами научных терминов и на основании этого высказать предположения относительно возможности (при активном использовании их в современной публицистике), либо маловероятности (при отсутствии такого использования) перехода научных терминов в активный словарный запас.

Примером обеспокоенности публициста современными ему политическими проблемами может служить метафорическое использование медицинского термина *конвульсии* в следующем контексте: *По всей видимости, кипят **предвыборных конвульсий** должен отстояться, и за близкими ликами, лживыми и циничными обещаниями мы должны увидеть настоящих кандидатов – тех, кого мы знаем и в ком уверены* (Сов. Белоруссия, апр. 1995). Большой толковый словарь русского языка определяет переносное значение данного термина как ‘последняя стадия существования

каких-л. социальных и т.п. явлений' [2, с. 435], что вполне согласуется и с содержанием примера из НКРЯ: *Что делать государству, озабоченному построением вертикали власти?... Народным избранникам, сотрясаемым лишь **политическими конвульсиями**? Учителям, мамам, папам?* (Комс. правда, янв. 2002) Очевидно, что в результате метафоризации научного термина в обоих случаях происходит коннотативное обогащение значения лексемы, реализуются авторские интенции негативной оценки действительности и тем самым достигается эффект экспрессивного воздействия на адресата. Аналогичным примером использования специальной лексики, но уже с дополнительным оттенком иронии, служат следующие контексты, где в переносном значении 'резкое неприятие кого-, чего-либо' [2, с. 35] используется медицинский термин *аллергия*: *Мне кажется, что наметившаяся было активность правительства по реформированию сложного и малоэффективного управленческого аппарата резко замедлилась. Мне кажется, что наши министры вновь впали в привычную **административную аллергию*** (Нар. газета, нояб. 1993) и ещё один пример – из НКРЯ: *Причиной тому стала, по-видимому, непробиваемая глухота, **«аллергия» власти к мыслям, рекомендациям выдающегося мыслителя по «обустройству» России в новых условиях...*** (Труд-7, май 2000).

Достаточно часто негативные оценочные смыслы развиваются в метафорических значениях слов, не зафиксированных в толковых словарях: *Политик также отметил идеологическую «всеядность» партий и **блоков-«гермафродитов»** (БЮТ, НСНУ, ПР), чью ориентацию вообще крайне сложно определить: они одновременно и «левые», и «правые», а на деле «заняты своими вопросами», к которым избиратель не имеет ни малейшего отношения* (Новый регион, нояб. 2007). Вероятно, что отправной точкой для развития негативной коннотации лексемы *гермафродит* 'животное или человек, соединяющее признаки мужского и женского пола' [2, с. 200] послужил один из дифференциальных признаков, составляющих содержание термина, а именно – представление о ненормальности явления, отклонении от нормы, что в обыденном сознании «плохо», тогда как соответствие норме – «хорошо». Подобный механизм развития негативной семантики специального слова *дистрофия* 'нарушение питания тканей и внутренних органов, организма в целом, ведущее к их истощению' [2, с. 262], базирующийся на бытовом представлении о слабости как «неправильности», «ненормальности» наблюдается и в следующем примере: *Единственно приемлемой формой будет создание кальки со всего политического поля и после некоторой **умышленной дистрофии**, облачение его представителей легитимностью: прямо как в притче – всякой твари по паре* (Ноев ковчег, июль 2007). [оба примера взяты из НКРЯ]

Высокая степень экспрессии нередко достигается при окказиональном использовании научной лексики: *То депутат Кучинский командным голосом потребует провести очередную прокурорскую проверку, то в очередной раз, на манер Данко, выхватит своё пылающее сердце депутат*

*Наумчик и попытается осветить им путь несколько осовевшим от этой политической кардиохирургии коллегам...* (Нар. газета, нояб. 1993). Индивидуально-авторское употребление термина *кардиохирургия* 'раздел медицины, занимающийся лечением сердца хирургическими методами' [2, с. 417] придаёт высказыванию яркую ироническую окраску, наполняет метафорическое значение слова негативным оценочным смыслом, а также создаёт неявно выраженный комический эффект. Комизм достигается не только тем, что термин попадает в несвойственный его употреблению контекст, но и самой разговорной, непринуждённой манерой повествования, которую подчёркивает просторечная лексема *осовевший*.

Активная гражданская позиция современного публициста заставляет его не только быть в курсе политических событий, но и уделять должное внимание более «насушным» проблемам жизни людей, освещать негативные явления социальной сферы, указывать (если такое возможно) пути выхода из кризисных ситуаций. Следуя логике подачи материала, обратимся к примерам детерминологизации специальных слов, переносное значение которых зафиксировано в словарях: *В отличие от машинистов, депутаты, конечно же, знают, что не только в нашей Беларуси, но и в более опытных с точки зрения демократических традиций странах, за искусственный паралич транспорта по головке не погладят* (Сов. Белоруссия, авг. 1995), также рассмотрим контекст из НКРЯ: *Люди, застрявшие под землей в метро и в лифтах домов, паралич общественного транспорта, массовая порча продуктов в холодильниках магазинов, прерванные операции в больницах, выброс нечистот в Москва-реку – все это позволяет говорить о чрезвычайной ситуации в столице* (Труд-7, июнь 2005). Не вызывает сомнения наличие семы отрицательной эмоциональной оценки в коннотативной зоне значения термина *паралич* 'состояние вынужденного бездействия, покоя' [2, с. 780] каждого из данных примеров, что обусловлено вполне объективными факторами: бездействие городского транспорта ставит под угрозу нормальное функционирование всего городского организма.

Негативное отношение говорящего к достаточно противоречивым переменам в сфере образования и – опосредованно, через использование самого термина *эвтаназия* 'добровольная, согласованная с врачом, смерть неизлечимо больного с помощью специальных обезболивающих средств' [2, с. 1512] к явлению, им обозначаемому, – находим в следующем контексте: *Не зная теории и практики педагогики...можно потратить массу сил и средств на создание элитарного образования для «достойных», а всех остальных, «недостойных», освободить от тяжелых мук познания. В итоге – ... интеллектуальная эвтаназия, которая оборачивается не возрождением, а деградацией нации* (Сов. Белоруссия, июль 1994). Нетипичное использование лексемы *эвтаназия* в следующем примере (контекст из НКРЯ) обусловлено общим позитивным настроением коммуникантов – допуская возможность закрытия юмористической программы, один из её авторов говорит: *Юрий Стоянов: а если почувствуем спад*

интереса к программе, или, не дай бог, друг к другу, то сделаем нашему «Городку» *эвтаназию* (Комс. правда, июль 2007). Подобное словоупотребление «снимает» негативный оценочный ореол метафоризованного термина, но не лишает его выразительности, напротив, семантическое ядро лексемы обогащается дополнительными оттенками смысла.

Одним из путей индивидуально-авторского использования специальных слов вне научной сферы является приём языковой игры. Так, имея своей целью акцентировать внимание адресата на «покорности» и «смирении» рядового гражданина, его зависимости от власть имущих, автор использует не совсем корректное с точки зрения грамматики терминологическое сочетание: *На руках у населения, особенно у его законопослушной части, сосредоточено значительное количество огнестрельного оружия и боеприпасов. ... А что делать нам, добропорядочным обывателям, чья жизнь проходит исключительно под знаком «страдательного наклонения»?* (Нар. газета, нояб., 1996) Совмещая в пределах одного словосочетания значения лингвистических терминов *наклонение* 'категория глагола, выражающая отношение к действительности' [2, с. 585] и *страдательный* 'обозначающий состояние лица или предмета, испытывающего на себе чьё-либо действие' [2, с. 1276], говорящий получает идиоматическое окказиональное образование в значении 'существование в полном повиновении кому-либо, рабское существование'. Следует также обратить внимание и на развитие в последние годы глаголом *наклонять* нового значения 'принуждать к совершению каких-либо действий с позиций силы', соответственно, существительное *наклонение* приобретает значение 'процесс по действию глагола *наклонять*' – если автор учитывал новое значение данной лексемы, то мы можем говорить о двойном семантическом плане высказывания, поскольку в этом случае имеет место намеренное столкновение узуального значения слова *наклонять* и его новой семантики. Нельзя не согласиться с тем, что подобные семантические трансформации позволяют генерировать в высказывании не только негативный оценочный смысл, но ещё и мощный экспрессивный заряд, позволяющий говорящему достигнуть значительного эстетического эффекта в воздействии на реципиента.

Таким образом, детерминологизация научных терминов на основе метафоризации раскрывает богатый семантический потенциал такого рода лексики, позволяет достигнуть яркой эмоциональной оценочности, создать запоминающиеся образы, актуализирующиеся благодаря семантической двуплановости, и реализовать тем самым такие опорные для публицистики лингвокреативные категории как экспрессивность, эмотивность и прагматичность в выражении оценки.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Капанадзе, Л. А. О понятиях «термин» и «терминология» / Л. А. Капанадзе // Развитие лексики современного русского языка. – М. : Наука, 1971. – 381 с.
- 2 Большой толковый словарь русского языка. – СПб. : «Норинт», 2000. – 1536 с.

## НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА КАК ИСТОЧНИК ИЗУЧЕНИЯ ФЕМИННЫХ НОМИНАЦИЙ В ДИАХРОНИЧЕСКОМ АСПЕКТЕ

Интерес современного человека к своему прошлому, к своим истокам заставляет его внимательно всматриваться в языковые единицы – трансляторы культуры. В связи с этим в последнее время возрастает интерес к исследованиям, раскрывающим ментальность народа через язык, выявляющим закрепленные в нем культурные традиции, определяющим роль внеязыковых данных в значении лексических и фразеологических единиц. Сказанное в полной мере относится к анализу номинаций лиц женского пола. Важно при этом учитывать системные связи отдельных лексико-семантических групп, отражающие системную организацию словарного состава языка в целом. Важным источником для подобного рода исследований являются ресурсы Национального корпуса русского языка.

В русской народной речи существуют слова, отдельно называющие разные возрастные классы мужчин и женщин. Эти слова обозначали не физический возраст, вернее не только его, но целую область правил, представлений и отношений, которые подразумевались за таким определением. Женские возрастные определения таковы: *дева/девка/девушка* (от рождения до замужества), *молодка/молодая/молодуха* (от замужества до первого ребенка или до первой дочери), *баба* (замужняя женщина, но не хозяйка в доме), *большуха* (хозяйка «сама») [3; с.41].

Первоначально лексема *дева* имела единственное значение – ‘девушка’. Лексема *девица* имело два значения, первое из которых было ‘девушка’, второе же являлось метафорическим: ‘вид пищали’ (*Пищал девица ядромь и длиною такова ж, весом 24 пуда*) [1; с.197]. Подобное метафорическое значение мы наблюдаем также и для лексем *девка*. *Девка* – ‘самая малая кость при игре в бабки’: *Панок-от литой, тяжелый, девка полегче* [2; с.192]. В этом случае метафорический перенос происходит на основе уподобления по форме, размеру, противопоставления «мужского» – «тяжелого» и «женского» – «легкого». Кроме этого, лексема *девка* имела еще три значения: ‘девушка, девочка’, ‘дочь’ (– *Ну, молодец девка моя Вера, – говорила мужу Марья Алексевна, удивленная таким быстрым оборотом дела: – гляди-ко, как она забрала молодца-то в руки!* Н. Г. Чернышевский. Что делать?) и ‘прислуга, дворовая женщина’: *Да и народ совсем взбунтовался, слушать перестают; девка моя и та грубить стала* (Л. Н. Толстой. Война и мир). Но в современном русском языке значение ‘прислуга, дворовая женщина’ устарело, а значение ‘дочь’ приобрело разговорный оттенок. Таким образом, лексема *девка* оставляет за собой только одно значение – ‘девушка, девочка’, которое, однако, является просторечным: *У нас там были прекрасные соседи по площадке, престарелые муж с женой, но*

рядом, за стеною, жила молодая девка, работавшая проводницей вагона (В. Астафьев. Затеси).

В более поздний период функционирования языка все рассмотренные выше лексемы (*дева*, *девица*, *девка*) переживают расширение своей семантики, что связано с расширением областей познания человека в различных сферах. Так, лексема *дева*, кроме прежнего значения ‘молодая незамужняя девушка’ приобретает значение ‘зодиакальное созвездие’: *Дева: ты готова к новому свиданию, и его вероятность на этой неделе предельно велика...* (Твой гороскоп // «Даша», № 10, 2004).

Лексема *девица*, кроме прежнего значения ‘молодая незамужняя женщина’ (*Молодой монах из кельи, / Притаясь, в окно глядит: Под окном, внизу, девица / Зарумянившись стоит* (М. Михайлов. Молодой монах) приобретает значение ‘фрейлина, девушка-придворная (*Тогда служащие девицы, всем собором, Царевну вновь одеть пришли*), что обусловлено изменениями, произошедшими в социально-культурной жизни людей.

Лексема *девка* сохраняет свои прежние значения ‘молодая незамужняя женщина’ и ‘служанка’, но вместе с тем приобретает и два новых значения, до этого не встречавшиеся в памятниках письменности: а) ‘публичная женщина (прост.)’: *Таська, есть в сауне такая девица-мастерица, их обслуживала и запомнила разговор* (Лев Корнешов. Газета); б) ‘о робком, стыдливом, стеснительном мужчине’: *Гришин вроде не красна девица, чтобы бизнесмены приходили и ему деньги давали* (Николай Леонов. Лекарство от жизни).

В современном русском языке все три рассматриваемые лексемы имеют лишь одно значение – ‘молодая незамужняя женщина’. Все остальные (употреблявшиеся ранее) значения устаревают либо приобретают просторечную стилистическую окраску.

В этих примерах очевидна тенденция расширения либо сужения семантики рассматриваемых лексем на протяжении исторического развития в связи с теми или иными социально-историческими и культурными событиями, происходившими в жизни людей на разных этапах.

*Молодуха* – женщина в первый год брака. Название *молодуха* женщина получала в ходе свадебного ритуала на следующий день после брачной ночи и сохраняла его вплоть до рождения первого ребенка. Включение *молодухи* в семью мужа, в новый для нее деревенский коллектив, а также в социовозрастную группу замужних женщин происходило в результате обрядов, которые были чрезвычайно разнообразны и носили как семейно-родственный, так и общинный характер [3; с. 360]. *Вид у новобрачного Лешки был тогда невеселый, а невеста, вернее молодуха, Борису очень понравилась* (Владимир Корнилов. Демобилизация); *И видим – идет не то девка, не то молодуха* (П. П. Бажов. Веселухин Ложок); – *Вот, вот, – шумно вздыхая поддерживает тетку Дуню молодуха Глаша.* (Л. Воронцова. Двадцать пять верст по Калязинскому уезду). В последнем контексте

уменьшительная форма имени собственного *Глаша* является показателем незрелости и несамостоятельности положения молодухи в русской деревне.

Однако на протяжении исторического развития в связи с теми или иными социально-историческими и культурными событиями, происходившими в жизни людей, также начала появляться тенденция расширения семантики лексемы *молодуха*. Так, в разговорной речи в последнее время всё чаще и чаще можно встретить употребление этого слова в несвойственном для него значении (особенно ярко это демонстрируют тексты газетного корпуса русского языка): *В местной ФСБ красивая молодуха, очень похожая на комиссаршу домкома из «Собачьего сердца», отперла мне тяжелую дверь и скомандовала мужским сильным голосом: «Стойте здесь! (Кто убил губернатора Колымы // Комс. правда, 2002.12.04); Но жители Псотина носа в притон не показывали, обходили стороной и «чертов дом», и его хозяйку: веселая «мать-одиночка» Ольга безвылазно прожила в деревне три месяца, но никто из соседей так и не мог сказать, ни чем живет молодуха, ни кого и зачем принимает, ни даже как зовут не посещавшую местную школу восьмилетнюю девочку (Андрей Родкин. В Подмоскowie зарублены топором семь человек // Комс. правда, 2002.02.12).*

Таким образом, значение, развивающееся у лексемы *молодуха* с течением времени в устной разговорной речи – ‘молодая женщина вообще’, ‘девушка’. Зачастую можно проследить пренебрежительный оттенок: *Фигуристая молодуха с большим ртом энергично рассказывала о погоде, о циклоне и антициклоне (Токарева Виктория. Своя правда // «Новый Мир», № 9, 2002); – А сколько вы нам заплатите? – спрашивает заплетаящимся языком молодуха. – Тридцатку! (Дмитрий Ковшуля. Три дня из жизни бензовоза // «Звезда», 2003).*

Лексема *большуха* имеет значение ‘старшая по положению в доме среди женщин’. Ею, как правило, являлась жена большака. Большуха обязательно должна была быть замужней женщиной, не очень старой, но имевшей детей. В ее обязанности входило следить за порядком в семье, а также готовить пищу для всех живущих в доме, тогда она называлась еще стряпухой: *Все руководство домашним хозяйством держала в руках большуха – женщина, жена и мать (О крестьянских истоках национального бытия (2002) // «Жизнь национальностей», 2002.10.14); Иногда таким авторитетом был наделен или дед, или один из сыновей, или большуха, тогда как формальное главенство всегда принадлежало мужчине, мужу, отцу, родителю (О крестьянских истоках национального бытия (2002) // «Жизнь национальностей», 2002.10.14).*

В ведении *большухи* находились все женские работы, которые она распределяла среди женской половины семьи: кормление и дойка коров, рубка дров, прядение, ткачество, прополка огорода, уборка, стирка, приготовление пищи. Последнее, как правило, выполняла сама *большуха*. *Большуха* по праву старшинства наделялась большой властью в доме: «своя воля – своя большая». Говорили: «Она всю большую забрала себе в руки».

В руках хозяйки сосредотачивалась власть над дочерьми и невестками, поэтому считалось большой удачей выдать дочь замуж в дом, где нет свекрови, снохи и золовок, в таких случаях молодая сразу становилась хозяйкой: *Баба начинает точить: «Ишь, сколько рубах за весну на работе спарил – не наготовишься на тебя, а тут и прясть некогда: большуха-то дома сидит, а ты то за водой, то корму скоту задай»* (А. Н. Энгельгардт. Письма из деревни).

Окончательный «выход на большинство» молодой хозяйки мог произойти только с того времени, когда она приобретет все качества, необходимые для ведения хозяйства: проживет больше года в семье, родит детей. И тем самым получит необходимую спору и жизненную силу, долю. А до этого момента она находилась в положении *молодухи* [3; с.66].

Но в разговорной речи в последнее происходит расширение семантики лексемы *большуха*. При анализе материала Национального корпуса русского языка было выявлено два новых значения: а) ‘старшая дочь’: *У старичка есть дочь, большуха, которая тоже неподвижно полулежит на лавке, под образами, и тоже не может сказать, когда она родилась, а помнит только что в тот год, когда была «некручина»* (М. Е. Салтыков-Щедрин. Современная идиллия); б) ‘взрослый, зрелый человек’, при этом в последнем значении пол лица не играет принципиальной роли: *Я уже такой большуха!*

Представленные в Национальном корпусе русского языка образцы устной речи очень важны, потому что люди пишут не так, как говорят; в особенности это различие ощутимо для языков с давней письменной традицией, за время существования которой нормы письменной и устной речи успевают разойтись достаточно сильно. Письменная речь всегда более консервативна; к тому же в русском языке письменная традиция формировалась под сильным иноязычным влиянием (первоначально старославянским, впоследствии немецким и французским), которое гораздо меньше затронуло устную речь. Таким образом, если мы хотим выявить наиболее динамичные структуры живого русского языка, мы должны обратиться к стихии устной речи, не скованной традицией и нормой.

Таким образом, Национальный корпус предназначен в первую очередь для обеспечения научных исследований лексики языка и процессов, происходящих в языке на протяжении небольших периодов – от одного до двух столетий. Другая задача корпуса – предоставление всевозможных справок, относящихся к указанным областям. Благодаря отражению материала разносторонне (газетный, устный, диалектный корпус) мы можем проанализировать не только контексты, в которых употребляется та или иная лексема, но также провести более развернутый анализ, позволяющий проникнуть в особенности менталитета русского народа, их быт и традиции.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Словарь русского языка 11–17 вв. Вып. 4 (Г-Д). – М. : Наука, 1975. – 404 с.



2 Словарь говоров русского севера. Т. 3 Г-Ж. – Екатеринбург : Издательство Уральского университета, 2005. – 388с.

3 Баранов, Д. А. Мужики и бабы: мужское и женское в русской традиционной культуре / Д. А. Баранов, О. Г. Баранова, Т. А. Зимина. – Спб. : Искусство, 2005. – 688с.

А. Г. Шукайлова

## СИМВОЛИКА ЗООНИМА *ЗАЯЦ* В РУССКОЙ И КИТАЙСКОЙ КУЛЬТУРАХ (по данным интернет-источников)

В лексике любого языка зоонимы представляют собой весьма специфический пласт, так как в этой группе слов ярче, чем в любой другой области языка, отражаются особенности осмысления внеязыковой реальности. Зоонимические представления восходят к глубинам человеческого сознания, его верованиям и мифологии. Многие зоонимы выступают в роли символов, то есть того, что служит условным знаком какого-нибудь понятия, явления, идеи. В силу огромного разнообразия свойств (движение, образ жизни, форма, окраска и т. д.) и отношения к природе и человеку, животные играют едва ли не важнейшее значение в символике. Её истоки традиционно рассматриваются в тесной связи с тотемизмом и поклонением животным. В самом общем виде, можно говорить, что животные символизируют инстинктивную жизнь, плодородие и изобилие, эмоциональные порывы, которые необходимо обуздать, прежде чем вступать в сферу духовного и многое другое. Достаточно часто выбор уподобления человека животному делается с целью выделения для его описания какого-либо присущего ему положительного или отрицательного признака, поскольку невозможно найти животное, которое не имело бы хоть какого-то сходства с человеком.

Животные могут получать положительную и отрицательную оценку. К числу негативно коннотируемых зоонимов в русской культуре можно отнести зоолексему *заяц*. В переносном значении этот зооним применяется для того, чтобы охарактеризовать трусливого человека, что отражено в поговорах: *Труслив как заяц; Дрожит как заяц; Вор что заяц и тени своей боится; Жаден, как волк, а труслив, как заяц; Труслив, что заяц, блудлив, что кошка; Лучшие умереть орлом, чем жить зайцем и др.* Распространено представление, что заяц из-за своей боязни всего и всех спит с открытыми глазами: *Спит, как заяц* – говорят о том, у кого чуткий сон. Трусливость зайца объясняют тем, что у него маленькое сердце. По легенде Бог вылепил ему слишком длинные уши, а на сердце глины не хватило. Тогда Бог оторвал ему хвост, оставив лишь небольшой отросток, и сделал маленькое сердце. Образ трусливого зайца наиболее популярен в современном сознании. С этим мотивом связан и *заяц* как безбилетный пассажир или дезертир.

В то же время заяц в русской культуре осмысливается как плодовитое животное, наделяемое в народных представлениях эротической (фаллической) символикой и демоническими свойствами. В песенном фольклоре встречается

мотив совокупления зайца с девушкой: *Зайнька серенький, / Да не ходи по селям, / Не топай ногою, / Я лягу с тобою. – Заяушка, с кем ты спал да ночевал? / – Спал я, спал я, пане мой, / Спал я, спал я, сердце мой, / У Катюхи – на руке, / У Марюхи – на грудях, / А у Дуньки вдовиной на всем животе.*

Сверхплодовитость и готовность к спариванию делают зайца символом похотливости. С этим связаны и представления о том, что сон о пойманном зайце предвещает женщине беременность и рождение сына. Кроме того, идея плодовитости выражена в объяснении детям, что их приносят зайцы. Заячья кровь используется в народной медицине в качестве средства от бесплодия, а заячий жир применялся для смазывания женских органов при трудных родах.

Вместе с тем заяц тесно связан с нечистой силой. Именно поэтому в народе люди называют зайца *косой*. Считается, что он находится в распоряжении лесных духов. Леший может проиграть зайцев соседнему лешему. Согласно некоторым поверьям, заяц создан чертом и служит ему. Черта иногда представляют себе с заячьим хвостом. Представлением о зайце как существе опасном и «нечистом» объясняется и примета о том, что заяц, перебежавший дорогу или встреченный на пути, сулит путнику несчастье, например, охотнику и рыбаку – неудачу в ловле, сплавщику плотов – мель или разбитые плоты и т. д.



Святая Феврония  
Муромская

В христианстве заяц также означает плодовитость и похоть. Несмотря на это, заяц весьма часто появляется в христианской иконографии. Белый заяц у ног Девы Марии означает триумф над похотью. Его беззащитность олицетворяла людей, уповающих лишь на Бога. Заяцы, отдающие детям гроздь винограда, являются чувственными образами душ, взятых в рай, где они могут в полной безопасности вкушать плоды вечной жизни. Вскользь встречаются представления о трех зайцах в круге, чьи уши образуют треугольник. Может быть, это указание на божественную тройственность либо скоротечность времени в его беге по кругу.

В китайской культуре зооним *заяц* имеет ярко выраженную положительную символику. поскольку заяц является символом долголетия. По легенде он – знаток целебных снадобий. Если хорошо попросить, заяц поделится эликсиром, который готовит в волшебной ступе на Луне. Кроме бессмертия, этот эликсир приносит счастье и удачу. Однажды заяц накормил собой Будду, бросившись в костер, и тем самым утолил его голод. За это Будда определил Заяца на почетное место – на Луну и посвятил его в тайны целебных рецептов. Лунный Заяц помогает детям получить покровительство Луны, а значит – хорошую семью и счастливую жизнь в будущем.

Фигурка кролика, символизирующая плодовитость, может оказать благотворное воздействие на семейные отношения. Кроме того, известно, что в учении Фэн-шуй традиционно используются старинные китайские

монеты. Одна из таких монет – китайская монета счастья *mǎo*. На «иньской» стороне монеты, в центре, находится инь-янь. На «янской» стороне монета имеет изображение – амулет одного из двенадцати животных китайского календаря – кролика. В дни праздника Луны белые фигурки кроликов выставляют в Китае на окнах. Необходимо заметить, что в символике заяц и кролик не различаются.

Вместе с тем, заяц в китайской культуре заяц является олицетворением очень хитрого и умного человека : *shī zì bó tú, yì yòng quán lì* (букв.: для того, чтобы льву бороться с зайцем, нужно использовать всю свою силу), *jǐāo tù sān kǔ* (букв.: у хитрого зайца три дома).

В некоторых пословицах актуализируется такой признак зайца, как быстрота. Например: *tù qǐ jīyī jù* (букв.: заяц скрылся, дикие утки улетели). Образ зайца также неразрывно связан со скоротечностью времени. В этом контексте интересно выражение *baí tù chī tù* (букв.: белый заяц, чёрная черепаха). Здесь заяц, как символ луны, ассоциируется с ночью, а черепаха – с днём.

Кроме того, заяц символизирует выгоду, которая может появиться сама собой. Известно выражение *chōu zhū dài tù* (букв.: сидя смотреть на пень и ждать зайца), т. е. хотеть получить выгоду без старания. Данная пословица связана с легендой о том, что однажды господин Лао увидел, как заяц, выбегая из леса, ударился о пень и издох. После этого случая господин Лао проводил все дни возле этого пня в ожидании очередного зайца.

Таким образом, можно сделать вывод о том, что зооним *заяц* имеет символическое значение как в русской, так и в китайской культуре. Подтверждением тому служат легенды, пословицы и поговорки, представленные в языках-источниках. В русской культуре зоолексему *заяц* можно отнести к числу негативно коннотируемых зоонимов. В китайской культуре зоолексема *заяц* имеет ярко выраженную положительную символику.

#### СПИСОК ЛИТЕРАТУРЫ

1 [Электронный ресурс] – Режим доступа: <http://studentu911.narod.ru/articles/fsh/2.html> - Дата доступа: 25.10.2010.

2 [Электронный ресурс] – Режим доступа: <http://moy-bereg.ru/simvolika-zhivotnyih/simvolika-nekotoryih-zhivotnyih.html> – Дата доступа: 28.10.2010.

3 [Электронный ресурс] – Режим доступа: <http://sigils.ru/seals/zaez.html> – Дата доступа: 25.10.2010.

4 [Электронный ресурс] – Режим доступа: <http://dic.academic.ru/dic.nsf/simvol/268> – Дата доступа: 25.10.2010.

5 [Электронный ресурс] – Режим доступа: <http://www.symbolsbook.ru/Article.aspx?id=221#1> – Дата доступа: 25.10.2010.

6 [Электронный ресурс] – Режим доступа: <http://www.pagan.ru/slowar/z/zayac0.php> – Дата доступа: 25.10.2010.

# КОРПУСНАЯ ЛИНГВИСТИКА: ПРОБЛЕМЫ РАЗРАБОТКИ, ВЕДЕНИЯ И ИСПОЛЬЗОВАНИЯ КОРПУСОВ

A. Andrzejczuk

## KORPUSY JĘZYKOWE W PRACY JĘZYKOZNAWCY: NOWE MOŻLIWOŚCI CZY DROGA NA MANOWCE?<sup>1</sup>

Korpusy językowe podbiły współczesne językoznawstwo. Świadczy o tym chociażby popularność takich przedsięwzięć, które mają na celu utworzenie korpusów. W zasadzie coraz więcej krajów dąży do tego, żeby stworzyć własny korpus narodowy.<sup>2</sup> W Korpusach tych zgromadzone są teksty z różnych odmian regionalnych i stylistycznych języka danego kraju. Popularne są też równoległe korpusy dwu- i wielojęzyczne, w których umieszczane są zwykle dzieła literackie w ich oryginalnym języku oraz ich przekłady. Powstają też próby tworzenia korpusów języka mówionego. Korpusem można również nazwać zasoby Internetu.

Korzyści płynących z umiejętnego korzystania z korpusów nie da się przecenić. Opis języka dzięki oparciu na autentycznych tekstach staje się wiarygodniejszy, a refleksje językoznawcze bardziej obiektywne od opisu opartego na tworzonych ad hoc przykładach. Trzeba jednak pamiętać, że korpusy to jedynie narzędzie. Same w sobie nie dadzą nam gotowych odpowiedzi. Wyniki uzyskane z ekscerpcji tekstów również będą bezwartościowe, jeżeli nie będziemy mieć do nich dystansu i zdroworozsądkowego krytycyzmu.

W swoim wystąpieniu chciałabym zaprezentować, w jaki sposób dzięki korpusowi możemy zweryfikować postulaty językoznawców. Wybrałam do zweryfikowania postulaty normotwórcze. Chciałam pokazać, w jakim stopniu są one w języku żywe, a w jakim są jedynie pobożnym życzeniem normatywistów. Muszę zaznaczyć, iż uważam – co chyba dość oczywiste – że norma językowa nie powinna stać w jawnej sprzeczności z uzusem. Bańko [1] pisze, iż niektórzy językoznawcy boją się, że akceptacja kryterium uzusowego doprowadzi do akceptacji form, które są powszechnie potępiane. Pokazuje na przykładzie form *\*szłem/szedłeś* i ich występowania w Korpusie Języka Polskiego PWN, a Andrzejczuk i Czupryniak [2] powtarzają tę kwerendę w Internecie, i w obu

<sup>1</sup> Praca naukowa finansowana ze środków na naukę w latach 2007–2010 jako projekt rozwojowy.

<sup>2</sup> Narodowy Korpus Języka Polskiego zawierający ok. 1 miliarda słów, Cambridge International Corpus - ok. 1,2 mld słów, British National Corpus – 100 mln słów, Český národní korpus – 300 mln słów, Nacional'nyj Korpus Russkogo Jazyka – 150 mln słów.

przypadkach niepoprawna forma \*szłem nie stanowi nawet 10 procent wszystkich wystąpień. Andrzejczuk i Czupryniak podsumowują: Już same liczby pokazują, że niepoprawna forma *szłem* to niecałe 8 procent wszystkich wystąpień, więc zasługiwałaby co najwyżej na status rzadkiej, a po analizie kontekstów – potocznej. Co więcej, gdy przejrzymy kilka pierwszych wyników, zauważymy, że znaczna część piszących ma świadomość, że jest to forma niepoprawna: Tak, możemy użyć tylko form *szedłem, szedłeś, szłaś, szła, szło, szliśmy*. Nigdy: *szłem, szłeś, szedłaś, szedła, szedło, szedliśmy*. Nawet w języku potocznym. (http:

//forum.gazeta.pl/forum/72,2.html?f=374&w=1268470&v=2&s=0)

Czasami forma niepoprawna służy stylizacji i wprowadzeniu humoru, np. w piosence Kabaretu Chatelet i Łony zatytułowanej *Do Ciebie szłem*, której tytuł i słowa można znaleźć na wielu stronach. Autentycznych przykładów użycia formy *szłem*, jako elementu idiolektu autora, znajdziemy więc znacznie mniej, niż się to wydaje na początku.

Bugajski [3] uważa, że zadaniem językoznawcy jest doskonalenie języka, a nie uleganie społecznym zwyczajom. Pytanie tylko, jaki sens ma tworzenie zasad poprawnościowych, które są w społeczeństwie martwe i ich zastosowanie budzi zdziwienie wśród native speakerów?

Jedną z takich reguł jest reguła mówiąca o połączeniach rzeczowników plurale tantum z liczebnikami. Zacytuję dwie definicje z *Wielkiego słownika poprawnej polszczyzny* [4]:

okulary *blp, D.* okularów «para szkieleł w jednej oprawie»: Nosić, włożyć, założyć okulary. Chodzić w okularach. Spojrzeć zza okularów. Okulary leżą na stole, *ale: dwoje, troje itd. okularów* leży na stole. Mieć dwie pary okularów.

urodziny *blp, D.* urodzin: Dzień urodzin. Obchodzić, wyprowadzić urodziny. W lutym obchodzimy w naszej rodzinie aż troje urodzin a. urodziny aż trzech osób (*nie: obchodzimy aż trzy urodziny*)

Zgodnie z normą wzorcową w polszczyźnie rzeczowniki PT powinny się łączyć z rzeczownikami zbiorowymi. Reguła ta jest logiczna, ale nie zmienia to faktu, że w praktyce jest ona żywa jedynie dla kilku specyficznych leksemów. Głównie liczymy w ten sposób *drzwi, skrzypce* i *sanie*. Sprawą wątpliwą pozostaje liczenie desygnatów o podobnej funkcji, mających nazwę pluralną, typu: *wrota, gęśle* czy *sanki*.

Połączenia rzeczownika *okulary, urodziny* czy *imieniny* z liczebnikiem zbiorowym w Narodowym Korpusie Języka Polskiego (dalej: NKJP) nie znalazłam. W Internecie zaś znalazłam kilkanaście przykładów frazy *dwoje okularów*, z czego kilka powtórzonych z tego samego tekstu. Ciekawe są dwa przykłady. Jeden jest ze słownika polsko-niemieckiego<sup>3</sup>, a drugi wystąpił jako przykład liczenia tego desygnatu w języku polskim i angielskim<sup>4</sup>. Dla porównania fraza *dwie pary okularów* występuje w Internecie ponad 400 razy. W NKJP również można ją znaleźć.

<sup>3</sup> <http://pl.pons.eu/polski-niemiecki/okular%C3%B3w>

<sup>4</sup> [http://linguaetextae.blogspot.com/2007\\_05\\_01\\_archive.html](http://linguaetextae.blogspot.com/2007_05_01_archive.html)

Trudniej w języku polskim znaleźć dowody na liczenie świąt czy uroczystości, których nazwa jest rzeczownikiem plurale tantum, ponieważ native speakerzy polszczyzny prędzej zrobią to w sposób opisowy niż użyją bezpośredniego połączenia z jakimkolwiek liczebnikiem. Pisała już o tym Jadacka [5].

Radzenie sobie z tym, co w Korpusach jest, tzn. odróżnianiem prawidłowych, częstych, powszechnych form od form błędnych, rzadkich, czy regionalnych to nie jedyne ciekawe zagadnienie. Interesującą kwestią są braki w Korpusie i podejście językoznawców do nich. Jadacka [5] postuluje:

Takie rzeczowniki, jak narta, wrotka, nartorolka czy pionierka, trampek, można opatrywać uwagą zwykle w lm, pod warunkiem podania wiarygodnego przykładu użycia tak oznaczonego wyrazu w liczbie pojedynczej, czyli udowodnienia, że liczba mnoga jest tylko częstsza, ale nie – wyłączna. Bez spełnienia tego wymagania cały opis, łącznie z postacią hasłową leksemu, rażąco narusza standardy naukowe.

Wydaje się, że Jadacka za bardzo chce zawierzyć Korpusom. Zapomina, że choć Korpusy są narzędziem niezastąpionym w pracy językoznawcy, to jednak nie idealnym. Przecież nie istnieje i nigdy istnieć nie będzie korpus zawierający wszystkie możliwe formy wyrazowe, czy konstrukcje składniowe. Nie należy zapominać o tym, o czym już pisał Saloni [6], że niektóre formy mogą zostać wyabstrahowane z tekstów nie na podstawie danych empirycznych, lecz na podstawie rozumowania, najczęściej przez analogię. Istnienie mogło być wywnioskowane z istnienia innych jednostek, do których w jakiś sposób podobna jest jednostka badana.

Gdyby poważnie potraktować postulat Jadackiej z zacytowanego wyżej fragmentu, to za „standardy naukowe” należałoby przyjąć bezkrytyczne szukanie wystąpień w każdej formie, usankcjonować rezygnację z korzystania wiedzy językowej i pozajęzykowej. Zresztą zapis ten musi budzić sprzeciw, wszak nietrudno się domyślić, że nazw butów zwykle używa się w liczbie mnogiej. Niemniej oczywiste jest, że każdą pojedynczą sztukę można określić odpowiednim rzeczownikiem w liczbie pojedynczej. Oczywiście zdaję sobie sprawę z tego, że jeśli nie mamy potwierdzeń w tekstach, to rodzi to czasem pewne problemy – choćby z rodzajem butów o nazwie *oficerki*. Jednakże, nawet gdybyśmy chcieli stwierdzić, że jest to wyłącznie PT, to i tak mamy problem z D. lm - nie mamy *oficerek* czy *oficerków*?

Z powodu zbytnej wiary w teksty zapisy słownikowe dochodzą do takich absurdów, że obiekty należące do tej samej klasy są w różny, zupełnie nieintuicyjny sposób notowane. Można to zauważyć, przeglądając niekonsekwentny sposób zanotowania rodzajów obuwia w ortograficznych słownikach języka polskiego.

Całkowicie inaczej do problemu występowania/niewystępowania form czasownikowych czasu przeszłego rodzaju nijakiego, które ze względów semantycznych są bardzo rzadkie, ale ze względów gramatycznych są niezwykle łatwe do utworzenia w polszczyźnie, podeszli Żmigrodzcy [7]. Zamiast

dywagować nad sensownością i komunikatywnością tych słów postanowili poszukać w tekstach przykładów użycia różnych czasowników zakończonych na *-toś, -tom*.

Okazało się, że całkiem sporo przykładów znaleźli. Mogli je poklasyfikować ze względu na funkcję, jaką pełnią one w zdaniu:

- forma orzeczenia dostosowuje się do rodzaju gramatycznego
- są wykorzystywane w sytuacjach, w których nadawca lub odbiorca komunikatu nie ma określonej płci, nie wiadomo jakiej jest płci lub nadawca chce ją ukryć przed odbiorcami
- wykorzystywane są, żeby podkreślić antropomorfizację
- pełnią funkcję ekspresywną
- wykorzystywane są w grach słownych, w celach zabawowych.

Znalezienie tych – należy podkreślić – bardzo rzadko występujących form w polszczyźnie było możliwe jedynie dzięki uogólnieniu problemu, szukaniu interesujących nas form nie jednego czy kilku wybranych leksemów, ale całej ich grupy. Takie podejście umożliwiło wysnucie wniosków logicznych, intuicyjnych i całkowicie zgodnych z wiedzą o języku i o świecie. Wnioski też są poza wszystkim oparte na autentycznych przykładach użycia, co powoduje, że są one nie tylko logiczne, ale i wiarygodne.

W dzisiejszym świecie chyba nikt już nie zaprzeczy, że przykłady korpusowe są obiektywnymi, przekonującymi argumentami. Dzięki nim mamy możliwość zweryfikować nasze indywidualne wyczucie, opinie i sądy dotyczące zjawisk językowych, zaktualizować przestarzałe czy nieprawdziwe informacje ze współczesnych słowników przejętych ze starszych opracowań, wyszukać nowe wyrazy, poznać ich pisownię i odmianę, zorientować się w jakich kontekstach są używane. Jednak korzystając z nich, zawsze należy pamiętać, że badacz nie może sobie pozwolić na rezygnację z analizy logicznej oraz odwoływania się do swojej wiedzy nie tylko językowej, ale i ogólnej. Wiedza, doświadczenie, intuicja, zdrowy rozsądek badacza nie mogą być zastąpione przez żadne narzędzia.

#### SPIS LITERATURY

- 1 Bańko, M. Z pogranicza leksykografii i językoznawstwa / M. Bańko. – Warszawa 2001.
- 2 Andrzejczuk, A. O wykorzystaniu zasobów internetowych w pracy językoznawcy / A. Andrzejczuk, M. Czupryniak. – Polonica, t. XXIX, 2009. – S. 41–45.
- 3 Bugajski, M. Językoznawstwo normatywne / M. Bugajski. – Warszawa, 1993.
- 4 Wielki słownik poprawnej polszczyzny PWN, pod red. A. Markowskiego. – Warszawa, 2006.
- 5 Jadacka, H. Pluralia tantum w opisie leksykograficznym / H Jadacka. – Warszawa 2006. – S. 161–169.
- 6 Saloni, Z. Co istnieje, a co nie istnieje we fleksji polskiej / Z. Saloni // Prace Filologiczne XXXVII, 1992. – S. 75–87.
- 7 Żmigrodzka, B. W sprawie form czasu przeszłego typu byłom, byłoś (na marginesie artykułu Krystyny Pisarkowej) / B. Żmigrodzka, Żmigrodzki P. // Język Polski LXXXV, 2005. – S. 59–65.

## ЭКСТРАЛІНГВІСТЫЧНЫ КАНТЭКСТ РЭПРЭЗЕНТАТЫЎНАСЦІ БЕЛАРУСКАГА КОРПУСА ТЭКСТАЎ

Як вядома, корпусная лінгвістыка займаецца вывучэннем корпусаў тэкстаў як моўных аб'ектаў. У сучаснай лінгвістыцы корпусныя даследаванні ўжо маюць свае традыцыі, прызнаных лідэраў, свае навуковыя цэнтры, метады і праблематыку. У той жа час, на беларускай глебе корпусная лінгвістыка – адносна новы напрамак, які яшчэ толькі афармляецца, набывае абрысы. Усталяванне корпуснай лінгвістыкі на Беларусі, у пэўнай ступені, пытанне часу.

Развіццё корпуснай лінгвістыкі ў свеце пачалося ў 60-я гады ХХ ст. паралельна са з'яўленнем камп'ютарных тэхналогій. Менавіта тады ўзнік корпусны стандарт у 1 млн словаўжыванняў. Сёння асобныя корпусныя рэсурсы дасягаюць некалькі мільярдаў адзінак, многія з іх прадстаўленыя ў Інтэрнэце. Стварэнне беларускага корпуса нават у 1 млн словаўжыванняў, тое над чым у апошнія 5 год працуюць беларускія навукоўцы, і яго з'яўленне ў Інтэрнэце – крок хаця і вельмі запаздалы, але неабходны. Беларуская корпусная лінгвістыка адстае ад сусветных тэндэнцый развіцця вельмі істотна, але не назаўсёды. Што дазваляе глядзець наперад з аптымізмам?

Уласцівая корпуснай лінгвістыцы падтрымка інфармацыйнымі і праграмнымі сродкамі дазваляе ў параўнальна кароткі тэрмін дасягнуць сусветных стандартаў у праграмным забеспячэнні корпусных даследаванняў нават з нуля. Усё дастаткова проста: навукова-тэхнічны прагрэс у геаметрычнай прагрэсіі скарачае дыстанцыю паміж віртуальнай і матэрыяльнай рэчаіснасцю. Глобалізацыя, у тым ліку ў лінгвістычнай сферы дазваляе ўжо сёння вырашаць амаль любыя пытанні праграмага забеспячэння пры мінімальным выдатках: многае ўжо зроблена і складзеныя папярэднікамі праграмы нецяжка прыстасаваць ці дапрацаваць. Выкарыстанне замежных распрацовак і набыццё пэўных ліцэнзійных прадуктаў дазволіла, напрыклад, польскім навукоўцам навучыцца эфектыўнаму выкарыстанню корпусных метадаў і выпрацаваць уласную стратэгію развіцця корпуснай лінгвістыкі на польскім матэрыяле.

У бліжэйшым асяроддзі, у Расіі і на Украіне, развіццю корпуснай лінгвістыкі ўдзяляецца належная ўвага, і шмат чаго, у выніку, зроблена. Асаблівай павагі заслугоўвае распрацоўка і ўдасканаленне Нацыянальнага корпуса рускай мовы (Нацыянальнага корпуса рускага языка) [6]. Названы праект знаходзіцца на ўзроўні замежных узораў.

Акрамя праграмных сродкаў развіццё корпуснай лінгвістыкі патрабуе эмпірычных матэрыялаў: слоўнікаў і тэкстаў. Акадэмічныя слоўнікі сучаснай беларускай мовы ёсць, але для корпусных распрацовак



патрабуецца іх прадстаўленасць на машынных носьбітах і максімальна спрошчаны доступ неабмежаванага кола навукоўцаў і проста патэнцыяльных карыстальнікаў да адпаведных электронных рэсурсаў. Пакуль што слоўнік-эталон беларускай мовы немагчыма знайсці ў Інтэрнэце. У «сусветным павуцінні», з іншага боку, можна знайсці аматарскія распрацоўкі і прататыпы слоўнікаў беларускай мовы рознага кшталту. У кожнага карыстальніка беларускамоўнага Інтэрнэта ў выніку – свой слоўнік: на базе тарашкевіцы ці трасянку, нярэдка на лацініцы. Адсутнасць (у электронным выглядзе і свабодным доступе) нарматыўных слоўнікаў, на жаль, перашкаджае стварэнню, напрыклад, Інтэрнэт-рэсурсаў адпаведных крытэрыям нацыянальнага корпуса. Наяўнасць жа ў Інтэрнэце ненарматыўнай беларускай мовы і адсутнасць электроннага эталона, аказвае проста катастрафічны ўплыў на стан беларускай гутарковай мовы і ў далейшым, між іншым, звязіць магчымасці выкарыстання сённяшняй беларускай мовы Інтэрнэта як крыніцы любога корпуса ўзуальнай беларускай мовы.

Збіранне тэкстаў на беларускай мове, у прынцыпе, не ўяўляе сабой невырашальнай задачы. Ёсць пытанне аўтарскага права. Але, па-першае, абмежаванні аўтарскага права ўжо амаль не закранаюць класіку беларускай літаратуры, па-другое, можа і не трэба імкнуцца спяшацца творы спрэчнай якасці ўключаць у прадстаўнічы корпус, калі аўтар супраць. Ёсць архівы перыядычных выданняў, спецыяльная літаратура. Ужо сёння праграма пошуку па беларускіх тэкстах, размешчаных, напрыклад, на сайце газеты «Звязда» ([www.zvyazda.minsk.by](http://www.zvyazda.minsk.by)) альбо на «Беларускай палічцы» ([www.knihi.com](http://www.knihi.com)), дае тыя ж магчымасці, што і любы неразмечаны (неанатаваны) камп'ютарны корпус, а значыць, можа выкарыстоўвацца для разнастайных лінгвістычных мэтаў [3, 4]. Уключэнне надрукаваных у дзяржаўных сродках масавай інфармацыі і выдавецтвах артыкулаў і мастацкіх твораў, відавочна, пры выкарыстанні такіх тэкстаў у нацыянальным корпусным праекце не будзе выклікаць асаблівых цяжкасцей з пункту гледжання аўтарскага права.

Дарэчы, маюцца прыклады паспяховай працы беларускіх даследчыкаў над корпусным праектам. Корпус, створаны ў навукова-даследчай лабараторыі інтэлектуальных інфармацыйных сістэм на факультэце прыкладной матэматыкі Белдзяржуніверсітэта (загадчык лабараторыі доктар фізіка-матэматычных навук прафесар І. В. Соўпель) па заказе Міністэрства інфармацыі, задумваўся як паралельны руска-беларускі корпус матэрыялаў пасяджэнняў Палаты прадстаўнікоў Нацыянальнага Сходу Рэспублікі Беларусь, заканадаўчых актаў і пад., а рэальна, пасля пашырэння зыходнай тэкставай базы, стаў рэпрэзентатыўным корпусам беларускай мовы, з якога пры дапамозе сродкаў камп'ютарнай падтрымкі лёгка вылучаюцца субкорпусы, прыдатныя для розных прыватных даследчых мэтаў. Хочацца спадзявацца, што названы корпус калі-небудзь з'явіцца ў Інтэрнэце.

Дзейнасць па стварэнні ці кампіляцыі корпуса тэкстаў вельмі разнастайная. Крытэрыі адбору тэкстаў для свайго корпуса стваральнік задае, зыходзячы з мэтаў сваёй практычнай ці навуковай дзейнасці: паказчыкам рэпрэзентатыўнасці для яго будзе служыць патрабаванне максімальна адлюстраваць у створаным корпусе вывучаемую з'яву. Нават аб'яднанне ў адным файле адабраных па пэўных крытэрыях тэкстаў адбываецца, як правіла, з улікам папярэдне вызначанай мэты, што немагчыма без усведамлення металінгвістычных умоў стварэння і функцыяніравання абраных матэрыялаў.

Экстралінгвістычны аспект корпусных распрацовак мае практычнае значэнне, аказваючы сур'ёзны ўплыў на даследчы працэс у корпусным рэчышчы. Паводле крытэрыяў адбору тэкстаў і рэпрэзентатыўнасці корпусы распадаюцца на два класы. У першы клас уваходзяць корпусы тэкстаў, мэты якіх – адлюстраваць аб'ектыўную карціну маўленчай дзейнасці. Толькі суцэльны, без суб'ектыўнага адбору корпус, напрыклад, нацыянальны корпус, дазваляе атрымаць поўную, сістэмную карціну стану мовы. Метадалогія пабудовы корпусаў першага тыпу так ці інакш грунтуецца на прынцыпе дэдукцыі – ад агульнага (аб'ектыўнай моўнай практыкі носьбітаў мовы) да абмежаванага корпуса тэкстаў, які адлюстроўвае гэта агульнае.

У другі клас уваходзяць корпусы, пабудаваныя для адлюстравання пэўнай з'явы. У гэтым выпадку прыходзіцца задаволіцца тымі ці іншымі выбаркамі з агульнай сукупнасці моўных дадзеных. Але гэтыя выбаркі павінны адпавядаць як універсальна-статыстычным, так і спецыяльна-прадметным крытэрыям якаснай прадстаўнічасці выбаркі для лінгвістычнага даследавання. Па гэтай прычыне метадалогія пабудовы корпусаў другога класа павінна займацца праблемай карэктнасці адлюстравання асобнага лінгвістычнага феномена ў корпусе тэкстаў, прызначаным адлюстраваць гэты феномен.

Спіс корпусаў, створаных замежнымі лінгвістамі для пэўнай прагматычнай мэты, проста бясконцы, але большая частка з іх, што характэрна, пазней выкарыстоўваецца для мэтаў, больш шырокіх за першапачатковыя.

Поўныя ці рэпрэзентатыўныя моўныя корпусы і корпусы, якія ствараюцца для прыватных лінгвістычных задач, рэдка існуюць незалежна – як правіла, мэты корпусаў першага і другога тыпу сінкрэтычныя. З гэтага лагічна вынікае відавочная і важная выснова аб выкарыстанні ўжо акумуляваных рэсурсаў. Нават калі для пэўнай мовы (напрыклад, беларускай) яшчэ не створаны прадстаўнічы агульнадаступны корпус, то прыватныя даследчыя корпусы трэба рабіць агульнадаступнымі, максімальна моўна нарматыўнымі, і найлепшы сродак для гэтага – Інтэрнэт. З іншага боку, агульнадаступнымі, вельмі пажадана, павінны быць нарматыўныя слоўнікі ў электронным выглядзе.

Пад рэпрэзентатыўнасцю разумеецца, у тым ліку, здольнасць корпуса тэкстаў адлюстроўваць усе ўласцівасці праблемнай вобласці, рэлевантныя для дадзенага тыпу лінгвістычнага даследавання, у пэўнай прапорцыі, якая вызначаецца частатой той ці іншай з'явы ў праблемнай вобласці. Гэта патрабаванне арыентуе «збіральнікаў» корпусаў тэкстаў на спецыялізацыю распрацаванага прадукта па ўзроўневым прынце: фанетычныя, марфалагічныя, сінтаксічныя, лексічныя, тэкставыя і інш. корпусы [1]. Што, дарэчы, не выключае агульную арыентацыю моўнага матэрыяла на вызначаныя стандарты.

Рэпрэзентатыўнасць любога корпуса, у прынце, не абмяжоўваецца ўласналінгвістычнымі параметрамі. Так, у кожным канкрэтным выпадку можа ўзнікнуць патрэба ўлічыць стылістычны, часавы, аўтарскі і іншыя элементы тэкставага масіву праблемнай вобласці. Экстралінгвістычныя характарыстыкі тэкстаў корпуса – своеасаблівая абалонка корпусных дадзеных. Менавіта ўлік такіх экстралінгвістычных фактараў на папярэднім этапе дазваляе дакладна фармуляваць задачы складальнікам корпусаў. Складаныя ўмовы функцыянавання сучаснай беларускай літаратурнай мовы проста абавязваюць улічваць экстралінгвістычны кантэкст рэпрэзентатыўнасці беларускага корпуса тэкстаў.

Дакладна спраектаваная рэпрэзентатыўнасць ператварае або не ператварае набор тэкстаў на машынным носьбіце ва ўнікальнае слоўнае адзінства – корпус тэкстаў. Гэта ўласцівасць корпуса настолькі важная, што часам гавораць пра рэпрэзентатыўнасць як пра вынік працэсу самаарганізацыі корпуса, разглядаемы, безумоўна, як метафара. Тады, па ідэі, на якой грунтуецца корпусная лінгвістыка, корпус тэкстаў адлюстроўвае аб'ектыўную карціну маўленчай дзейнасці незалежна ад жадання яго стваральніка [7; 8].

Такім чынам, для паспяховага развіцця беларускай корпуснай лінгвістыкі ёсць не толькі перашкоды ў выглядзе спазнення на пяцьдзсят год і ўнікальнага гістарычнага і функцыянальнага рознагалосся мовы, але і магчымасці ўлічыць усе памылкі і выкарыстаць усе дасягненні складальнікаў корпусаў іншых моў. Свядомы ўлік экстралінгвістычнага кантэкста сённяшняга развіцця беларускай мовы дазволіць паспяхова вырашаць пытанні рэпрэзентатыўнасці не толькі будучага нацыянальнага корпуса беларускай мовы, але і шматлікіх спецыялізаваных корпусаў, стварэнне якіх з дапамогай сучасных тэхналогій хутка можа стаць рэчаіснасцю лінгвістычных даследаванняў у Беларусі.

#### СПІС ЛІТАРАТУРЫ

- 1 Баранов, А. Н. Введение в прикладную лингвистику / А. Н. Баранов. – М. : Эдиториал УРСС, 2001. – 358 с.
- 2 Беларуская мова ў Інтэрнэт [Электронны рэсурс]. – Рэжым доступу: <http://mova.by.ru/>. – Час доступу: 10.08.2010.
- 3 Беларуская палічка [Электронны рэсурс]. – Рэжым доступу: [www.knihi.com](http://www.knihi.com). – Час доступу: 10.08.2010.

4 Звезда [Электронны рэсурс]. – Рэжым доступу: [www.zvyazda.minsk.by](http://www.zvyazda.minsk.by). – Час доступу: 10.08.2010.

5 Компьютерный фонд белорусского языка и перспективы создания белорусского лингвистического портала / Н. К. Рубашко, Г. П. Невмержицкая, И. В. Совпель // Слово и словарь. *Vocabulum et vocabularium*: сб. науч. тр. по лексикографии. – Гродно: ГрГУ, 2007. – С. 44–46.

6 Национальный корпус русского языка [Электронны рэсурс]. – Рэжым доступу: [www.ruscorpora.ru](http://www.ruscorpora.ru). – Час доступу: 10.08.2010.

7 Holmes-Higgin, P. *Assembling and Viewing a Corpus of Texts: Self-organisation, Logical Deduction and Spreading Activation as Metaphors* / P. Holmes-Higgin, K. Ahmad // *Euralex'96 Proceedings*. – Stockholm: Stockholm Univ. Press, 1996. – P. 47–91.

8 McEnery, T. *Corpus Linguistics* / T. McEnery, A. Wilson [Electronic resource]. – Edinburgh: Edinburgh Univ. Press, 1999. – Mode of access: <http://www.ling.lancs.ac.uk/staff/andrew/data.htm>. – Date of access: 10.08.2006.

И. Г. Гомонова

## ИССЛЕДОВАНИЕ ПАРЕМИОЛОГИЧЕСКИХ ЕДИНИЦ С ИСПОЛЬЗОВАНИЕМ МАТЕРИАЛОВ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА

Использование Национального корпуса русского языка (НКРЯ) [1] в повседневной практике в настоящее время стало нормой для большинства лингвистов. Тем не менее возможности корпуса в исследовании самых разнообразных лингвистических явлений по-прежнему удивляют и восхищают.

Богатейший материал предоставляет НКРЯ и исследователю-паремиологу, особенно в тех случаях, когда объектом исследования являются пословицы и поговорки, входящие в активный паремиологический фонд.

Паремию, как единицу, имеющую устойчивый характер, во многих случаях целесообразно искать в корпусе, используя функцию поиска точных форм. Так, например, осуществляя подобным образом поиск паремии *Язык мой – враг мой* <наперед ума лепечет / прежде ума (наперед ума) глаголет / прежде ума (моего) рыщет, беды ищет>, мы убедились в том, что это выражение бытует прежде всего в виде простого биноминативного предложения и именно так, как правило, и представлено в литературно-художественных и публицистических текстах разных авторов XIX, XX и XXI в.в. Только в двух контекстах (причем из разновременных и разностилевых произведений) из 22 случаев вхождения данной паремии в корпус она включает в себя поясняющую часть, зафиксированную в паремиологических словарях и имеющую, на наш взгляд, архаичный характер: – *Вы мне когда-то говорили, что он (Нил Андреич) племяннику обобрал, в казне воровал... – Помолчи, помолчи об этом, – торопливо отозвалась бабушка, – помни правило: «Язык мой – враг мой, прежде ума моего родился!»* (И.А.Гончаров. Обрыв (1869)); ...*осужденный, поехали, зво'нит,*

квар'тал, блага', ходата'йствовать... И несть числа этим «жучкам»! Вот когда уместно вспомнить пословицу: **«Язык мой – враг мой, прежде ума глаголит»**. Все эти «колорадские жучки» навели меня на мысль, как обеспечить сохранение чистого русского языка. (Евгений Весник. Дарю, что помню (1997)). Известная нам по данным паремиологических словарей вариативность местоимения-посессива (*мой / наш / твой*) также получила подтверждение в корпусе: *Совершается немало решительных действий, чтобы не дать языку заговорить: и наступают на язык, и укорачивают язык, и стараются не дать языку воли. Потому что считается, что язык наш – враг наш, что он страшнее пистолета, что он длинный, злой и вообще плохо подвешен. О добрых языках почему-то в народе умалчивают.* (Ф. Кривин. Хвост павлина (1981-1987)) – всего два вхождения в корпус данной паремии с посессивом *наш*; *С дядьками сдружился, врал им разную околесицу, и больше все-таки молчал, памятуя завет отца, у которого была любимая пословица: – Язык твой – враг твой, прежде ума твоего рыщет.* (В. А.Гиляровский. Мои скитания (1927)) – всего три вхождения в корпус паремии с посессивом *твой*.

По данным НКРЯ, анализируемая пословица часто входит в контекст в виде автономного предложения, которое у ряда авторов является эмоционально и интонационно окрашенным: *Отчего я молчу? я самого себя спрашиваю: виноват ли я перед вами? Точно, каюсь: виноват. Язык мой – враг мой. Но послушайте, Вера Николаевна...* (И.С. Тургенев. Где тонко, там и рвется (1851)); – *О, простите меня! Язык мой – враг мой! Ваша сестра очень глупа!* (А.С. Грин. Бегущая по волнам (1926)). В то же время более активно анализируемая паремия представлена в сопровождении «метаоператоров» – показателей интертекстуальности пословицы [см. об этом: 2, с. 47–51], подтверждающих мысль о том, что «говорящему необходимо ввести в речь пословицу, чтобы подчеркнуть, что его мнение основывается не на оторванных от человеческой жизни представлениях» [2, с. 49], «чтобы создать иллюзию стоящего за ним плотного и гомогенного социального пространства» [3, с. 312]. Это могут быть лексические и синтаксические (предикативные единицы) метаоператоры типа *считается, что...; истинно, что...; недаром говорят; старая истина; уместно вспомнить пословицу* и др.: *Так даже скажу: если и напередки тебе на этот счёт языком побаловать захочется, так ты вспомни пословицу: язык мой – враг мой, и, вспомнивши, плюнь.* (М. Е. Салтыков-Щедрин. Святочный рассказ (1858)); *Воистину верно говорят: «Язык мой – враг мой!»* (В. Доценко. Срок для Бешеного (1993)); *Подтвердилась старая истина «язык мой – враг мой»: видимо, кто-то из присутствовавших «стукнул» куда следует, что по тем временам было, увы, нормой советского общества.* (Б. Ефимов. Десять десятилетий (2000)).

Оформив в строке поиска точных форм запрос «чья бы корова», получаем 17 контекстов с паремией *Чья бы корова мычала, а твоя (бы) молчала*. Осуществляя поиск этой пословицы, нецелесообразно вводить

предикат, так как при этом теряются ее усеченные трансформации, например: – *Ваш брат метко его окрестил. Голиаф. В смысле – такой большой и такой ненужный. – Чья бы корова...* – не удержался Курчев. (В. Корнилов. Демобилизация (1969–1971)), а также случаи, в которых между подлежащим и сказуемым первой части предложения-паремии содержатся какие-либо компоненты, например вводные конструкции в роли метаоператоров: – *Уж чья бы корова мычала, как говорится, а твоя бы молчала, – в упор глядя в материны подслеповатые глаза, едко сказала Анна.* (Г. М. Марков. Строговы. Кн.1. (1936–1948)). Таким образом, паремия, имеющая структуру сложноподчиненного предложения, легко распознается в корпусе по метафорически переосмысленному посессивному комплексу. Впрочем, в виде сложного предложения данная пословица представлена в 10 случаях из 17, а в восьми из них, в свою очередь, с типичной для нее корреляцией посессивов *чей – твой*: – *И ты хорош, заводский варнак! – вскипел Анфим. – Сколько разов-то вывалил сам? Чья бы корова мычала, а твоя бы молчала... Сам ты баба!* (Д. Н. Мамин-Сибиряк. Хлеб (1895)); – *Я тебе тоже на полном серьезе: пошел к черту! Культурный нашелся. Уж чья бы корова мычала, а твоя бы молчала. Культурный – по чужим бабам шастать.* (Василий Шукшин. Живет такой парень (1960–1964)). В то же время в корпусе представлены и варианты анализируемой паремии, связанные с заполнением позиции посессива во второй части местоимениями *мой* и *его*: *Тут братья-сестры к Кате приступить стали, почему не сказала про Колывань-то. Только от Кати тоже немного добились. Сразу отрезала: – Чья бы корова мычала, моя бы молчала. Мало я вам сказывала, что Данило живой. А вы что?* (П. П. Бажов. Горный мастер (1939)); *Уж чья бы корова, как говорится, мычала, а его бы молчала.* (В. Медведев. Баранкин, будь человеком! (1957)). Несомненно, узнаваемой и свободно реализующей свой прагматический потенциал является данная паремия в тех случаях, когда она «сворачивается» до первой части. Авторы при этом, тем не менее, могут указать на недоговоренность: *Защитники напоминают, что рабочее движение всегда было частью политики, а насчет денег спецслужб – чья бы корова мычала...* (С. Мехов. Переметчик (2003) // «Совершенно секретно», 2003.04.03).

Материалы корпуса иллюстрируют тот факт, что в качестве интертекстемы анализируемая паремия выступает, как правило, в виде автономного предложения, не сопровождается метаоператорами, а если таковые и употребляются, то являются невыразительными в данном отношении вводными конструкциями типа *как говорится* или *что называется*: *А что до преследований за убеждения, в которых обвинял маккартистов Советский Союз, так это, что называется, чья бы корова мычала...* (В. Быков, О. Деркач. Книга века (2000)).

Помимо поиска точных форм, НКРЯ дает возможность поиска слова во всех или нескольких возможных формах, а также сегмента слова. Такой

вариант целесообразно использовать, если паремиологическая единица содержит компонент, являющийся основой для формирования ее образной составляющей и, соответственно, очень значимый для организации и интерпретации паремии. В данном случае в качестве таких компонентов рассматриваются посессивные комплексы *своя рубашка / рубаха* и *моя хата / изба*, узнаваемость которых не вызывает сомнений и подтверждается, в частности, возможностью вычленения из паремии сочетания с посессивом в качестве ее ключевого компонента: *С точки зрения социолога, совершенно очевидно, что многие человеческие качества и способности, жизненно необходимые для общества, проявляют себя наиболее сильно именно у «не-взрослого», «недосозревшего» человека... он легче воспринимает новое, у него меньше вздорных стереотипов в голове, он еще не проникся той опустошающей житейской мудростью, которая воплощена в поговорках о своей рубашке и хате с краю.* (Н. Наумова. О зрелости подлинной и мнимой // «Юность», 1971).

Поиск по посессивному сочетанию *своя рубашка / рубаха* (существительные можно оформлять как варианты или использовать функцию поиска сегмента *руба*\*) в основном корпусе дал результат в виде 110 вхождений, из которых 53 (!) представляли собой искомую паремию. Необходимость вручную отбирать нужный материал (что само по себе для человека заинтересованного является увлекательным процессом) компенсируется тем, что при таком варианте поиска мы не теряем самые разнообразные трансформации паремии, связанные с усечением и, наоборот, расширением ее компонентного состава, заменой компонентов, их инверсией и т. д., а также случаи встраивания внутрь паремиологической единицы метаоператоров. В анализируемом материале таких контекстов 24 из (напомним) 53. Например: *Посудите сами, время праздничное, подумают ещё, как следствие затеется, что его здесь и убили у вас; прикажите его, говорю вам, отвезти скорее, бог с ним, своя рубашка ближе...* (Д. В. Григорович. Бобль (1847)) – усечение состава; *«Станишники! Братцы!» – кричал вслед Степан, да где уж там – своя рубашка, а не чужая к телу липнет.* (М. А. Шолохов. Тихий Дон. Книга вторая (1928–1940)) – расширение состава и замена компонента; *Нет, не одни лишь судьбы страны решались здесь, судьбы народов и стран в конечном счете подвластны времени, но всякий раз на незримых весах взвешивалась, а то и колебалась судьба каждого. А своя рубашка, что ни говори, ближе к телу.* (Г. Я. Бакланов. Жизнь, подаренная дважды (1999)) – использование метаоператора.

Добавим, что подобные поразительные результаты дает и поиск по посессивному компоненту паремии *Моя хата (изба) с краю* <(я) ничего не знаю>. Выявленные благодаря материалам НКРЯ самые разнообразные особенности функционирования пословиц, ключевыми для которых являются образы *своей рубашки* и *моей хаты*, могут составить предмет отдельного исследования.

В заключение же скажем, что богатейший материал, предоставляемый исследователю паремиологических единиц корпусом, помогает в решении самых разнообразных задач – общих и частных. Так, например, фактически невозможно решить в абстрактной форме, без учета конкретных факторов реализации паремии, вопрос о том, какие именно функции являются типичными для тех или иных пословичных изречений.

Несомненный интерес вызывают также способы и средства включения паремий в контекст. Прочитав в этой связи Т. М. Николаеву, которая в 1995 г. высказала мысль о том, что «темой интересного специального исследования, которое еще ждет своего часа, может быть определение того, когда, в каком коммуникативном и/или в каком текстовом окружении могут появляться пословицы в речи современного человека» [3, с. 313]. Благодаря корпусу такой *час*, видимо, пробил.

Кроме того, в корпусе можно найти материал, позволяющий сделать выводы о типичных для того или иного автора способах «встраивания» паремий в тексты произведений, о представленности определенной паремиологической единицы в текстах разной стилистической, жанровой, хронологической, наконец, при желании, гендерной маркированности.

#### СПИСОК ЛИТЕРАТУРЫ

1 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: [www.ruscorpora.ru](http://www.ruscorpora.ru) – Дата доступа: 29.10.2010.

2 Даниленко, Л. Интертекстуальные реляции паремий / Л. Даниленко // Грани слова: Сб. науч. ст. к 65-летию проф. В.М. Мокиенко. – М. : ООО «Изд-во ЭЛПИС», 2005. – С. 47–51.

3 Николаева, Т. М. Обобщенное, конкретное и неопределенное в паремии / Т. М. Николаева // Малые жанры фольклора: Сб. ст. памяти Г. Л. Пермякова. – М., 1995. – С. 311–324.

А. В. Жулего

### **СОВРЕМЕННОЕ СОСТОЯНИЕ ЕВРОПЕЙСКИХ НАЦИОНАЛЬНЫХ ЯЗЫКОВЫХ КОРПУСОВ**

Для оценки языкового корпуса прибегают, прежде всего, к таким важнейшим критериям, как репрезентативность (представительность, representative) и сбалансированность (balanced) [1]. В. П. Захаров выделяет следующие наборы признаков-характеристик: тип данных, язык текстов, «параллельность», «литературность», специфичность, жанр, доступность, назначение, динамичность, разметка и характер разметки, объем текстов, хронологический аспект, «общность», структура [2].

Как известно, языковые корпуса (здесь и далее будем использовать московскую норму постановки слова во множественном числе, употребляемую на сайте Национального корпуса русского языка [ruscorpora.ru](http://ruscorpora.ru)) разрабатываются как для конкретных языков отдельно, так



существуют и параллельные корпуса одновременно нескольких языков. В нашей работе речь пойдет о состоянии *национальных* корпусов выбранных западноевропейских языков.

Национальный Корпус Польского Языка (Narodowy Korpus Języka Polskiego, адрес: <http://www.nkjp.pl>) – проект, финансируемый Министерством науки и высшего образования Республики Польша. Корпус создается четырьмя институтами, которые еще до возникновения проекта обладали собственными корпусами: Институт польского языка Польской Академии наук, Институт основ информатики ПАН, Лодзинский университет и научное издательство PWN. Совместным опытом и с использованием материалов всех партнеров-участников проекта удалось достаточно быстро опубликовать первую демонстрационную версию НКЖР. В настоящее время опубликована вторая демонстрационная версия, содержащая 450 млн. слов. В эту версию вошли уже тексты, собранные специально для НКЖР.

В плане представительности конечной целью является доведение количества слов до 1 миллиарда. Часть корпуса будет доступна для публичного использования по адресу [www.nkjp.pl](http://www.nkjp.pl), в то время как другая часть будет только для внутреннего пользования, т. е. непосредственно по месту пребывания четырех институтов, создающих НКЖР. Подкорпус объемом в 300 млн. слов будет сбалансирован относительно состава печатных текстов. Наибольшую долю в нем – вплоть до 50 % – будут составлять публицистические тексты. В оставшуюся половину входят: художественная литература (ок. 16 %), документальная проза (5,5 %), информационно-справочные тексты (5,5 %), научно-дидактические труды (2 %), небеллетристические и неклассифицированные тексты (1 %), иные письменные тексты (в т. ч. деловая документация, 3 %). Кроме того, будет добавлен незначительный процент скоротечных текстов, таких как объявления, реклама, тексты агитационно-пропагандистского характера, краткие инструкции, письма. Не стоит также обходить вниманием тексты из интернета. Около 7 % таких текстов войдет в сбалансированный корпус. Среди них будут как тексты, спонтанно создаваемые пользователями (форумы, чаты и т. п.), так и продуманные тексты со статических web-страниц. Вдобавок в НКЖР планируется включение порядка 10 % записанных устных текстов. Большая их часть будет записью медийных интервью и репортажей. Но будет также и небольшой процент записей разговорной речи. Для поиска в НКЖР используется корпусные менеджеры Poliqarp, а также менеджер, подготовленный проектом PELCRA в Лодзи ([www.1.uni.lodz.pl/pelcra](http://www.1.uni.lodz.pl/pelcra)).

Чешский национальный корпус (Český národní korpus, ČNK, адрес: [www.korpus.cz](http://www.korpus.cz)) – академический проект, имеющий целью создание объемного электронного корпуса главным образом письменного чешского языка. Автор проекта: Институт Чешского национального корпуса (Ústav Českého národního korpusu) на факультете философии и искусств Карлова университета в Праге. Проект стартовал с 1994 г. Структура корпуса: 1) подкорпус газет и журналов 1995–2007 гг. (2009 г., 700 млн.); 2) подкорпус газет и

журналов 1990–2004 гг. (2006 г., 300 млн.); 3) сбалансированный подкорпус текстов 2000–2004 гг. (2005 г., 100 млн.); 4) сбалансированный подкорпус текстов 1990–1994 гг. (2000 г., 100 млн.); 5) социолингвистически сбалансированный подкорпус разговорного чешского языка (2008 г., 1 млн.), диахронический подкорпус DIAKORP (1,6 млн.) и др. Для поиска используется корпусный менеджер Manatee/Bonito.

Словацкий национальный корпус (Slovenský národný korpus, SNK, адрес: [www.korpus.juls.savba.sk](http://www.korpus.juls.savba.sk)), создаваемый Академией наук Словакии, является базой современных словацких текстов различных функциональных стилей, морфологически размеченных с достаточно мощной поисковой системой. Объем всего корпуса на сегодняшний день (в версии prím-4.0) насчитывает свыше 520 млн. токенов. 65 % корпусной базы составляют публицистические тексты, 17 % – художественная литература, 16 % – специализированные тексты и 2 % – прочие. В качестве корпусного менеджера используется Bonito.

Хорватский национальный корпус (Hrvatski nacionalni korpus, HNK, адрес: [www.hnk.ffzg.hr](http://www.hnk.ffzg.hr)) является реализацией идеи создания национального корпуса хорватского языка, возникшей в стенах Института языкознания факультета философии Загребского университета еще в 1967 г. (когда был создан первый хорватский компьютерный корпус). Для поиска используется корпусный менеджер Manatee. Web-интерфейс корпусного менеджера на сайте HNK отсутствует, для поиска в корпусе используется свободно распространяемая программа Bonito. Сама же база данных располагается на сервере. В идеале, будущая структура Хорватского национального корпуса, объемом в 100 млн. словоупотреблений, будет выглядеть следующим образом: 74 % – информационные тексты (в т. ч. 37 % – газеты, 16 % – журналы, 21 % – публицистические книги, брошюры, корреспонденция), 23 % – художественная проза и 3 % – прочие тексты и тексты смешанных стилей и жанров.

Частью идеи создания национального корпуса *словенского* языка является проект “Nova Beseda” (адрес: [www.bos.zrc-sazu.si/a\\_about\\_si.html](http://www.bos.zrc-sazu.si/a_about_si.html)) – словенский языковой корпус, база которого насчитывает на данный момент 162 млн. словоупотреблений, 4158 текстов.

Институтом Болгарского языка совместно с Секцией компьютерной лингвистики и Секцией болгарской лексикологии и лексикографии создается Национальный корпус болгарского языка (адрес: [www.search.dcl.bas.bg](http://www.search.dcl.bas.bg)), в котором будут объединены отдельные электронные корпуса, разработанные в период с 2001 по 2009 гг. Национальный корпус насчитывает в настоящее время 320 млн. словоупотреблений, включенных в порядка 10 тыс. текстов. Материал корпуса иллюстрирует состояние болгарского литературного языка в период с 1945 г. по наши дни. Доступ к корпусу бесплатный, разделенный на два уровня: ограниченный доступ и полный доступ для зарегистрированных пользователей. Корпусный менеджер имеет довольно удобный интерфейс пользователя,

позволяющий конечному пользователю конкретизировать поиск наглядным образом, не прибегая к использованию специальных тегов.

Образцом национального языкового корпуса признан Британский национальный корпус (British National Corpus, BNC, адрес: [www.sara.natcorp.ox.ac.uk](http://www.sara.natcorp.ox.ac.uk)), общим объемом 100 млн. словоупотреблений. Корпус отражает состояние английского языка (его британского варианта) 2 половины XX в. Последняя на сегодняшний день редакция корпуса вышла в свет в 2007 г. 90 % текстовой базы корпуса являются образцы литературного письменного языка, 10 % – транскрипты разговорной речи. При поиске в корпусе используется корпусный менеджер XAIRA – поисковый инструмент, построенный на XML. Доступ к полной версии корпуса осуществляется на коммерческой основе, бесплатно осуществляется доступ к случайной выборке 50 результатов запроса.

Среди множества проектов *американских* корпусов единственным проектом с открытым публично доступом к базе данных является Корпус современного американского английского языка (Corpus of Contemporary American English, COCA, адрес: [www.americancorpus.org](http://www.americancorpus.org)), общим объемом свыше 410 млн. словоупотреблений из 160 тыс. текстов. Создателем корпуса является Марк Дейвис (Mark Davies), профессор корпусной лингвистики Университета Бригама Янга. Последнее обновление корпуса осуществлялось летом 2010 г. 85 млн. токенов представляют собой транскрипты устной речи, взятой из электронных масс-медиа; 81 млн. – художественная литература; 86 млн. – популярные журнальные издания; 81 млн. – газеты и 81 млн. – академические издания.

Коллекция COSMAS (адрес: [www.corpora.ids-mannheim.de/ccdb](http://www.corpora.ids-mannheim.de/ccdb)), поддерживаемая Институтом *немецкого* языка, представляет собой самое большое в мире электронное собрание современных немецкоязычных текстов. Хотя коллекция нацелена на охват максимально возможного объема текстов, а не на сбалансированность их состава, в ней представлены различные типы текстов: художественная проза, научная и научно-популярная литература, периодика и др. В части корпуса произведена автоматическая морфологическая разметка текстов. Имеется также подкорпус устной речи (разговорная речь, записи диалектной речи и др.). Ресурс включает созданную на основе корпуса базу данных по сочетаемости слов немецкого языка. Общий объем корпусной базы – около 2 млрд. слов. Доступ к полной версии корпуса имеют только сотрудники Института немецкого языка. Сторонним пользователям предоставляется возможность поиска по словоформе или лексеме по части корпуса, отражающей письменный язык (свыше 1,1 млрд. словоупотреблений), но выдаваемые при этом контексты не превышают 5 слов справа и слева от искомого и не выходят за границы предложения. В свободном доступе частично находится база данных по сочетаемости слов немецкого языка.

Корпус французских текстов Bibliothèque Universelle ([www.abu.cnam.fr](http://www.abu.cnam.fr)) включает произведения французской литературы, а

также некоторые переводные тексты на французском языке. Возможен просмотр целых текстов и поиска слова или группы слов по всему корпусу. Предполагается расширить возможности поиска: поиск заданной единицы можно будет осуществлять не только по всему массиву текстов, но и по выбранному пользователем его подмножеству. Доступ к корпусу свободный.

Представительный корпус испанского языка (Corpus de Referencia del Español Actual, CREA, адрес: [www.corpus.rae.es/creanet.html](http://www.corpus.rae.es/creanet.html)) отражает язык во всех его национальных вариантах за период с 1975 по 1999 гг. 50 % корпуса образуют европейские тексты, 50 % – латиноамериканские. 90 % объема корпуса занимают письменные тексты, 10 % приходится на долю записей устной речи. Корпус характеризуется использованием целых текстов. Тексты снабжены метаразметкой. Поиск в корпусе может осуществляться по словоформе, ее части или группе словоформ на базе всего корпуса или заданного пользователем подкорпуса. Подкорпус можно выбрать по различным параметрам: автор, название произведения, время создания текста, страна его происхождения, тип издания (книга, газета и т. п.), тип и тематика текста.

#### СПИСОК ЛИТЕРАТУРЫ

1 Andrzejczuk A. Narodowy Korpus Języka Polskiego: Prawdy, mity, potrzeby // Z materiałów konferencji: Conference on Translation, Interpreting and Comparative Legi-Linguistics 2009, organizowanej przez Instytut Językoznawstwa Uniwersytetu im. Adama Mickiewicza. – Poznań, w druku.

2 Захаров, В. П. Корпусная лингвистика: Учебно-методическое пособие / В. П. Захаров. – СПб : Издательство СПбГУ, 2005.

А. Ю. Крохмальник

### **ПОИСКОВЫЙ ЗАПРОС КАК СПОСОБ ВЫЯВЛЕНИЯ ЯЗЫКОВОЙ ПРЕДСТАВЛЕННОСТИ НЕВЕРБАЛЬНОЙ КОММУНИКАЦИИ В ХУДОЖЕСТВЕННЫХ ТЕКСТАХ**

Под корпусом понимается информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме. Национальный корпус представляет данный язык на определенном этапе (или этапах) его существования и во всем многообразии жанров, стилей, территориальных и социальных вариантов и т. п.

Национальный корпус создается лингвистами (специалистами по так называемой *корпусной лингвистике*, быстро развивающейся современной области языкознания) для научных исследований и обучения языку. Большинство крупных языков мира уже имеет свои национальные корпуса (различающиеся по полноте и уровню научной обработки текстов). Общеизвестным образцом является, в частности, Британский национальный корпус (BNC) [1]: на него ориентированы многие другие современные корпуса.

Национальный корпус [2] имеет две важные особенности. Во-первых, он характеризуется представительностью, или сбалансированным составом текстов. Это означает, что корпус содержит по возможности все типы письменных и устных текстов, представленные в данном языке (художественные разных жанров, публицистические, учебные, научные, деловые, разговорные, диалектные и т. п.), и что все эти тексты входят в корпус по возможности пропорционально их доле в языке соответствующего периода.

Во-вторых, корпус содержит особую дополнительную информацию о свойствах входящих в него текстов (так называемую разметку, или аннотацию). Разметка [3] – главная характеристика корпуса; она отличает корпус от простых коллекций (или «библиотек») текстов, в изобилии представленных в современном интернете, в том числе и на русском языке.

В ряде случаев поисковый запрос помогает установить модели вербализации определенных явлений внеязыковой действительности. Например, запрос, состоящий из слов *взгляд* (либо *глаза*) и *говорить* выявляет три варианта вербализации такого явления, как невербальная коммуникация при помощи глаз, что находит свое отражение в трех моделях: «взгляд/глаза + говорить», «взглядом/глазами + говорить» и «во взгляде/в глазах + говорить».

В модели «взгляд/глаза + говорить» взгляд выступает грамматическим субъектом действия и указывает на коммуникативную силу визуального воздействия: – *Надо развод? Я напишу ему. Я вижу, что я не могу так жить... Но я поеду с тобой в Москву. – Точно ты угрожаешь мне. Да я ничего так не желаю, как не разлучаться с тобой, – улыбаясь, сказал Вронский. Она видела этот взгляд и верно угадала его значение. «Если так, то это несчастье!» – говорил этот его взгляд. Это было минутное впечатление, но она никогда уже не забыла его* (Л. Н. Толстой. Анна Каренина).

В контексте реализуется как неконгруэнтность речи и визуального выражения, так и сугубо индивидуальная интерпретация взгляда (при помощи словосочетания *минутное впечатление*) и условность (при помощи союза *если*).

*Князь Василий вдруг пробурлил что-то и вышел. Пьеру показалось, что даже князь Василий был смущен. Вид смущенья этого старого светского человека тронул Пьера; он оглянулся на Элен – и она, казалось, была смущена и взглядом говорила: «что ж, вы сами виноваты». «Надо неизбежно перешагнуть, но не могу, я не могу», думал Пьер, и заговорил опять о постороннем, о Сергее Кузьмиче, спрашивая, в чем состоял этот анекдот, так как он его не расслышал* (Л. Н. Толстой. Война и мир).

В данном случае демонстрируется непричастность субъекта речи к действиям партнера по коммуникации. Следует отметить, что взгляд Пьера выражает также невербальный вопрос, поиск поддержки, что явилось следствием его состояния, переданного лексемой *тронуть* ‘вызвать

в ком-н. сочувствие, привести в умиление'. Визуальный ответ Элен отличается некоторой степенью условности ввиду наличия частицы *казалось*. Характеристика персонажа в контексте проявляется в слове с оценочной модальной характеристикой – *виноваты*.

Отдельно выделяются контексты, реализующие реакцию акционерного вида, т. е. при помощи конкретных действий:

*Удивительно посмотрел на меня Капитан Клюквин. Во взгляде его были и печаль, и досада, и лёгкое презрение ко мне. «Мне от вас ничего не надо», – говорил его взгляд. Да, Капитан Клюквин имел гордый характер, и я не стал с ним спорить, сдался, бросил семечко в кормушку* (Ю. Коваль. Капитан Клюквин).

Аналогичные модели определяются и при задании поискового запроса, состоящего из слов *жест* и *говорить*. Запрос, состоящий из слов *жест* и *говорить* выявляет три варианта вербализации такого явления, как невербальная коммуникация при помощи жестов, что находит свое отражение в трех моделях: «жест + говорить», «жестом(-ами) + говорить» и «в жесте(-ах) + говорить».

*Вместо ответа старый дрессировщик потрепал меня по шее. Я воспрянул: этот жест сказал мне больше, чем десятки похвал. По счастью, нам было легко подменить Багиру* (В. Запашный. Риск. Борьба. Любовь). Жест является ответной реакцией на вербальное обращение и выражает ободрение. Также передается эмоциональный ответ на него, что иллюстрируется при помощи глагола *воспрянул*. Как и в предыдущем случае, жест относится к скрытым репрезентивам, поскольку указывается на способность экспликации определенной информации при отсутствии содержательной стороны кинемы.

*Бим завилал хвостом, говоря этим жестом: «Ружье! Ружье! Знаю ружье!»* (Г. Троепольский. Белый Бим Черное Ухо). Очевидно, что собака выражает свое опознание предмета, а точнее – результат когнитивного процесса. Процесс опознания вербализуется при помощи повтора слова *ружье*, применительно к данному контексту можно говорить о приеме градации. Последний компонент градационного ряда – словосочетание *знаю ружье* – является наиболее сильным экспрессивным элементом, поскольку он не столько передает информацию, сколько интерпретирует положительные эмоции. Приведенные рассуждения дают основание говорить об экспрессивной целевой установке в рассматриваемом контексте.

*Вы умеете убеждать и руководить другими, заставляя выполнять работу в срок. Но иногда что-то в тоне, взгляде, жесте ваших подчиненных говорит: «Дай передохнуть!»* (В. Шахиджанян. 1001 вопрос про ЭТО (№№ 501–1001). В контексте воссоздается комплексное невербальное поведение персонажей, проявляемое также в жестовой форме. Директивный характер выражается при помощи конструкции *дай передохнуть*, аналогом формы повелительного наклонения. Жест является источником, способом

выражения определенных эмоций, выступающих в качестве неопределенного субъекта действия.

С другой стороны, в ряде случаев поисковый запрос состоит лишь из одного слова, которое, в свою очередь, может являться компонентом языковой модели. Например, лексема со значением невербальной коммуникации *жест* входит в состав языкового выражения целого ряда моделей, например: *Волосы у Мишеля стояли чубом, как у Караяна, носил он, как и Караян, водолазки. Он повернулся к Альберу и жестом показал: «Потрясающе!»* (С. Спивакова. Не всё). В данном случае в контексте репрезентируется модель «показать + жест» со значением изобразительной функции жеста.

*Мало быть от природы музыкальным человеком. Надо иметь и особую пластику, и умение передать жестом то, что написано в партитуре. Многие композиторы занимаются дирижированием.* (Р. Щедрин: «Даже самый очарованный странник всегда стремится в Россию»). В приведенном примере находит свое выражение модель «передать + жест» со значением референтной функции жеста.

*Ольга жестом заставила всех замолчать и сказала спокойно и холодно: – Ничего подобного. Я попросила бы вас, господа, похитить его и привести ко мне. – Ух ты! – воскликнул Гоша* (В. Белоусова. Второй выстрел). Модель «заставить + жест» со значением суггестивного проявления жестовой коммуникации проявляется в выделенном нами контексте.

*Я не собирался спорить. Ни к чему было и учить кого бы то ни было вежливости. Поэтому я только засмеялся и поднял обе руки жестом безоговорочно сдающегося. – Подождите, Нина Михайловна, подождите!* (А. Волос. Недвижимость). В контексте репрезентируется модель «жест + субстантив либо субстантивированный компонент» со значением передачи через жестикуляцию ситуационного состояния субъекта.

*Всё это он отбарабанил впоыхах как по писаному, а потом замолчал и вдруг положил руку мне на плечо жестом сожаления; но тут открылась дверь в дальнем конце коридора, и звонкий женский голос прокричал: «Косталенко! Косталенко!..»* (А. Волос. Недвижимость). В данном случае реализуется модель «жест + субстантив с абстрактным значением душевного состояния».

При поисковом запросе *в + взгляде* обнаруживается ряд моделей, в которых данная лексема выступает в качестве источника получения информации партнером по общению.

*Мне кажется, я попал на ленту Мёбиуса... Не в каком-нибудь там высоком смысле, а вот именно в том самом... в Петькином... Он умолк и посмотрел на меня. На этот раз в его взгляде явственно читалось желание от меня отделаться* (В. Белоусова. Второй выстрел).

*В её взгляде сквозила явная неприязнь и даже больше того – опасность. Между прочим, как я теперь понимаю, именно её отношение ко мне навело наиболее сообразительных из нас на некоторые размышления, а*

размышления плюс наблюдения позволили прийти к правильным выводам (В. Белоусова. Второй выстрел).

В приведенных двух контекстах функционирует модель «в/во взгляде + читаться + им. сущ. с абстрактным значением», выражающая возможность интерпретации взгляда собеседником на основании характера взгляда.

*Мне было бы лучше горевать в одиночку – горевать своё собственное горе; да, горевать своё и не делить его с ними – потому что в их взглядах читалось, будто я что-то у них отнимаю. Однако было бы глупо пытаться обнародовать свои чувства: я бы только смутил этих людей и расстроил течение поминок* (А. Волос. Недвижимость).

В данном контексте реализуется модель «в/во + взгляде + сквозить» со значением проявления определенных чувств, эмоций при помощи взгляда, что становится очевидным для партнера по общению.

Таким образом, поисковый запрос может состоять из двух компонентов, при помощи которых реализуются различные модели представления определенного действия, и одного компонента, позволяющего выявить включающие его модели.

#### СПИСОК ЛИТЕРАТУРЫ

1 Британский национальный корпус [Электронный ресурс]. – Режим доступа: <http://www.natcorp.ox.ac.uk/index.html> – Дата доступа: 12.10.2010.

2 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscopra.ru> – Дата доступа: 12.10.2010.

3 Язык разметки [Электронный ресурс]. – Режим доступа: <http://ru.wikipedia.org/wiki/> – Дата доступа: 11.10.2010.

С. Б. Кураш, О. И. Ревуцкий

#### ОТ МЕТАФОРЫ КОРПУСА – К КОРПУСАМ МЕТАФОР

Согласно М. Фасмеру, слово *корпус* отмечается в русском языке начиная с 1705 г., и пришло оно через польское *korpus* или немецкое *Korpus* из латинского *corpus* «тело, организм» [17]. В латинском языке, помимо названного и прочих, у слова *corpus* отмечается и такое значение, как «единое целое, общая масса, общий состав, совокупность» [11, с. 225].

Именно эти семы доминируют в пришедшем в русский язык из английского в 60-х гг. XX века лингвистическом термине *корпус* (а также терминологических словосочетаниях *языковой корпус*, *корпус текстов* и т. п.). Ср. в английском языке: *corpus* 1) formal a collection of writing, for example all the writings of one person; 2) linguistics a collection of written and spoken language stored on computer and used for language [22]. Подобные толкования слова *корпус* (как и сведения о корпусной лингвистике) имеются также в англоязычных лингвистических словарях [23, с. 84–93], свободной веб-энциклопедии «Википедия» [7], однако в словарях русского языка фиксируются не всегда, в том числе и лингвотерминологических.



Из словарей, фиксирующих данное значение слова *корпус*, можно указать на «Большой толковый словарь русского языка» под редакцией С. А. Кузнецова (1998), где приводится 10 значений полисеманта *корпус*, в числе которых на 3-ей позиции фиксируется и то, которое легло в основу лингвистического термина: «совокупность множества каких-л. однородных предметов, образующих целое; массив (*основной к. поэтических текстов*)» [3, с. 459]; на «Большой толковый словарь современного русского языка (онлайн-версия)», представляющий собой переработанное и существенно дополненное издание 4-томного «Толкового словаря русского языка» под редакцией Д. Н. Ушакова (Полное собрание, цельный свод каких-н. текстов (спец.). *Корпус стихотворений Пушкина*) и нек. др. [4]. В качестве лингвистического термина слово *корпус* представлено и в известном словаре О. С. Ахмановой [1, с. 209].

«Под названием *лингвистический, или языковой, корпус текстов* понимается большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [6, с. 3].

В применении к современному лингвотерминологическому аппарату понятие корпуса, таким образом, представляет собой пространственную метафору, в целом весьма характерную для языковой репрезентации многих основополагающих концептов сознания, в том числе и концепта «язык». В числе аналогичных пространственных метафорических терминов можно отметить такие, как *семантическое поле, смысловая зона, смысловое пространство, центр (периферия) семантического поля* и даже *семантическая деревня* (В. З. Демьянков), каждый из которых «опирается на представление о языке как о пространстве» [14].

Метафора корпуса, как видно, достаточно новая для современного русскоязычного лингвотерминологического аппарата, но при этом утвердилась в нём весьма прочно и породила соответствующее концептуальное поле, в котором можно отметить такие терминологические номинации, как *размер, разметка (tagging, annotation), метка (tag), тэггеры (taggers), парсеры (parsers), корпусный менеджер (конкордансер), конкорданси* др.

Примечательна своим названием и содержанием размещённая в Интернете лекция известного российского лингвиста В. А. Плунгяна, прочитанная им в 2009 году, на тему «Почему современная лингвистика должна быть лингвистикой корпусов». Автор в достаточно живой и аргументированной манере обосновывает следующий тезис: «...человек должен знать слово «корпус». Знать, что это такое, для чего это нужно, почему это важно. ... Люди, настроенные романтически, даже могли бы сказать, что в лингвистике произошла корпусная революция. После появления корпусов эта наука стала совсем другая. (...) корпусная лингвистика в узком смысле – это всего лишь наука о том, как создавать корпуса и как ими пользоваться, но она претендует на гораздо большее, на роль новой идеологии науки о

языке» [13]. В качестве одного из риторических приёмов аргументации В. А. Плунгян использует метафору, причём весьма характерно, что у него в связи с этим опять-таки возникает метафора пространства: «теперь для овладения языком человеку нужны не две, а три вещи: словарь, грамматика и корпус текстов данного языка. Потому что и словарь, и грамматика, в общем-то, бесполезны вне этого живого пространства, где язык, собственно, и функционирует» [13].

Некоторые специалисты вообще говорят о самой идее корпуса и её основных составляющих (самоорганизации, способности к развитию) как о метафорической по своему генезису [21].

Можно, таким образом, прогнозировать дальнейшее развитие лингвистической метафоры корпуса, учитывая перспективность компьютерной лингвистики в целом.

Однако не только корпусная лингвистика обязана языковому метафорикону термином для своего обозначения, но и метафорология как междисциплинарная наука о метафоре в самом широком её понимании в качестве одной из важнейших перспектив развития, без сомнения, имеет смыкание с идеологией корпусности.

С. А. Хахалова в связи с этим отмечает, что «метафорический корпус любого языка представляет собой благодатный материал для исследования когнитивной реальности эгоцентрической категории метафоричности, поскольку в нем концептуализированы не только знания о собственно человеческой наивной картине мира и все типы отношений субъекта к ее фрагментам, но как бы запрограммировано участие этих языковых сущностей вместе с их употреблением в межпоколенной трансляции эталонов и стереотипов национальной культуры» [18, с. 177]. Другой известный специалист в области лингвометафорологии О. Н. Лагута в качестве первоочередных задач данной науки называет «проведение системных исследований индивидуальных языковых метафорик: на материале как живых языков с богатой книжно-письменной традицией и развитыми функционально-стилевыми подсистемами, так и младописьменных языков с диффузными функционально-стилистическими характеристиками, а также на материале мертвых языков, зафиксированных на широком пространстве текстов, с целью последующего выявления корпуса наиболее актуальных моделей метафоризации и их сопоставительного описания» [9, с. 105]. Этот же автор с сожалением отмечает тот факт, что «из-за дискретности научных сфер лингвисты практически не используют достижений психологической и психотерапевтической метафорологии (корпус психотерапевтических текстов, используемых регулярно в клинической практике, почти не имеет лингвистического, особенно когнитивистского, описания). Однако только нейрологические исследования могут пролить свет истины на связь языка – естественного коммуникативного кода, коммуникативной (с включенной в нее речевой) деятельности и ментального языка с работой высшей нервной системы. В свою очередь, нейрологические исследования

результатов структурно-семантического описания метафор и их когнитивной интерпретации могут раскрыть ряд загадок полушарной асимметрии, а исследование результатов сопоставительных работ по изучению языковых метафорик может способствовать решению проблемы естественной и языковой категоризации» [9, с. 106].

Таким образом, идеология корпусности способствует дальнейшей интеграции различных наук и направлений, их методолого-методического аппарата, что в итоге может привести к получению принципиально нового знания о, казалось бы, давно и неплохо изученных объектах действительности. Можно констатировать, что к настоящему моменту в этом направлении сделан ряд шагов и получены теоретически и практически значимые, весьма интересные научные результаты и продукты. Однако здесь приходится отмечать тот факт, что в реализации корпусного подхода к метафоре первенство принадлежит западноевропейским учёным. Отметим, в частности, изданный не так давно сборник «Корпусные исследования метафоры и метонимии» [20], отразивший ряд важных научных результатов в интересующей нас сфере.

Что касается русскоязычных корпусно-лингвометафорологических исследований, то, не претендуя на полноту обзора, мы назовём лишь некоторые из них.

Из тематических групп метафор в первую очередь следует отметить корпусное описание политических метафор, поскольку именно политическая лингвометафорология, взявшая за основу когнитивный подход Дж. Лакоффа – М. Джонсона [10], явилась лидером в выстраивании союзнических отношений с корпусной лингвистикой. Здесь прежде всего следует отметить работы А. Н. Баранова, Э. В. Будаева, Н. А. Санцевич, А. П. Чудинова и др. [2; 5; 16; 19].

В соответствии с задачами исследования иных тематических групп метафор были составлены и другие соответствующие метафорические корпуса – юридических, педагогических, медицинских, психотерапевтических метафор, метафор информационных технологий, эмоциональных состояний человека и др. Кроме того, методика корпусного исследования активно применяется при изучении индивидуально-авторских метафориконов.

Особо следует отметить возможности использования национальных языковых корпусов, в частности, Национального корпуса русского языка (НКРЯ) [12] в решении задач лингвометафорологических исследований. К числу основных мы бы отнесли такие: семантическая разметка (определяет возможность поиска употребления слова в составе иного таксономического класса, т. е. вне прямой референции); синтаксическая разметка, дающая возможность поиска по заданным синтаксическим моделям (а метафора возникает именно в контексте, причём некоторые модели контекстов, можно сказать, специализируются на рождении метафорической семантики, например, генитивные словосочетания, словосочетания с творительным падежом и др.); разметка по параметрам

текстов (позволит подключить так называемые жанрово-тематические ожидания: например, больше метафорических контекстов со словами типа *звезда, огонь, луч* и т. п. можно ожидать от поиска в подкорпусе поэтических текстов, в сравнении с другими подкорпусами). Заметим при этом, что непосредственно маркер «метафора» (как и какие-либо иные тропеические маркеры) отсутствует в метаразметке текстов НКРЯ.

Наконец, НКРЯ позволяет проследить эволюцию самого концепта «метафора» в разных сферах его функционирования, дополнить его содержание [8].

Вместе с тем по отношению к применению корпусного подхода к исследованию метафоры учёными признаётся ряд ограничений: «Во-первых, эффективность анализа корпусов метафорических контекстов явно снижается, когда исследователь сталкивается с таким феноменом, как инновативная метафора (*innovative metaphor*). Во-вторых, проводимый анализ крупных корпусов по своей сути представляет интроспективную деятельность человека познающего, направленную на рассмотрение определенного объекта исследования «извне», а не «изнутри», что значительно снижает ценность теоретических обобщений в инференциальной составляющей полученного научного продукта. В-третьих, возникает проблема репрезентативно-иллюстративного представления практического материала в рамках того или иного экспериментального проекта» [15]. Однако заметим: данные ограничения касаются не столько идеологии корпусности в принципе, сколько говорят о молодости и связанных с ней «болезнях роста» данного подхода; с развитием корпусной лингвистики, с появлением и совершенствованием новых корпусов таких опасений, на наш взгляд, должно становиться всё меньше и меньше.

#### СПИСОК ЛИТЕРАТУРЫ

1 Ахманова, О. С. Словарь лингвистических терминов / О. С. Ахманова. – М. : Сов. энциклопедия, 1966. – 606 с.

2 Баранов, А. Н. Проблема репрезентативности корпуса данных (на примере политической метафоры) / А. Н. Баранов // Труды Международного семинара Диалог '2001 по компьютерной лингвистике и ее приложениям. – Аксаково, 2001 [Электронный ресурс]. – Режим доступа: <http://www.dialog-21.ru/Archive/2001/Baranov.zip>. – Дата доступа: 14.07.2010.

3 Большой толковый словарь русского языка / сост. и гл. ред. С. А. Кузнецов. – СПб. : Норинт, 1998. – 1536 с.

4 Большой толковый словарь современного русского языка (онлайн-версия) [Электронный ресурс]. – Режим доступа: <http://www.classes.ru/all-russian/russian-dictionary-Ushakov.htm>. – Дата доступа: 17.07.2010.

5 Будаев, Э. В. Когнитивно-дискурсивный анализ метафоры в политической коммуникации / Э. В. Будаев, А. П. Чудинов // Политическая лингвистика. – 2008. – № 3 (26). – [Электронный ресурс]. – Режим доступа: [http://journals.uspu.ru/i/inst/ling/ling26/ling\\_6\(26\)2008\\_budaev2.pdf](http://journals.uspu.ru/i/inst/ling/ling26/ling_6(26)2008_budaev2.pdf). – Дата доступа: 04.06.2010.

6 Захаров, В. П. Корпусная лингвистика: уч.-метод. пособ / В. П. Захаров. – СПб. : СПбГУ, 2005. – 48 с.

- 7 Корпусная лингвистика: статья // Википедия [Электронный ресурс]. – Режим доступа: [http://ru.wikipedia.org/wiki/Текстовый корпус](http://ru.wikipedia.org/wiki/Текстовый_корпус). – Дата доступа: 25.10.2010.
- 8 Кураш, С. Б. Нетерминологические употребления слова «метафора» в современном русском языке (по данным Интернет-источников) / С. Б. Кураш // Русский язык: система и функционирование (к 70-летию филол. ф-та): материалы IV междунар. науч. конф.: в 2 ч. Ч.1. – Минск: РИВШ, 2009. – С. 184–187.
- 9 Лагута, О. Н. Лингвометафорология: основные подходы / О. Н. Лагута. – Новосибирск: Изд-во Новосибирского гос. ун-та, 2003. – 114 с.
- 10 Лакофф, Дж. Метафоры, которыми мы живём / Дж. Лакофф, М. Джонсон. – М.: Едиториал УРСС, 2004. – 256 с.
- 11 Латинско-русский словарь / авт.-сост. К. А. Тананушко. – Минск: Харвест, 2008. – 1040 с.
- 12 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru>. – Дата доступа: 12.07.2010.
- 13 Плуныян, В. А. Почему современная лингвистика должна быть лингвистикой корпусов / В. А. Плуныян [Электронный ресурс]. – Режим доступа: <http://www.polit.ru/lectures/2009/10/23/corpus.html>. – Дата доступа: 17.07.2010.
- 14 Полиниченко, Д. Ю. Естественный язык как лингвокультурный семиотический концепт (на материале русского и английского языков): дис. ... канд. филол. наук: 10.02.20 / Д.Ю. Полиниченко. – Краснодар, 2004. – 179 л.
- 15 Самигуллина, А.С. Теория метафоры в современной англистике: принципы, подходы, перспективы / А.С. Самигуллина // Вестник Челябинского государственного университета. Сер. Филология. Искусствоведение. Выпуск 19. – 2008. – № 9. – С. 115–120.
- 16 Санцевич, Н. А. Моделирование вариативности языковой картины мира на основе двуязычного корпуса публицистических текстов: метафоры и семантические оппозиции: дис. ... канд. филол. наук: 10.02.21 / Н. А. Санцевич. – М., 2003. – 268 с.
- 17 Фасмер, М. Этимологический словарь русского языка; под ред. Б. А. Ларина / М. Фасмер. – Изд. 2-е, стереотип: в 4 т. – М.: Прогресс, 1986. – Тома I–IV. [Электронный ресурс]. – Режим доступа: <http://dic.academic.ru/dic.nsf/vasmer/41622>. – Дата доступа: 15.07.2010.
- 18 Хахалова, С. А. Когнитивная реальность эгоцентрической категории метафоричности / С. А. Хахалова // Языковая онтология семантически малых и объемных форм: Вестник ИГЛУ. Сер. Лингвистика. – Иркутск, 2000. – Вып. 1. – С. 173–182.
- 19 Чудинов, А. П. Корпусное исследование политической метафоры / А. П. Чудинов // Актуальные проблемы лингвистики и терминоведения. – Екатеринбург, 2007. – С. 219–224.
- 20 Corpus-based approaches to metaphor and metonymy / Edited by A. Stefanowitsch, St.Th. Gries. – Berlin. – New York: Mouton de Gruyter, 2006. – 319 p.
- 21 Holmes-Higgin, P. Assembling and Viewing a Corpus of Texts: Self-organisation, Logical Deduction and Spreading Activation as Metaphors / P. Holmes-Higgin, K. Ahmad // EURALEX '96: Proc. I-II, Part I. – Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden. – Göteborg: Göteborg University. – P. 109–120.
- 22 Macmillan English Dictionary [Электронный ресурс]. – Режим доступа: <http://www.macmillandictionary.ru/dictionary/british/corpus>. – Дата доступа: 15.07.2010.
- 23 The Linguistics Encyclopedia / Edited by Kirsten Malmkjær. – Second Edition. – London. – New York: Routledge is an imprint of the Taylor & Francis Group, 2002. – 643 p.

## ИССЛЕДОВАНИЕ СЕМАНТИКИ НАЦИОНАЛЬНО МАРКИРОВАННЫХ ЕДИНИЦ ПРИ ПОМОЩИ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ ПИСЬМЕННЫХ И УСТНЫХ ТЕКСТОВ

Лингвистические корпуса письменных и устных текстов все чаще рассматриваются как основной источник материала для различного рода языковедческих исследований. Пользователей корпусов, как правило, интересует не содержание конкретных текстов, а их метатекстовая информация и примеры (контексты) употребления тех или иных языковых элементов и конструкций.

Известно, что корпуса предназначены в первую очередь для обеспечения научных исследований лексики и грамматики языка, а также процессов языковых изменений, происходящих в языке на протяжении сравнительно небольших периодов [1]. На основе полученных материалов можно осуществлять категоризацию языкового материала и создавать проекции языка на различные подязыки. С помощью размеченного корпуса можно решить ряд прикладных задач в области лексикографии, преподавания и самостоятельного изучения языков, машинного перевода и информационного поиска.

Лингвистические корпуса предоставляют исследователям реальные контексты, точные статистические данные на больших объемах текстов, иллюстрируют сочетаемость языковых единиц. Наиболее ценными среди предоставляемых национальными мегакорпусами данных являются: типичные и нетипичные употребления и характерные сочетания слов, коллокации (collocations), контексты, частоты (абсолютные и относительные), частоты по коллокациям, статистика по жанрам, стилям, авторам и т. д. [2, с. 22].

Лингвистический корпус является бесценным источником как общезыковой, так и национально-маркированной информации. Следовательно, его можно использовать при исследовании семантики лингвокультурем. При этом, как показывает практика, корпуса одного языка не всегда достаточно. Эффект можно получить лишь работая параллельно как минимум с двумя корпусами. Именно тогда полученные в результате обработки корпусов данные позволят судить о лингвокультурных отличиях [2, с. 23].

Для проведения исследования семантики целевых единиц «обязанность – duty» были отобраны материалы электронных мегаресурсов Национальный корпус русского языка (НКРЯ) и Британский национальный корпус (BNC). BNC является статичным одноязычным корпусом, представляющим синхронный срез современного британского английского языка. Наиболее ранние произведения, включенные в корпус, датируются 1964 годом. Для большей объективности проводимого исследования в НКРЯ был задан подкорпус с временным ограничением «1950–2009 годы».

4630 документов НКРЯ предоставил пользователю на запрос «обязанность». Для работы с соответствующей единицей в английском языке BNC выдал 7862 документа на запрос «duty».

Анализ русскоязычных контекстов, содержащих концепт «обязанность» позволил выделить «гражданско-правовую обязанность» и «административную обязанность»: *У вас даже тени понятия нет о таких вещах, как гражданский долг, о том, что, в конце концов, просто первая обязанность мужчины – отстаивать интересы своей родины и своего народа* (М. Бутов. Свобода // «Новый Мир», 1999).

Поведение гражданина, как члена социума, предполагает соответствие ряду административных требований. Так, на основании контекстов была выделена категория «профессиональный долг и функции»: *Ведь это его обязанность как капитана оказывать членам своего экипажа моральную поддержку.* (Василий Аксенов. Звездный билет // «Юность», № 6, 7, 1961). При этом, обязанность, определяемая спецификой профессии, зачастую характеризуется как «нудная», «тягостная»: *Весь день Незванный и Старшов знакомились с обстановкой, распределяли участки и сектора обстрелов, а закончив эту привычную нудную обязанность, стали ждать, строго-настрого приказав не ввязываться в активный бой.* (Б. Васильев. Дом, который построил Дед (1990–2000)). Контекстов с положительной коннотацией найдено не было.

Если в категории «гражданско-правовая обязанность» частотными стали контексты с сочетаниями «право и обязанность», то в примерах, отнесенных к категории «профессиональный долг и функции», лингвокультурема «обязанность» встречается в сочетании с лингвокультуремой «долг»: *Ведь он писатель, а писателю нужно писать книги, в конце концов, это его долг и обязанность.* (В. Войнович. Иванькиада, или рассказ о вселении писателя Войновича в новую квартиру (1976)).

Помимо категории «профессиональный долг и требования», следует выделить еще одну категорию – «задание/поручение». *Митрофанову вменялась единственная обязанность. После двенадцати нужно было выключить какой-то рубильник.* (С. Довлатов. Заповедник (1983)).

В ходе анализа языкового материала был отмечен контекст с синтагмой «мужская обязанность по дому»: *С тех пор единственную мужскую обязанность по дому – выносить ведро – Иван целиком предоставил ей самой.* (О. Славникова. Стрекоза, увеличенная до размеров собаки (1995–1999)). Следует отметить, что примера, содержащего непосредственное сочетание «женская обязанность (по дому)» среди выдач НКРЯ найдено не было. Однако контекст, передающий аналогичный смысл был отобран: *Ее обязанность приготовить нам скудный обед, щи с кусочком сала и картошку на молоке, иногда постирать белье (иногда стирает женщина), иногда вымыть полы* (М. М. Пришвин. Дневники (1922)). Можно сделать вывод, что обязанность женщины по дому воспринимается россиянами как нечто должное, а вот

любое действие мужчины по хозяйству даже с помощью лингвистических средств отмечается как явление редкое, непривычное.

Для выделенной категории «моральный долг» характерны примеры, содержащие синтагмы «обязанность хранить таинство», «святая /святая обязанность», «обязанность молить Бога», ярко отражающие духовность русского народа: *Но у каждой личности есть не только право, а даже обязанность хранить таинство внутри себя.* (А. Слаповский. Любовь по-нашему // «Знамя», 2003). Синтагмы «обязанность товарища», «обязанность перед предками» отражают трепетное отношение россиян к дружбе, родственным узам: *Ведь это моя обязанность как вашего старшего товарища...* (В. Скворцов. Сингапурский квартет (2001)).

В ряде контекстов, помимо рассматриваемой единицы «обязанность», встречаются также несколько лингвокультурологических единиц, не менее важных в ценностной картине мира россиян: любовь, ненависть, долг, жизнь. Примерами таких контекстов являются следующие отрывки: *Для таких дней не стоило бы и просыпаться, но что поделаешь, просыпаться приходилось, приходилось каждый раз признавать за собою обязанность жить.* (С. Залыгин. Бабе Ане – сто лет // «Новый Мир», 2000).

Некоторые контексты позволяют выделить категорию «обязанность, связанная с семьей»: *Отчасти Софья Андреевна признавала за собой материнскую обязанность предупредить дочь, каким именно неприятным способом осуществляется брак.* (О. Славникова. Стрекоза, увеличенная до размеров собаки (1995–1999)).

При работе с материалами BNC для лингвокультуремы «duty» был выделен ряд семантических категорий, как схожих с категориями русского языка, так и отличных. Как и в русском языке, в английском представляется возможным выделение двух категорий: «общественные требования» и «внутренний долг».

Административные требования обуславливают наличие семантической подкатегории «гражданско-правовая обязанность». *H81 813 Citizens thus had a legal duty to reveal felonies known to them.*

В сфере административного урегулирования обязанностей англичан можно выделить две подкатегории: «профессиональный долг и требования» и «задания/поручения». *HXV 15 There has been a breach of duty by an engine driver in failing to sound his whistle.* Довольно многочисленными являются контексты с оборотом «to be under a duty»: *FBK 158 The police are under a duty to report all crimes to the prosecutor who is charged with the duty of conducting criminal investigations.*

К вышеупомянутым контекстам, одновременно включающим не одну, а несколько лингвокультурем нашего исследования, можно добавить следующий отрывок: *JXH 336 It is the responsibility and duty of investors and businesses to take a long-term view and they increasingly do so.* Следует отметить, что для русского языка такое сочетание («ответственность и обязанность») является нетипичным.



Для подкатегории «задания и поручения» характерными являются следующие синтагматические отношения: «be on duty», «duty period», «duty scheme», «night duty»: *CE9 639 She could not sleep and, remembering that Ronald Travis would be on duty, decided to wait until her parents had gone to bed, then slip outside and go to the signal box to see her lover.*

Не все категории совпадают по своей семантике с выделенными категориями для исследуемой лингвокультуры в русском языке. Так, в английском языке широкую контекстуальную базу для формы единственного числа имеет категория «военные обязанности»: *K3H 132 Paul, 23, wrote and proposed to 19-year-old Gillian Spencer at Christmas while he was on a six-month tour of duty with the UN.* Была также выделена принципиально отличная от русского языка семантическая подкатегория «церковная служба». *K97 7320 Because good works were required of a puritan as a duty of thanksgiving, the appeal of Bacon's arguments to the intellectual leaders of the English Revolution can be readily understood.* Ряд контекстов послужили основанием для выделения семантической категории «система налогообложения»: *K5C 393 There was no duty free...*

Как и россиянами, англичанами зачастую движут внутренние побуждения, обусловленные моральными и нравственными ценностями народа в целом. Категория «моральный долг» включает ряд контекстов, где целевая лингвокультура была отмечена в сочетаниях «sense of duty», «call of duty», «virtues of duty», «duty to ourselves»: *HJ3 6042 Furious host officials last night accused Charles and Diana of putting their personal problems before their sense of duty.*

Отдельную категорию – «обязанность, связанную с семьей», также актуальную и для русского языка, представляют контексты, содержащие сочетания «duty to obey», «husband duty», «family duty»: *B3F 139 While she worked she told Sara how fortunate she was to be going to Portugal and to have been found a husband and how, when she was married, it would be her duty to obey him and remain utterly faithful to him, taking the Blessed Virgin Mary as her model and guide in all things.*

Анализируя эмотивный аспект полученных материалов, можно сделать вывод, что обе нации зачастую рассматривают понятие «обязанности» как нечто обременительное, усложняющее жизнь. Среди русскоязычных материалов был найден лишь один контекст с синтагмой «почетная обязанность». Для англичан и россиян обязанность/обязанности зачастую представляются не долгом, а лишь заданием, некой функцией, обусловленной работой, профессией.

#### СПИСОК ЛИТЕРАТУРЫ

1 Национальный корпус русского языка [Электронный ресурс]. – 2003 – 2010. – Режим доступа: <http://ruscorpora.ru/>. – Дата доступа: 10.09.2010.

2 Рычкова, Л. В., Национальные мегакорпусы в изучении лингвокультур / Л. В. Рычкова, Е. Р. Кныш // Материалы IV Международной научной конференции 3–5 декабря 2009 г. Часть первая. – Минск, 2009. – С. 21–25.

## КОРПУС АГИОГРАФИЧЕСКИХ ТЕКСТОВ СКАТ: РАЗМЕТКА ЭЛЕМЕНТОВ СОДЕРЖАНИЯ

СКАТ – электронный корпус агиографических церковнославянских текстов XVI–XVII вв., созданный на кафедре математической лингвистики СПбГУ. Работа над ним началась еще в 70-х годах XX века с создания картотеки житий святых русской церкви. Тогда же, в конце 70-х, началась работа по вводу текстов житий в компьютер. На данный момент корпус содержит более 50 житий общим объемом свыше 500 тысяч словоупотреблений.

Параллельно формированию базы данных было начато изучение грамматики и словообразования конкретных текстов. К концу 1996 г. были опубликованы обобщающие работы, в которых содержится систематическое описание именного склонения, глагольного спряжения и именного словообразования памятников русской агиографической литературы XVI вв. В 1980–90-х гг. было опубликовано несколько полученных на ЭВМ словоуказателей к введенным текстам. (Подробнее об истории проекта см. [1])

С конца 90-х годов XX в. на кафедре математической лингвистики СПбГУ реализуется широкомасштабный проект по изданию серии текстов «Памятники русской агиографической литературы». В каждой книге серии текст одного или нескольких связанных между собой житий сопровождается указателем словоформ, текстологическим очерком, историческими сведениями о святых и основанных ими обителях. На данный момент вышло в свет уже 9 выпусков серии, содержащих 17 житий вологодских святых, десятый выпуск находится в печати.

Тексты опубликованных житий размещаются на сайте проекта [2] и доступны для бесплатного просмотра и скачивания. Помимо текстов на сайте представлена информация о самом проекте: история его создания, имена участников, принципы представления рукописного текста, особенности представления словоформ в словоуказателе.

Тексты житий представлены в формате PDF и формате XML. Разработка XML-формата для представления текстов корпуса началась на кафедре в 2004 году. Разметка осуществляется на основе международных норм оформления электронных изданий текста, в частности Text Encoding Initiative (TEI) [3].

На сайте проекта также представлен электронный словоуказатель, который позволяет осуществлять поиск словоформ по всему корпусу текстов, при этом можно искать как словоформу целиком, так и ее фрагмент. Для найденных словоформ указывается адрес: рукопись, номер листа и строки. Этот адрес соответствует формальному членению текстов на листы – колонки – строки – слова, которое отражено в структуре XML-файлов. Ведется также работа по разметке морфологических и синтаксических характеристик текстов.

Помимо формального деления на страницы и строки, можно провести также деление на смысловые части. Прежде всего, следует отметить, что во многих рукописях автор сам разбивает текст на главы, выделяет их заголовками или буквицами. Разметив эти заголовки, можно создать для каждого жития оглавление по авторским разделам.

Однако для эффективного поиска по текстам этого деления недостаточно. В одной авторской главе может оказаться несколько различных сюжетов. Например, в авторской главе «О пострижении святого» рассказывается не только о пострижении, но и о переходах святого из монастыря в монастырь, о желании удалиться в пустынь, о создании собственной обители и т. п. Поэтому возникла необходимость ввести более дробное смысловое деление.

Это возможно, поскольку композиция житий подчиняется определенным канонам. При написании жития автор ориентировался на уже существующие тексты. Готовый текст использовался как шаблон, который заполнялся новыми образами, персонажами, элементами сюжета [4]. За счет этого приема каноническая схема жития заимствуется и переходит из текста в текст. Поэтому можно вывести общую сюжетную схему, характерную для большинства из житий.

По результатам исследования текстов корпуса СКАТ была выделена общая схема развития сюжета, состоящая из характерных для всех исследуемых житий мотивов [5]. В соответствии с этой схемой производится XML-разметка текстов житий. Такая разметка позволяет составить для каждого текста оглавление, показывающее, какие элементы общей схемы представлены в конкретном тексте. В оглавлении также предусмотрено указание точного адреса начала раздела, что позволит осуществлять быстрый переход к нужному фрагменту текста.

Помимо повторяющихся сюжетных мотивов в житийных текстах присутствуют устойчивые литературные клише, которые заимствуются из текста в текст, используются для описания определенных ситуаций и характеризуются «закрепленностью за определенным элементом композиции» [6, с. 62].

В рамках проекта СКАТ ведется работа по выявлению и разметке таких клише. Это позволит не только составить глоссарий, но и указать точные адреса вхождений таких устойчивых формул. Для каждого клише в глоссарии указывается название жития, в котором оно встретилось, и точный адрес: указание страницы, столбца и строки. Кроме того, указывается тот элемент сюжетной схемы, в котором употребляется клише. Это дает возможность быстро проследить, где именно в тексте встречается данная фраза и для описания какой ситуации она употребляется.

Например, если проследить употребление фразы «в сон тонок сведен бысть», можно увидеть, что она характерна для двух ситуаций. Первая это рассказ о знамении, полученном святым при выборе места для монастыря.

Вторая – рассказ о чуде, когда, наоборот, образ святого являлся для кого-то знаменем.

Разметка клише удобна для ряда задач. Во-первых, полученный глоссарий дает возможность быстрого поиска с целью дальнейшего анализа текста. Используемый вариант разметки позволяет без труда пополнять глоссарий и вводить в него новые фразы. Во-вторых, выявление таких клише, характерных для определенных компонентов сюжета, может в дальнейшем упростить выявление сюжетной схемы. Их можно использовать как формальный маркер при разметке: клише сразу указывает на то, к какому элементу схемы относится данный фрагмент текста. Кроме того, существуют обороты, обозначающие переход от одного раздела к другому. Например, фраза «мы же на подлежащее возвратимся», по которой можно автоматически размечать завершение одного компонента сюжета и начало другого. Использование таких фраз-маркеров поможет автоматизировать разметку содержательной структуры новых текстов.

#### СПИСОК ЛИТЕРАТУРЫ

1 Азарова, И. В. Разметка текстовых фрагментов в корпусе агиографических текстов СКАТ / И. В. Азарова, Е. Л. Алексеева, Л. А. Захарова // Труды международной конференции «Корпусная лингвистика – 2006». 10–14 октября 2006 г. – СПб., 2006. – С. 16–24.

2 Санкт-Петербургский корпус агиографических текстов СКАТ [Электронный ресурс]. – Режим доступа: <http://project.phil.spb.ru/skat/> – Дата доступа: 20.10.2010.

3 Международный консорциум по выработке норм электронной разметки текстов // Режим доступа: <http://www.tei-c.org/> – Дата доступа: 18.10.2010.

4 Панченко, О. В., Поэтика уподоблений (к вопросу о «типологическом» методе в древнерусской агиографии, эпидейктике, гимнографии) // ТОДЛР. – СПб., 2003. – Т. 54. – С. 491–534.

5 Рогозина, Е. А. XML-разметка содержательной структуры агиографических текстов в корпусе СКАТ // Материалы XXXVIII Международной филологической конференции (Санкт-Петербург, 16–21 марта 2009 г.). Секция прикладной и математической лингвистики / Под ред. Т. Г. Скребцовой. – СПб., 2009. – С. 48–54.

6 Руди, Т. Р. Топика русских житий (вопросы топологии) // Русская агиография. Исследования. Публикации. Полемика / Отв. ред. С. А. Семячко. – СПб., 2005. – С. 59–101.

Л. В. Рычкова, С. Н. Киеня

### СОЦИОКУЛЬТУРНАЯ ЗНАЧИМОСТЬ НАЦИОНАЛЬНЫХ КОРПУСОВ

Понятие социокультурной значимости лингвистических корпусов многоаспектно. Во-первых, сам национальный (в иной терминологии – «генеральный» [1, с. 316]) корпус в силу свойства репрезентативности представляет собой сокровищницу языка, во-вторых, материал, включенный в корпус, является культурным достоянием общества, наконец,

социокультурная значимость лингвистических корпусов может рассматриваться в прагматическом аспекте, т. е. с точки зрения использования корпусов различными категориями пользователей.

Национальный корпус – это представительный электронный лингвистический ресурс, отражающий состояние определенного языка в целом. Важной особенностью является то, что создатели национальных корпусов стремятся к представлению генеральной совокупности языкового материала во всем его стилевом и жанровом разнообразии. Национальные корпуса могут включать подкорпусы, представляющие различные формы существования языка: литературный язык, территориальные и социальные варианты языка, языки для специальных целей. Следовательно, помимо литературной, кодифицированной, формы языка, они могут отражать все разнообразие проявлений и богатство общенародного языка. Эта особенность национальных корпусов позволяет рассматривать их как хранилища разнообразного языкового материала, своеобразный «срез» состояния общенародного языка на протяжении определенного периода времени.

Язык является самоорганизующейся и динамической системой, внутри которой могут происходить различного рода изменения. Корпусы, в зависимости от хронологического аспекта, могут быть диахроническими / динамическими / «открытыми» (отражать эволюцию языка) и синхроническими / статичными / закрытыми (отражать состояние языка в определенный период времени). Динамические корпуса предоставляют возможность проследить изменения, происходящие в языке. Если корпус является постоянно пополняемым, то он будет отражать и новые появившиеся в языке единицы или конструкции (окказионализмы, неологизмы, заимствования разной степени освоенности), а также изменения в системе языковых номинаций в определенный период времени, и, опосредованно, состояние общества на данном этапе, языковую и ценностную картины мира. Так, в зависимости от изменения системы общественных ценностей можно наблюдать сдвиг семантической просодии слова, как вербальной номинации определенного понятия, от сугубо отрицательной к положительной (как в случае, например, с *агрессивный, корпорация, корпоративный*) либо полную смену семантической просодии на противоположную (*плюрализм* и др.).

Исходя из предложенного Л. Ельмслевым членения – «схема – норма – узус – акт» – рассмотрим, что именно находит и может находить отражение в корпусах. Схема, по Л. Ельмслеву, – это чистая форма, «схема чистых отношений». Лишенная материальной манифестации, схема не может быть положена в основу корпуса, хотя может найти отражение в системе лингвистической разметки. Норма материальна, социально обусловлена, но она не зависит от манифестации (проявления) в речи. Корпусы, отражающие «норму», могут формироваться только искусственно, специально конструироваться в различных социокультурных целях, например, образовательных. Узус – это совокупность навыков, принятых в данном (со)обществе и обусловленных определенными манифестациями. Корпусы

неизбежно отражают узус, или существующие узуальные практики, стереотипы, шаблоны. Это отмечается многими исследователями и связано с проблемой репрезентативности (сбалансированности) корпуса. Так, например, «основной корпус» Национального корпуса русского языка отражает сложившиеся в разное время узуальные практики, а «учебный корпус» приведен в соответствие с действующими нормами и, соответственно, отражает только кодифицированную форму языка. Металингвистическая разметка позволяет отбирать языковой материал, порожденный в разные годы, и, используя основной корпус Национального корпуса русского языка, эффективно осуществлять процедуру сравнения для определения узуальных тенденций, в том числе, в целях коррекции нормы. Данный аспект весьма важен при определении социокультурной значимости корпусов.

Корпусы принято рассматривать как специальным образом организованный материал, но, как представляется, они могут и должны рассматриваться также как фактор культуры, отражение определенного периода социокультурного развития общества. С этой точки зрения, возрастает значимость «интеллектуальной разметки». Благодаря наличию в корпусе разметки, извлечение языкового материала, обладающего нужными характеристиками, становится легким. Материал, включенный в корпус, представляет собой культурное наследие народа. Тексты разных временных периодов, входящие в корпус, содержат слова-концепты с определенной культурной семантикой. Культурные модели обладают высокой воспроизводимостью, они обнаруживают свое присутствие также и в стершихся метафорах, которые, сохраняя внутреннюю форму, сохраняют также культурную модель и способность к периодическому «оживанию». Поскольку одним из имманентных свойств языка является устойчивость, связанная с воспроизводимостью и с общественной памятью (прецедентные тексты транслируются из поколения в поколение), очень важно, чтобы культурные конструкты были зафиксированы. А собрание текстов, многие из которых отражают эти культурные конструкты, в одном месте (в данном случае в корпусе), приводит к сохранению и трансляции культурного наследия и может способствовать закреплению и формированию определенных ценностей [2, с. 13].

Культурные конструкты находят отражение в рамках стереотипов. Стереотипы могут формироваться на основе практически любого признака: возраста, пола, расы, профессии и т. д. Но, независимо от вида стереотипов, все они закрепляются средствами языка. Так, например, современному обществу свойственно стереотипное мышление относительно гендерных отличий. Гендерные стереотипы – внутренние установки в отношении места мужчин и женщин в обществе, их функций и социальных задач. Что касается собственно лингвистического аспекта гендерных стереотипов, то его часто связывают с ожиданием определенного типа речевого поведения, различного для мужчин и для женщин, и с реакцией на нарушения этого ожидания. Гендерная стереотипизация проявляется, таким образом, в ожидании того, что речь женщин отличается от речи мужчин.

Так, метаразметка Национального корпуса русского языка включает признак пола автора, а возможности Британского национального корпуса неоднократно использовались для выявления лексических предпочтений представителей разных полов. Например, выявлено, что в аналогичных речевых ситуациях для передачи значения ‘хорошо’, ‘замечательно’ британцы предпочитают сказать *fine*, а британки, как правило, используют *nice* и очень редко *fine*.

Корпус как многоцелевой ресурс рассчитан не только на определенный узкий круг лиц (чаще говорят о лингвистах), но и на использование самыми различными категориями пользователей для решения широкого круга разнообразных задач. Например, специалисты по общественным наукам (историки, социологи) могут изучать свои объекты через язык, используя такие параметры текстов, как период, автор или жанр. Корпусы также используются для разработки и настройки различных автоматизированных систем (машинный перевод, распознавание речи, информационный поиск) [3, с.11]. С нашей точки зрения, круг пользователей корпусов не ограничен и будет с годами расширяться.

Доступность корпусов (например, Национального корпуса русского языка) широкому кругу пользователей за пределами лингвистического сообщества существенно влияет на изменение социокультурной значимости не только корпусной лингвистики, но и языкознания в целом.

#### СПИСОК ЛИТЕРАТУРЫ

1 Рычкова, Л. В. Корпус как совокупность многослойных полнотекстовых баз данных / Л. В. Рычкова // Труды международной конференции «Корпусная лингвистика – 2004». – СПб. : Изд-во С.-Петерб. ун-та, 2004. – С. 316–323.

2 Брагина, Н. Г. Социокультурные конструкты в языке / Н. Г. Брагина. – М. : Гос. ИРЯ им. А.С. Пушкина, 2005. – 360 с.

3 Захаров, В. П. Корпусная лингвистика: уч. пособие / В. П. Захаров. – СПб., 2005. – 48 с.

И. В. Серикова

### **О МУЛЬТИМЕДИЙНОЙ ПРЕДСТАВЛЕННОСТИ ПУБЛИЦИСТИЧЕСКОГО ПОДКОРПУСА В НАЦИОНАЛЬНОМ КОРПУСЕ РУССКОГО ЯЗЫКА**

Национальный корпус русского языка (НКРЯ, или Корпус), два этапа работы над которым уже завершены (2003–2005 г.г., 2006–2008 г.г.) и который был открыт для свободного доступа в Интернете в 2004 году, продолжает развиваться, а разработчики Корпуса ставят новые задачи для дальнейшего усовершенствования данного проекта.

Таковыми задачами, как указывает Е. В. Рахилина, является отражение в НКРЯ всех «хронологических срезов языка, всех его региональных и социальных вариантов, а вариативность по этим параметрам в русском языке

достаточна велика. Полноценное отражение такой вариативности – первая задача. В некоторых случаях варианты превращаются почти что в отдельные подъязыки, для которых нужно строить свои подкорпуса со своей специально настроенной на них системой разметки. Игнорировать такие слои русского языка никак нельзя: чем сложнее они устроены, тем больше их значимость для системы в целом. Значит, это вторая задача» [1, с. 9].

Решение этих задач привело к реализации в рамках НКРЯ Корпуса звучащей речи, который получил название «Мультимедийный русский корпус» (МуРКо). Особенность данного корпуса состоит в том, что он не имеет аналогов в мировой практике и создаётся в качестве национального мультимедийного корпуса [2, с. 187].

Ядро МуРКо должен составить «кинематографический материал, накопленный к настоящему моменту в ходе пополнения подкорпуса устной речи и акцентологического подкорпуса в рамках НКРЯ» [2, с. 175–176] и «материал теле- и видеозаписей» [2, с.185]. Данные мультимедийные тексты (ММТ) – это особые «тексты, спаянные со зрительным и звуковым рядом, представляющие не письменную и не электронную речь, поскольку имеют в основном устную форму бытования, но, с другой стороны, это и не устная речь per se: основные характеристики последней – спонтанность и невозпроизводимость, а мультимедийная речь в достаточной степени подготовлена (иногда уровень подготовленности мультимедиа существенно выше, чем уровень подготовленности письменной речи) и, безусловно, не просто воспроизводима, а рассчитана на воспроизводимость» [3, с. 233–234].

Если говорить об объеме МуРКо, то в конечном итоге он будет иметь не менее 3 млн. словоупотреблений. По количественной характеристике МуРКо превосходит мультимедийные корпуса в Европе и США, «на создание которых накладываются определённые экстралингвистические ограничения, что ведёт к разумному ограничению объёмов корпусов, ни один из которых не достигает и одного миллиона словоупотреблений (обычно же мультимедийный корпус в 100 000 словоупотреблений считается большим)» [2, с. 184]. На наш взгляд, объём МуРКо может увеличиться за счёт материала, который находится в личных картотеках лингвистов. Пополнение корпуса – «задача крайне насущная» [1, с. 12], отмечает Е. В. Рахилина и далее выражает пожелание, чтобы «материалы, используемые в диссертациях, монографиях, хрестоматиях, предоставлялись в Корпус, где бы они обрабатывались и становились общедоступными при поиске» [1, с. 12].

Возможно, для разработчиков будет представлять интерес тот пласт «мультимедиа», который используется журналистами в периодике, не только российской, но и русскоязычной прессе Беларуси. Собранный ручным способом методом сплошной выборки, материал охватывает 2001–2005 г.г., частично 2006–2008 г.г. Мультимедийные тексты, извлечённые из общественно-политических газет, могут расширить (хоть и незначительно) объём не



только МуРКо, но и Публицистического подкорпуса, Поэтического подкорпуса и Корпуса названий.

ММТ как семиотическое образование используется журналистами в прессе в качестве претекста. Диапазон вовлечения в процесс текстопостроения и творческой интерпретации претекста весьма широк и разнообразен: использование в новых текстах и в заголовочных комплексах (ЗК). Нас интересует активное употребление ММТ журналистами в ЗК газетных текстов различных по жанровой характеристике русскоязычной прессы Беларуси. В ЗК (собственно заголовок, подзаголовок, Лид, афишка) отмечены названия кино- и телефильмов, мультфильмов, реплики героев кино- и телефильмов, мультфильмов, названия песен, строки из песен, названия опер, названия телепередач, телереклама. Данные ММТ в позиции заголовка газетного текста имеют следующие особенности. Необходимо учитывать, что журналистский текст – это прежде всего письменный текст. Таким образом, ММТ в позиции ЗК демонстрируют тесную связь мультимедийной речи с письменной. ММТ как семиотическое образование могут быть наделены соответствующей логикой и экспрессией, – например, иконический текст обладает динамизмом в процессе знакопостроения (театр, кино). Вербальный элемент – заголовочный комплекс – служит связующим звеном с иконическим текстом, например, с кинотекстом.

ММТ в позиции ЗК встречаются чаще всего в трансформированном виде, т.е. изменяется их структура и семантика или только семантика, а также служат моделями для построения целой серии фраз. Например, кинопаремия *Вооружён и очень опасен* представлена следующими преобразованиями: *Вооружён. А опасен?* (Республика, 11.10.02); *Вооружены и очень опасны* (Республика, 30.11.02); *Вооружён и очень глуп* (СБ (Советская Беларусь), 21, 06.02); *Вооружён и носит... женский парик* (СБ, 20.09.01); *Возбуждены и очень опасны* (СБ, 21.08.03); *Осведомлён – значит защищён* (СБ, 20.06.02); *Предупреждён значит вооружён* (СБ, 23.07.02); *Разоружены и уже неопасны* (СБ, 7.03.02); *Пойман и уже не опасен* (СБ, 25.09.03). Трансформируются рекламные слоганы *Имидж – ничто, жажда – всё: Имидж – ничто* (СБ, 25.02.03); *Имидж – всё* (СБ, 13.02.02); *Жажда – ничто, имидж – всё* (Совершенно секретно, № 1, 2003), *Жажда – всё* (СБ, 17.12.02); *Разведка – ничто, её имидж – всё* (СБ, 10.08.02).

Различаются ММТ в хронологическом плане и по степени цитирования. Необходимо также учитывать, что не только исходные ММТ (например, мультимедийная реклама), но и их трансформы имеют ограниченный срок жизни.

Таким образом, если проводить отбор материала, то встанет вопрос, в какой корпус – МуРКо или Публицистический – будут входить ММТ, активно используемые в прессе? Анализ текстов, находящихся в личной картотеке, показал, что в позиции ЗК активно изменяются определённые ММТ. Возможно, если при отборе материала учитывать разнообразную

тематику и направленность изданий, жанровый состав текстов, то определится ядро регулярно трансформирующихся «газетных» ММТ.

Техническая подготовка любого корпуса включает метаразметку текстов. По-видимому, «газетные» ММТ должны иметь указание не только на автора мультимедийного текста, который «может быть неизвестен, либо коллективен, либо далеко не единственный» [3, с. 240], но и на автора газетного текста, который известен всегда. Можно также точно узнать его пол и возраст, а «данные пункты при метаразмётке ММТ, скорее всего, останутся не заполненными» [3, с. 240]. Если говорить о сфере функционирования, то ММТ в позиции ЗК – это будут «газетные» мультимедийные тексты, а ММТ-источники в зависимости от того, с каким текстом они будут ассоциироваться в тезаурусе реципиента, будут обозначаться как «мультимедийная реклама» или «художественные мультимедиа» [3, с. 240].

Так же будет определяться и жанр текста. ММТ-заголовки являются частью текста, и жанр будет определён именно для того или иного газетного текста. Мультимедийная реклама и художественные мультимедиа «имеют собственное подразделение на жанры» [3, с. 241]. Для того чтобы понять, с какой целью автор изменил ММТ, «для заголовков необходимо предусмотреть функцию расширения контекстов, по-видимому, рассматривать первую фразу текста» [3, с. 247]. Однако для ММТ-заголовков необходимо приводить весь текст, т.к. именно он чаще всего даёт ключ к расшифровке трансформированной фразы.

#### СПИСОК ЛИТЕРАТУРЫ

1 Рахилина, Е. В. Корпус как творческий проект / Е. В. Рахилина // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. – СПб. : Нестор-История, 2009. – С. 7–26.

2 Гришина, Е. А. Мультимедийный русский корпус (МУРКО): проблемы аннотации // Е. А. Гришина // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. – СПб. : Нестор-История, 2009. – С.175–214.

3 Гришина, Е. А. Два новых проекта для национального корпуса: мультимедийный подкорпус и подкорпус названий / Е. А. Гришина // Национальный корпус русского языка: 2003–2005. – С. 233–250.

Н. И. Шабулдаева

### **О ФУНКЦИОНИРОВАНИИ КОНСТРУКЦИИ ХЛЕБ-СОЛЬ В РУССКОМ ЯЗЫКЕ**

**(на материале Национального корпуса русского языка)**

Национальный корпус русского языка (далее – НКРЯ) даёт возможность исследовать языковую единицу многогранно; при этом возможна количественная характеристика употреблений по авторам и временным отрезкам [1]. Предметом нашего рассмотрения является конструкция

*хлеб-соль* 'угощение (первоначально – обязательный атрибут ритуала встречи в виде хлеба и соли)' [2, с. 1444].

В текстах разного характера, отображённых в НКРЯ, обнаружено следующее количество употреблений сложносоставного наименования *хлеб-соль*: в прозаических произведениях – 525 (вхождений – 988), в диалектных текстах – 10 (вхождений – 15), поэтических текстах – 71 (вхождений – 85), в записях устных бесед – 18 (вхождений – 25), в газетных статьях – 192 (вхождений – 209).

На основании данных НКРЯ можно проследить, какие писатели, сколько раз, в каких своих произведениях и в каком году употребляли данную конструкцию. Мы приводим данные по использованию её от семи раз. Так, В. Я. Шишков в «Емельяне Пугачеве» использует её 28 раз; Мельников-Печерский в «В лесах» (Книга первая (1871–1874)) – 20; А. К. Толстой в «Князе Серебряном» (1842–1862) – 18; Н. М. Карамзин в «Истории государства Российского» (Том 12 (1824–1826)) – 16; Н. И. Костомаров в «Русской истории в жизнеописаниях ее главнейших деятелей» – 16 раз; М. Шолохов в «Тихом Доне» – 14; И. С. Никитин в стихах (1854) – 11; Ф. М. Достоевский в «Преступлении и наказании» (1866) – 8. А. Ф. Членов («Как Алешка жил на Севере» (1978)) прибегает к ней 11 раз. М. Н. Загоскин в своих описаниях довольно часто обращается к данной конструкции («Москва и москвичи» (1842–1850) – 7; «Рославлев, или Русские в 1812 году» (1830) – 7; «Юрий Милославский, или русские в 1612 году» (1829) – 10). П. И. Мельников-Печерский («На горах». Книга первая (1875–1881)) и И. М. Долгоруков («Повесть о рождении моем, происхождении и всей моей жизни, писанная мной самим и начатая в Москве, 1788-го года в августе месяце, на 25-ом году моей жизни» / Часть 4 / 1799–1806 (1788–1822)) употребили по 7 раз указанную конструкцию.

Степень спаянности слов *хлеб, соль* в выявленных текстах разная: 1) как неделимое целое – (*Родители благословляли жениха иконой, хлебом-солью...* (Свадьба тюменских старожилов // «Народное творчество», № 5, 2004), оформленные раздельно, но представляют собой одно целое – *В последний-то раз отплыл от берега.., остановился и как... крикнет по-своему. Дескать: «Спасибо за хлеб, за соль!..».* (Д. Н. Мамин-Сибиряк. Приемьш (1893)); *...благодарю за хлеб за соль* (Д. В. Григорович. Бобыль (1847)); *Но как прожить, скажем, учителю начальных классов, если его зарплата едва-едва хватает на хлеб и на соль.* (Ю. Азаров. Подозреваемый (2002)); *Благодарим за хлеб соль.* (А. Белянин. Свирепый ландграф (1999)); 2) как независимые друг от друга слова, объединённые общим контекстом (*Объясняю, что такое хлеб, соль, почему все свадебные пироги круглые.* (Роща из... теста // «Народное творчество», № 5, 2003); 3) одно из них синтаксически зависит от другого (*Отец брал булку хлеба с солью в солонке и три раза проводил ею над головами молодых...* (Свадьба тюменских старожилов // «Народное творчество», № 5, 2004)).

С использованием НКРЯ возможно исследование семантики даже при не снятии омонимии. В прозаических и поэтических текстах обнаружено употребление *хлеб-соль* в переносном значении:... *Сокольников*, для которой составление и разоблачение заговоров было, как говорила она сама, *хлебом и солью жизни*, то есть источником заработка и острейшим наслаждением, тут же увидела за матросским визитом подвох. (Д. Быков. Орфография (2002)); *Тому из двух в одном со мной делившим Беседы простодушной хлеб и соль...* (В. И. Иванов. «Тому из двух в одном со мной делившим...» (1923)); *Хлеб с солью дружен... Так подчас Болтаем мы иль просто мелем...* (Д. Д. Минаев. Хлеб и соль: «Хлеб с солью дружен... Так подчас...» (1880)).

В ходе исследования выявлено несколько случаев инверсионного употребления данной конструкции: *В первую ночь в доме оставался только стол, поставленный в центре и накрытый белой скатертью, с солью и хлебом.* (С. А. Еремеева. Лекции по русскому искусству (2000)); *Коли уж мы с вами стали состязаться в знании русских пословиц, то на этот счет есть такая: без соли без хлеба худая беседа* (С. Романов. Парламент (2000)); *Пошла с дарами соли-хлеба Просить: «Останься, наш отец!»...* (Н. А. Некрасов. Притча о «Киселе»: «Жил-был за тридевять земель...» (1865.08.21)).

Семантика данного сложносоставного слова довольно разнообразная. Оно используется в качестве приветствия: – *Добрый вечер, хозяева! Хлеб-соль вам.* (Б. Можяев. Живой (1964–1965)). *Как мне теперь быть, и ума не приложу: сразу ли назад пятки или «хлеб да соль» говорить...* (Е. Пермяк. Бабушкины кружева (1955–1965)). *Стрельбицкий узнал его именно по кожуху с обрезанными рукавами. – Хлеб-соль! – сказал Стрельбицкий, опуская на землю сваю.* (В. П. Катаев. Порт (1957)). – *Здорово, дядя Степан. Хлеб да соль. – Милости просим.* (А. Веселый. Россия, кровью умытая (1924–1932)). *С любезнейшей улыбкой, на чистом русском языке, приветствовал он нас. – Хлеб-соль!* (С. Д. Мстиславский. Крыша мира (1905)). Конструкция является символом гостеприимства: *Рассказывали нам, что воры в Сенне до того уважают хлеб, соль, как символ гостеприимства, что если, забравшись ночью в дома для грабежа, им случится наступить на один из этих предметов; то они бросают свое предприятие, уходят из этого дома и никогда уже более не забираются туда для воровства.* (Е. И. Чириков. Путевой журнал (1849–1852)). Рассматриваемая единица обозначает дружеские отношения: *В Крайске он водил хлеб-соль и с семейством Куприяновых и с Иннокентием Филатычем.* (В. Я. Шишков. Угрюм-река. Ч. 1–4 (1913–1932)); *Дитя мое, вам бы не мешало помнить, что водить хлеб-соль с честными дураками, подобными вашей сестре и мне, гораздо безопаснее...* (П. И. Ковалевский. Император Петр III (1900–1910)); *Итак, мы с Домной Платоновной всё водили хлеб-соль и дружбу.* (Н. С. Лесков. Воительница (1866)); *Я там с Парфеном старостою давно хлеб-соль*

*вожу* (М. Н. Загоскин. Искуситель (1838); *Ведет хлеб-соль с поляками* (М. Н. Загоскин. Юрий Милославский, или русские в 1612 году (1829)). Используется как пожелания приятного аппетита: ...*поздоровалась заведующая губоно. – Садитесь. Хлеб да соль. – Спасибо! – ответил хор голосов.* (Г. Г. Белых, Л. Пантелеев. Республика ШКИД (1926)); *Вафля, здравствуй! Приятного вам аппетита, хлеб да соль.* (А. П. Чехов. Леший (1888)); *Потом он шагнул к костру, улыбнулся еще светлее и сказал: «Хлеб да соль, братцы!»* (А. П. Чехов. Степь (1888)); ...*прибавил он, усаживаясь подле пообедавших свое кушанье: «Хлеб да соль!* (Ф. М. Достоевский. Записки из мертвого дома (1862)); *Я, право, и не заметил, что вы уже обедаете. Ну, хлеб да соль!* (В.Т.Нарежный. Два Ивана, или Страсть к тяжбам (1825)). Обозначает пищу вообще: *Вот мы им тюремного хлеба-соли принесли.* (С. Маршак. Умные вещи (1945–1964)); *Снегурочка просит Леля спеть, она готова заплатить пастуху за песню хлебом-солью...* (Т. Поташинская. Краткое содержание оперы «Снегурочка» (1964)); *Родне, прибывшей издалеча, Повсюду радостная встреча, И восклицанья, и хлеб-соль.* (С. М. Волконский. О декабристах (1921)); *На протяжении восьми лет вместе ели депутатские хлеб-соль и много что еще», – говорит спикер.* (Итоговый выпуск (вечерний) – 27.04.05 18:25 – Екатеринбург // Новый регион 2, 2005.04.28).

Богатейшие ресурсы Национального корпуса русского языка позволяют изучать грамматические категории данной языковой единицы. Сложносоставное слово *хлеб-соль* имеет особенности в употреблении грамматических категорий. Оно обычно употребляется в форме единственного числа. Однако выявлены случаи использования его и во множественном числе: ...*традиционные хлеб-соль. Были и здесь торжественные митинги, теплые приветствия...* (Вл. Янелис. Горячий север // Техника – молодежи, 1982); *хлеб-соль готовые, кров теплый, поел казак да на бок.* (Н. С. Лесков. Старые годы в селе Плодомасове (1862)); *На протяжении восьми лет вместе ели депутатские хлеб-соль и много что еще», – говорит спикер* (Итоговый выпуск – 27.04.05 18:25 – Екатеринбург // Новый регион 2, 2005.04.28).

Категория рода в данном случае обнаруживает колебания. Мужской род с препозицией определения выявлен в 8 случаях: *И кончает письмо: «Я бы хотел, чтобы вы приехали к нам на Алтай. Отведать нашего хлеба и соли.* (В. Некрасов. Три встречи (1959–1967)); *Не оставьте напередки, братики, на наш хлеб-соль, на нашу кашу!* (В. А. Гиляровский. Мои скитания (1927)); *Граф Артур, прошу вас сделать мне честь, откусать моего хлеба-соли.* (Н. Г. Гарин-Михайловский. Студенты (1895)); *А в закончении письма стояла просьба: «не пренебречь нашим хлебом-солью...».* (Н. С. Лесков. Заячий ремиз (1894)); *Нашего хлеба-соли отведай, Мосей.* (Д. Н. Мамин-Сибиряк. Три конца (1890)); *Не побрезгуй хоть ты-то нашим хлебом-солью.* (А. Ф. Писемский. Взбаламученное море (1863)); *Прошу*

милости к **нашему хлебу-соли**. (Н. С. Лесков. Старые годы в селе Плодомасове (1862)). В качестве определения выступают притяжательные местоимения *наш, мой*. Не выявлено ни одного случая употребления *хлеб-соль* в мужском роде с определением в постпозиции.

В том случае, когда данное сложносоставное слово является женского рода, то определения значительно разнообразнее и многочисленнее. Употребление определения в препозиции (9 примеров): *...женихи Наталье не нравились, и пропала даром **сватовская хлеб-соль***. (М. А. Шолохов. Тихий Дон. Книга первая (1928–1940)); *За **твою хлеб-соль** оставлю ей, надолго хватит*. (П. П. Бажов. Малахитовая шкатулка (1938)); *Ну, вот еще что скажешь, **какая тебе хлеб-соль**, я всегда тебе рад, – говорил Василий...*(П. С. Романов. Звезды (1927)); *...посещал имения своей супруги и принимал **подносимую** ему местными мужиками **хлеб-соль**...* (Ф. К. Сологуб. Публицистика разных лет (1904–1918)); *Родня родней, а **старую хлеб-соль** забывать тоже нехорошо*. (Д. Н. Мамин-Сибиряк. Золото (1892)); *Принимаю с удовольствием **вашу хлеб-соль**...*(Д. В. Григорович. Замшевые люди. (Заноза) (1891)); *Кушай, князь, не брезгай **нашей хлебом-солью!*** (А. К. Толстой. Князь Серебряный (1842–1862)); *Вот **моя хлеб-соль** на дорогу* (А. И. Герцен. Кто виноват? (1841–1846)). Примеры использования определения в женском роде в постпозиции (10 случаев): *...**хлебом-солью** моею брезгует, ну ладно, упомянет, молодец!* (Ю. П. Герман. Россия молодая. Часть вторая (1952)); *Ей чего, она баба здоровая. Помянула добром **хлеб-соль нашу***. (А. С. Эфрон. Мироедиха (1952)); *Услышав это, Илья Муромец выражает желание «голову сложить за государеву чашу и молитвы и за ево **хлеб-соль великую**»*. (Н. К. Гудзий. История древней русской литературы. (XVI–XVII вв.) (1938)); *Но несколько верных старинных друзей Звал русский на **хлеб-соль простую**...* (Ю. Н. Тынянов. Архаисты и Пушкин (1926)); *А все-таки приятели: **хлеб-соль одна*** (А. М. Ремизов. Крестовые сестры (1910)); *За **хлеб-соль** вашу я платил трудом и впредь платить буду...*(М. Горький. Мещане (1901)); *...не обессудили бы **хлеб-соль хозяйскую**...* (П. И. Мельников-Печерский. На горах. Книга вторая (1875–1881)); *...с ними у него **хлеб-соль одна***. (Ф. И. Буслаев. Русские духовные стихи (1861)); *...покорнейше просим принять **хлеб-соль нашу** на дорогу...* (В. А. Сологуб. Тарантас (1845)). У Ф. М. Достоевского обнаружены оба рода у данного слова даже в одном произведении: 1) мужской: *В день похорон мужа гонят с квартиры, после **моего хлеба-соли**, на улицу, с сиротами!* (Ф. М. Достоевский. Преступление и наказание (1866)); 2) женский: *...не говорят у нас что ли: «**Старая хлеб-соль** не помнится?»* (Ф. М. Достоевский. Дневник писателя. 1880 год. (1880)); *То есть сесть за **чужую хлеб-соль**; Выдумав раз **эту хлеб-соль**, она уже ей свято верила*. (Ф. М. Достоевский. Преступление и наказание (1866)); *Хлеб-соль-то наша, говорит, **родственная** возопияла к тебе!* (Ф. М. Достоевский. Ползунков (1848)).

Обнаружена утрата словоизменения данного слова: *Помню одну карикатуру, на которой мировое еврейство принимает блюдо с хлеб-солью...* (В. В. Набоков. *Другие берега* (1954)); *Заготовьте место для хлеб-соли.* (И. Ильф, Е. Петров. *Золотой теленок* (1931)).

Исследование конструкции *хлеб-соль* убедительно показывает, что использование ресурсов Национального корпуса русского языка позволяет проводить многосторонний анализ языкового материала.

#### СПИСОК ЛИТЕРАТУРЫ

1 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru> – Дата доступа: 28.10 2010.

2 Большой толковый словарь русского языка / Сост. и гл. ред. С. А. Кузнецов. – СПб.: «Норинт», 2000. – 1536 с.

РЕПОЗИТОРИЙ ГГУ имени Ф. СКОРИНЫ

# ПРИКЛАДНАЯ ЛЕКСИКОГРАФИЯ. АВТОМАТИЗИРОВАННЫЕ ЛЕКСИКОГРАФИЧЕСКИЕ СИСТЕМЫ

В. А. Житко, И. Т. Давыденко

## ЕСТЕСТВЕННО-ЯЗЫКОВОЙ ИНТЕРФЕЙС ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ

В связи с проникновением компьютерной техники во все сферы человеческой деятельности, остро встает проблема перехода от графического пользовательского интерфейса к интерфейсам программных систем на естественном языке. Это обусловлено имеющимися у естественно-языкового интерфейса преимуществами: минимальной подготовкой пользователя для работы с системой, простотой задания произвольных запросов к пользовательскому интерфейсу, высоким уровнем модели предметной области. Все эти преимущества значительно повышают производительность пользовательского интерфейса.

При рассмотрении пользовательского интерфейса вопросно-ответных систем логично рассматривать пользовательский интерфейс как некую специализированную интеллектуальную систему решающую задачу взаимодействия пользователя с компьютером [1]. Таким образом, естественно-языковой пользовательский интерфейс можно рассматривать как специализированную интеллектуальную систему обеспечивающую взаимодействие пользователя с компьютером на естественном языке.

Естественно-языковой интерфейс вопросно-ответной системы должен обеспечивать диалог пользователя с системой на естественном языке. Полная поддержка естественного языка является сложной комплексной задачей в области компьютерной лингвистике. Поэтому в данной работе будет рассматриваться ограниченный и псевдо естественный язык. Псевдо естественный язык – это язык, текст на котором, без каких либо усилий, воспринимается носителем исходного языка, а так же не требует длительного изучения для приобретения навыков составления текстов этого языка. В качестве выводного языка естественно-языкового интерфейса можно так же использовать псевдо естественный язык [1].

Функционал естественно-языкового интерфейса:

- ввод сообщений от пользователя в естественно-языковой форме;
- анализ сообщений от пользователя и формирование запросов к системе на sc-языке вопросов [1];
- трансляция ответов от системы в тексты естественного языка;
- отображение естественно-языковых сообщений системы пользователю.



Естественно-языковой интерфейс представляет собой независимый компонент вопросно-ответной системы, включающий в себя базу знаний, машину обработки знаний и элементы графического пользовательского интерфейса.

База знаний естественно-языкового интерфейса включает в себя лингвистическую базу знаний используемого языка, лингвистическую базу знаний предметной области и техническую базу знаний естественно-языкового интерфейса.

Машина обработки знаний включает в себя операции анализа различных информационных структур, операции синтеза и стилизации текстов на естественном и псевдо естественном языке. Общая задача компонентов машины обработки знаний является трансляция естественно-языковых запросов пользователя на sc-язык вопросов, а так же трансляция ответов предметной интеллектуальной системы на естественный язык.

Графический интерфейс естественно-языкового пользовательского интерфейса включает в себя просмотрщики и редакторы естественно-языковых текстов.

Трансляторы естественно-языкового пользовательского интерфейса должны осуществлять обработку информационных структур полагаясь исключительно на описание синтаксиса используемого языка и описание семантики предметной области [2]. Описание синтаксиса используемого языка приводится в лингвистической и предметной лингвистической базах знаний. В предметной лингвистической базе знаний описаны специфические для данной области знания о языке. Семантика предметной области задается предметной базой знаний.

Цикл работы естественно-языкового интерфейса начинается с ввода пользователем сообщения на естественном языке. Введенный в систему текст разбивается по предложениям, для каждого предложения строится sc-конструкция, отображающая данное предложение в память системы. Все введенные предложения анализируются в порядке следования, и предшествующие результаты анализа используются для анализа текущего предложения.

На этапе морфологического анализа для каждой словоформы, в sc-конструкции предложения, строятся отношения в соответствии значений её грамматических категорий, т. е. совокупности грамматических значений (род, падеж, склонение и т. д.). Грамматические категории могут присутствовать в базе знаний естественно-языкового интерфейса явно (словоформа присутствует в словаре с описанием её грамматических категорий), либо грамматическая категория может выявляться на основании морфологических правил используемого естественного языка. Такой подход позволяет разработчику, при проектировании естественно-языкового интерфейса, выбирать между быстродействием и объемом проектируемой базы знаний.

На этапе лексического анализа строятся отношения синтаксических связей внутри предложения, выделяются главные и второстепенные члены

предложения, выявляется тип предложения и т. д. Такой анализ протекает поэтапно “наращивая” sc-конструкцию предложения, используя информацию, полученную на этапе морфологического анализа. На этом этапе используются лексические и синтаксические правила используемого языка. В общем случае, такие правила описывают критерии существования синтаксических связей в предложении, при анализе предложения проверяется соответствие таким критериям, и в зависимости от результата создаются либо удаляются соответствующие связи в предложении [3]. Результатом этапа лексического анализа является sc-конструкция лексического разбора предложения в памяти естественно-языкового интерфейса.

На этапе семантического анализа строятся отношения эквивалентности элементов предложения и узлов в памяти системы. На этом этапе решаются проблемы синонимии и омонимии, в этом немаловажную роль играет тот факт, что лингвистическая и предметная база знаний в готовой системе представляют собой единую базу знаний [4]. Таким образом, во время семантического анализа используются знания из предметной области, такие например как область действия отношений, иерархия понятий и отношений, существующие связи между понятиями предметной области и др. Все это помогает отсеять неверные трактовки анализируемого вопроса.

Данный этап является наиболее сложным этапом обработки естественно-языкового текста. В рамках sc-технологии, семантический анализ, представляет собой обработку семантической сети, отражающую результат анализа естественно-языкового текста на всех предыдущих этапах, а также присутствующих в системе знаний о языке и предметной области основной системы. На этом этапе происходит соотношение лингвистической конструкций (совокупность всех sc-конструкций построенных на предыдущих этапах обработки естественно-языкового текста) и конструкций в памяти системы (знания о используемом языке, а также знания системе о предметной области), для выявления отношения соответствия, эквивалентности и т. д.

Результатом семантического анализа является конструкция запроса к системе. Сложность в получении такого результата заключается в семантической неполноте исходного естественно-языкового вопроса. Например, вопрос «В какой класс многоугольников входит понятие квадрата?» в развернутом виде будет выглядеть следующим образом: «Какое понятие является подмножеством понятия многоугольника и надмножеством понятия квадрат?». Из примера видно, что, в процессе развертки, произошли следующие подстановки: высказывание «входит» было развернуто в «является подмножеством», «квадрата» – «понятие квадрата», «класс многоугольников» – «понятие, являющееся подмножеством понятия многоугольника», «многоугольник» – «понятие многоугольника». Такие преобразования протекают на этапе семантического анализа и кодируются на SC-языке.

Таким образом, вопрос, сгенерированный естественно-языковым интерфейсом, будет выглядеть как на рисунке 1. Такие вопросы обрабатывает

универсальный решатель вопросно-ответных систем, являющийся частью вопросно-ответной системы по соответствующей предметной области. Стоит заметить, что аналогичный универсальный решатель есть и в системе естественно-языкового интерфейса, таким образом, система естественно-языкового интерфейса может отвечать на вопросы связанные с используемым естественным языком, в рамках его описания в базе знаний естественно-языкового интерфейса.

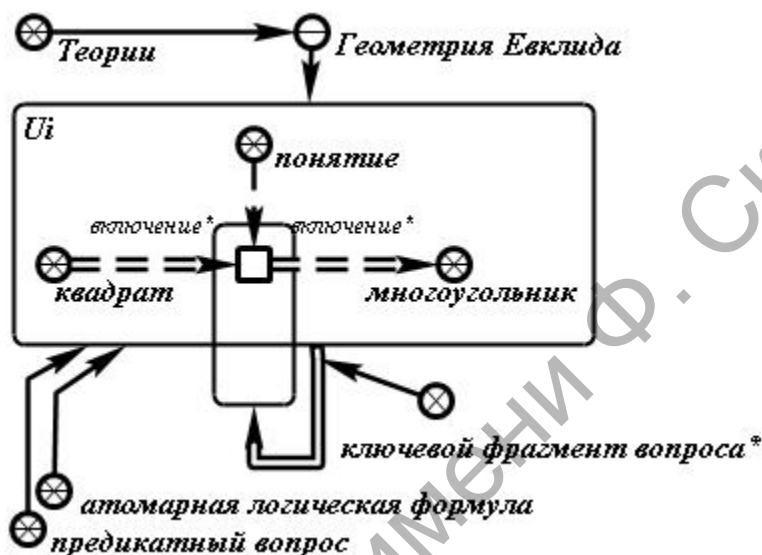


Рисунок 1 – Пример сгенерированного вопроса к системе

Данная модель естественно-языковых интерфейсов предоставляет единую основу для анализа естественно-языковых текстов, начиная с ввода текста, морфологического, лексического, семантического анализа и генерации текста. Так же, семантическая модель естественно-языковых интерфейсов интеллектуальных вопросно-ответных систем, позволяет устанавливать отношения между семантически эквивалентными лингвистическими конструкциями и структурой внутренней памяти системы.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 OSTIS – Open Semantic Technology for Intelligent Systems [Electronic resource]. – Mode of access: <http://www.ostis.net/>. – Date of access: 28.10.2010.
- 2 Апресян, Ю. Д. Избранные труды, том II. Интегральное описание языка и системная лексикография / Ю. Д. Апресян. – М. : Школа «Языки русской культуры», 1995.
- 3 Болдасов, М. В. Генерация содержания запроса пользователя системой InBASE в виде естественно-языкового высказывания / М. В. Болдасов, Е. Г. Соколова // Международной семинар Диалог'2002 по компьютерной лингвистике: сб. науч. тр.: в 2 т. – Протвино : 2002.
- 4 Голенков, В. В. Представление и обработка знаний в графодинамических ассоциативных машинах / В. В. Голенков [и др.]. – Минск : БГУИР, 2001.

## ЧАСТОТНЫЕ СЛОВАРИ И АНАЛИЗ ТЕКСТА

Вопросы методологии и практики изучения художественных текстов нашли широкое освещение в исследованиях как языковедов, так и литературоведов. Эти вопросы отражены в работах Л. В. Щербы, А. М. Пешковского, В. В. Виноградова. Л. В. Щерба предложил методику комплексного изучения литературного произведения, которая «предполагает органическое единство лингвистического, литературоведческого, культурно-исторического аспектов анализа» [1, с.28].

Художественный текст представляет собой «многомерное смысловое пространство», в исследовании которого многие ученые выделяют три основных уровня: изучение фактологической (информационной) стороны текста, исследование композиционной структуры произведения и анализ художественных средств и приемов, посредством которых эта информация передается [2, с. 28]. В последние десятилетия весьма актуально использование точных методов исследования языка конкретного автора или произведения. Особое место занимает составление частотных словарей и анализ полученных данных.

Современные информационные технологии дают возможность изучать язык конкретного автора или произведения путем ввода в компьютер текста и его обработки, поэтому многие традиционные проблемы анализа текста (начиная со сбора фактического материала) в настоящее время решаются по-новому и результаты являются весьма убедительными. В настоящее время наблюдается качественный скачок в создании частотных словарей конкретных авторов, произведений. Материалы частотных словарей активно используются в процессе обучения студентов, учащихся, в криминологии (при определении авторства текстов), психолингвистике. Наряду с информативной ценностью некоторые ученые отмечают их ограниченность, заключающуюся в отсутствии контекстуальных иллюстраций, введение которых значительно увеличило бы объем словаря. Статистическая характеристика текстов является формальным показателем, который способен привести не только к более детальному изучению языка произведений, но и к исследованию динамической картины мира писателя, поэта.

Изданные частотные словари, и словари, созданные на основе отдельных текстов с использованием программы Wordstat, позволяют детально рассмотреть и описать лексические, морфологические, словообразовательные особенности анализируемого текста. Материалы частотных словарей дают точную информацию о степени представленности тематических, лексико-семантических групп слов, стилистически окрашенной лексики, указывают на особенности формообразования и т. д. Анализ данных словарей позволяет ученым делать аргументированные выводы о разном и общем в языке конкретных произведений или авторов. Особое

место в числе современных частотных словарей занимает Частотный словарь языка произведений М. Ю. Лермонтова [3]. Этот словарь отличается своей структурой, подачей материала, интересными комментариями, демонстрирующими глубину анализа текста на материале словаря такого типа. Вместе с данными «Словаря языка Пушкина» [4] предлагаемый словарь позволяет судить как о строении русской лексики, так и о «тематических пристрастиях» эпохи. С другой стороны, частота слов в Словаре отражает и специфические особенности языка Лермонтова. Как отмечают авторы, возможности использования Частотного словаря могут расширяться по мере появления подобных словарей языка других писателей [3, с. 717]. Сравнение с ними языка М. Ю. Лермонтова позволит, например, выделить специфические «лермонтовские» слова. Пока такое сравнение возможно только со «Словарем языка Пушкина», но даже такое попарное сравнение дает исследователям интересные результаты. Сопоставляя материал обоих словарей, ученые определяют слова, которые у Лермонтова встречаются реже, чем у Пушкина. У Лермонтова слабее представлена традиционная «высокая» поэтическая лексика: *воитель, длань, краса, нега* и т. д., меньше используются традиционные слова-символы поэзии: *богиня, волшебник, дань, держава* и многие другие слова. Изучение материалов двух частотных словарей авторов или произведений позволяют выделить слова-лейтмотивы, специфические семантические поля, свойственные конкретному автору или произведению, например, авторы Частотного словаря языка Лермонтова выделяют две специфические «пушкинские» области, мало характерные для Лермонтова: 1) Веселое застолье, шалости, лень: *обед* 21:84 (первая цифра – частота употребления слов в произведениях Лермонтова, вторая цифра – в произведениях Пушкина): *вино* 60:154, *Вах* 2:34, *бокал* 4:30, *пир* 55:119, *бутылка* 8:37, *кубок* 7:31, *пунш* 1:15, *рюмка* 2:25, *веселье* 55:96, *забава* 27:70, *затея* 2:29, *лень* 12:74 и т. п. 2) Поэзия: *лира* 10:129, *муза* 14:148, *поэт* 49:230, *певец* 38:141, *стих* 54:193, *лавр* 2:32, *элегия* 2:20 и т. д. [3, с. 717]. Слова, чаще встречающиеся у Лермонтова, чем у Пушкина, авторами словаря тоже объединяются в несколько групп, в своей совокупности характеризующих Лермонтова-романтика. Таковы группы: Любовь (*грудь, сердце, дорогой, страсть, любить, объятие, ласка, поцелуй* и т. п.); Страдание (*встревожить, тревога, грустить, жаль, зло, мрачный, страх, проклятье, смятение, сожаленье, мука* и т. п.); Ложь (*ложь, обман, стыд, стыдно, честь, клятва*); Смерть (*умирать, смерть, труп*); Бог (*ад, рай, земля, ангел, келья, святыня, храм, творец*); Небо (*звезда, метеор, облако, тучка, солнце, молния, ветер*) и др. [3, с.718]. Анализ такого материала убедительно указывает на устойчивые лейтмотивы (семантические инварианты) творчества Лермонтова.

Частотные словари, созданные нами на основе отдельных текстов произведений Лермонтова с использованием программы Wordstat, позволяют, пользуясь достаточно скромным математическим аппаратом, определить

коэффициент богатства языка конкретного произведения и (при сопоставительном анализе нескольких произведений) коэффициент ранговой корреляции, что сделает выводы об особенностях языка произведений Лермонтова более убедительными.

Говоря о богатстве языка произведения, надо иметь в виду богатство словаря произведения, определяемое при помощи величины ( $R$ ), которую условно назовем коэффициентом богатства словаря. Он может показывать отношение количества слов ( $K_c$ ) к общему количеству словоупотреблений ( $U$ ), однако такое соотношение этих величин не является точным. Для введения более точного показателя богатства языка необходимо учитывать наблюдения и выводы венгерского ученого Ф. Папа, который еще в шестидесятые годы двадцатого века, проследившая употребление писателем новых слов, не встречавшихся ранее в тексте, сделал вывод о том, что количество новых слов на каждые сто словоупотреблений зависит от объема текста: чем больше увеличивается объем текста, тем меньше новых слов приходится на каждые сто словоупотреблений [5, с.101]. После деления любого текста на следующие друг за другом отрезки в сто словоупотреблений каждый, определяется, сколько слов, не встречавшихся до сих пор, содержится в данном отрезке текста. Определив количество новых слов в каждом отрезке текста указанного объема и внося в формулу поправку  $P$ , которая равна  $\sum d_i$ , где  $d_i$  – количество новых слов на каждые сто словоупотреблений, можно рассчитать коэффициент богатства конкретного произведения по формуле:

$$R = 1 + \frac{K_c - O_c}{K_c},$$

где  $K_c$  – количество слов в тексте,  $O_c$  – количество новых (ожидаемых слов). На материале частотных словарей таких произведений Лермонтова, как «Аул Бастунджи», «Измаил-бей», «Песня про купца Калашникова», «Штосс», «Мцыри», «Демон», «Маскарад», «Ашик-Кериб» подсчитаны коэффициенты богатства названных текстов. Самый высокий коэффициент богатства языка у драмы «Маскарад» ( $R=1,53$ ), самый низкий – у произведения «Ашик-Кериб». Эти показатели тесно связаны со временем написания произведений, зрелостью поэта, сюжетом, количеством героев и т. д.

При сопоставительном анализе двух произведений, объединенных темой, идеей, временем коэффициент ранговой корреляции позволит сделать обобщения, выводы более точными и аргументированными. Такое сравнение проводится по всем интервалам частот, близость частот слов для каждых двух рангов оценивается специальной числовой характеристикой – коэффициентом ранговой корреляции. Удобство данного коэффициента заключается в том, что он дает возможность проводить сравнение различных пар текстов и выделять более коррелированные и менее коррелированные пары. Существуют разные виды коэффициентов корреляции, и

выбор определенного вида коэффициента ранговой корреляции определяется конкретными условиями задачи. Этот коэффициент используется тогда, когда неудобно или невозможно точно измерить наличие какого-либо признака у объектов, корреляция между которыми устанавливается. В этом случае ограничиваются тем, что располагают, «ранжируют» эти объекты в соответствии с «количеством данного признака» и в дальнейшем каждый объект представлен лишь своим местом («рангом») в последовательности других объектов. Степень корреляции выражается числом, которое и называется коэффициентом ранговой корреляции

$$\rho = 1 - \frac{6 \sum d^2}{n^3 - n},$$

где  $\rho$  – коэффициент ранговой корреляции,  $d^2$  – разность между рангами,  $n$  – количество сопоставляемых пар.

Например, при сопоставлении существительных с самой высокой частотой в произведениях «Аул Бастунджи» и «Измаил-бей» коэффициент ранговой корреляции указывает на положительную корреляцию между произведениями –  $\rho \approx 0,8$ . Коэффициент ранговой корреляции между словами аналогичной группы в «Измаил-бее» и «Демоне» значительно меньше:  $\rho \approx 0,3$ .

Материалы частотных словарей позволяют более точно описать богатство языка произведений, написанных в разные периоды творчества автора, отметить сокращение и рост частоты употребления некоторых семантических групп слов, определить тенденции в употреблении языковых средств. Результаты, полученные в процессе изучения частотных словарей, являются составляющей анализа текста, дополняют выводы и делают их более аргументированными.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Щерба, Л. В. Опыт лингвистического толкования стихотворений / Л. В. Щерба // Избранные работы по русскому языку. – М. : Учпедгиз, 1957. – 348 с.
- 2 Аверинцев, С. С. Филология. Русский язык: Энциклопедия / под ред. Ф. П. Филина. – М. : Сов. Энциклопедия, 1979. – 746 с.
- 3 Частотный словарь языка М. Ю. Лермонтова / Под ред. В. В. Бородина, А. Я. Шайкевича; Сост. Авдеева А. А., Бородин В. В., Быкова Н. Я., Козокина С. М., Гордеева Н. А., Макарова Л. А., Шайкевич А. Я. // Лермонтовская энциклопедия / АН СССР. Ин-т рус. лит. (Пушкин. Дом); Науч.-ред. совет изд-ва «Сов. Энцикл.» – М. : Сов. Энцикл., 1981. – С. 717–774.
- 4 Словарь языка Пушкина: в 4 т. / Отв. ред. акад. АН СССР В. В. Виноградов. – 2-е изд., доп. / Российская академия наук. Ин-т рус. яз. им. В. В. Виноградова. – М. : Азбуковник, 2000.
- 5 Пап, Ф. Количественный анализ словарной структуры некоторых русских текстов / Ф. Пап // Вопросы языкознания, 1961, № 6 – С. 98–106.

**ФРЕЙМ ТЕКСТА ОПИСАНИЯ ЗАМКА  
ДЛЯ РУССКОГО И АНГЛИЙСКОГО ЯЗЫКОВ  
(на материале дескрипторного словаря лексем)**

Значение туризма в современном мире постоянно возрастает, что связано с важностью его развития как для экономики отдельной страны, так и для социальной и других сфер жизни общества. Каждый год миллионы людей совершают туристические поездки. Так, в 2004 году в Республике Беларусь зарегистрировано 67517 прибытий и 514593 отбытий<sup>1</sup>, и с каждым годом туристический поток только возрастает.

Являясь источником валютных поступлений, а также средством обеспечения занятости населения, индустрия туризма приводит к появлению в стране различного рода организаций: фирм, предлагающих туристические услуги турагентов, туристических оптовых фирм, которые превратились в туроператоров, предлагающих полный комплекс услуг в виде инклюзив-туров. Сюда же можно отнести крупные корпорации, основанные на капитале транспортных, торговых, страховых компаний и банков, осуществляющие операции по представлению туристических услуг клиентам.

Указанные выше организации, имея целью привлечение все большего количества клиентов, нуждаются в разработке соответствующих продуктов, в частности туристических маршрутов, карт-туров, а также описаний туристических объектов. Описание туристического объекта рассматривается нами как особый вид дискурса, имеющий определенные статусно-ролевые, ситуативно-коммуникативные и другие характеристики, а также вполне конкретные мотивы, цели и стратегии реализации коммуникации.

Основной целью любой коммуникации является оказание определенного воздействия на адресата. Успешность реализации текста описания туристического объекта определяют как используемая структура, так и его лексический состав. Разрабатываясь стихийно, такие тексты создаются часто без участия лингвистов, а, скорее, специалистами в области рекламы.

Имея целью установление общих критериев, которым должно соответствовать описание туристического объекта, а также процедуры по вербальному наполнению такого текста, мы проанализировали более 40 текстов описаний определенного типа туристического объекта, а именно замка. Наличие большого количества туристических объектов различных типов, времени создания и места нахождения, и соответственно огромного числа текстов, описывающих данные туристические объекты, обусловило невозможность отражения всей генеральной совокупности. В связи с этим и возникла необходимость в выборе конкретного типа объекта, для чего были разработаны определенные критерии отбора, среди которых наличие

---

<sup>1</sup> По данным: Национальная программа развития туризма в Республике Беларусь 2006–2010 гг. – с. 6 и Республика Беларусь. Статистический ежегодник. – с. 271–272



у данного туристического объекта «статуса объекта Всемирного наследия» ЮНЕСКО (<http://whc.unesco.org/en/list/>), а также определенного времени создания.

Таким образом, был создан корпус, в который вошли тексты описания на различных языках следующих объектов: Мирский замок (Беларусь), замок Кронборг (Дания), замок Литомишль или Литомышль (чеш. Litomyšl, бывш. нем. Leitomischl) (Чехия), а также замок Мальборк (Мариенбург, польск. Malbork, нем. Marienburg) (Польша) и крепостные сооружения Беллинцоны (Швейцария). Для каждого из текстов описаний был создан дескрипторный словарь лексем, на основе анализа которого были составлены примерные статические фреймы текстов описаний замка для русского и английского языков (понимая под *фреймом* «некоторую структуру данных», граф или модель, отражающую и служащую «для представления стереотипной ситуации» [1, с. 2], здесь текста описания замка; фрейм также можно интерпретировать в виде «сети, состоящей из узлов и связей между ними» [1, с. 2]).

Центральный слот, а именно понятие всегда справедливое для текста описания замка, в рассматриваемой фреймовой структуре занимает понятие «castle» (англ.) – «замок» (русс.). На следующем, более низком уровне фрейма как для английского, так и для русского языков располагаются вершины-терминалы, несущие данные о названии замка (например, «malbork» (англ.) – «мальборк» (русс.), его местонахождении (например, «europa» (англ.) – «европа» (русс.); «place» (англ.) – «место» (русс.) и др.), истории данного объекта и основных исторических личностях, принимавших участие в его создании, восстановлении, реставрации.

Интересным представляется отличие рассматриваемых фреймов в слотах, содержащих данные об истории объекта. В русском варианте обнаружена большая конкретизация данных и, соответственно, большее количество лексических единиц, что, перегружая его обилием информации, не являющейся необходимой для адресата, с нашей точки зрения является скорее недостатком для текста описания замка. Кроме того, в данной подструктуре фрейма обязательны ячейки, содержащие информацию о дате или времени происходящих событий (обязательны понятия «year» (англ.) – «год» (русс.), «century» (англ.) – «век» (русс.) и т. п.), а также личностях, принимавших в них участие. Здесь имеются слоты с данными о событии из истории объекта, где встречаются такие понятия, как, например, «renovation», «modernization», «rebuilding» (англ.) – «ремонт», «реставрация/реставрационные работы», «перестройка» (русс.).

Важной особенностью текстов описаний рассматриваемого типа туристического объекта является наличие информации о его значимости, где обязательно обнаруживают себя ячейки «UNESCO heritage list» (англ.) – «список наследия ЮНЕСКО» (русс.), «unique» (англ.) – «уникальный» (русс.), «world wide» (англ.) – «всемирное» (русс.) и др.

Понятию «castle» (англ.) – «замок» (русск.), заполняющему центральный слот фрейма, сопутствуют понятия, определяющие и называющие рассматриваемый туристический объект. Например, «building» (англ.) – «здание, постройка» (русск.); «complex» (англ.) – «комплекс» (русск.); «palace» (англ.) – «дворец» (русск.); «residence» (англ.) – «резиденция» (русск.).

К вершинам-терминалам более низкого структурного уровня фрейма следует отнести такие немаловажные данные, как структура объекта, его внешний и внутренний вид (например, «tower» (англ.) – «башня» (русск.); «room» (англ.) – «комната, покой» (русск.); «façade» (англ.) – «фасад» (русск.)), где отметим значительное совпадение лексического наполнения для английского и для русского языков.

Особый интерес вызывает наполнение ячеек в подструктурах фрейма, несущих информацию о признаке объекта, его элемента. С одной стороны, имеются определенные соответствия: сюда относятся понятия, характеризующие расположение объекта относительно сторон света («northern» (англ.) – «северный» (русск.); «southern» (англ.) – «южный» (русск.); «eastern» (англ.) – «восточный» (русск.) и др.); относительно других элементов («central» (англ.) – «центральный» (русск.) и др.); характеризующие материал, из которого он выполнен («wooden» (англ.) – «деревянный» (русск.), «brick» (англ.) – «кирпичный» (русск.) и др.), а также его размеры («large» (англ.) – «большой» (русск.) и др.).

С другой стороны, явно отмечается обилие понятий, дающих оценочную характеристику, для английского языка по отношению к аналогичной подструктуре фрейма для русского («unique», «exceptional», «impressive» и др.). Также отметим богатство лексического наполнения ячеек, содержащих характеристику формы объекта, элемента для русского языка по отношению к английскому (например, «полуциркульный», «квадратный», «овальный» и др.). Обратим также внимание на отсутствие цветообозначений в рассматриваемых фреймах.

В случае описания замка как объекта маркетинга возникает подструктура, ячейки которой содержат данные о часах открытия объекта, близлежащих отелях, турах, программа которых включает посещение описываемого замка.

На основе рассматриваемых фреймов разрабатывается определенный набор критериев и принципов создания соответствующего текста, учет которых обеспечит успешность реализации описаний определенных типов туристических объектов (в данном случае замка), в том числе повысив их привлекательность для клиента.

#### СПИСОК ЛИТЕРАТУРЫ

1 Минский, М. Фреймы для представления знаний / М. Минский. – М. : Мир, 1979. – 217 с.

## СТРУКТУРНО-СЕМАНТИЧЕСКАЯ ХАРАКТЕРИСТИКА КОМПЬЮТЕРНЫХ ЭМОЦИОНОВ

Электронная коммуникация открыла принципиально новые возможности для общения. Потребность полноценного общения в Интернет вызывает к жизни новые знаковые системы. Одной из таких знаковых систем являются эмодзи.

Эмодзи (эмоцион, эмодзи, смайл, смайлик) – идеограмма, изображающая эмоцию, чаще всего составленная из типографических знаков. Эти идеограммы получили распространение в Интернет и мобильном Интернет, однако в последнее время используются повсеместно (рекламные логотипы, логотипы компаний и т. д.)

Впервые эмодзи – классический вариант – появились в Америке. Европейские страны заимствовали идеограммы, незначительно изменив их структуру, в результате чего система получила новые виды смайликов. Однако принципиально новыми образованиями стали азиатские смайлики – каомодзи. Азиатские страны заимствовали средства изображения смайликов, полностью изменив их структуру, усложнив и дополнив ее новыми символами. Предполагаем, что со временем стили эмоционов выработают единый для всех культур вариант в графическом изображении смайлов. Основанием для этого служит редко встречающееся гибридное образование, графический облик которого включает элементы классического и азиатского вариантов эмоционов.

В ходе анализа было отмечено, что эмодзи, как и слова, могут быть охарактеризованы с семантической, грамматической и структурной точек зрения. Грамматическое значение эмоционов можно объединить в группы. Преобладают эмоционы-существительные, которые можно отнести к «собственным» и «нарицательным». «Нарицательные» идеограммы имеют значение конкретное и абстрактное. «Собственные» идеограммы можно классифицировать как персонажные. Отмечены эмоционы-прилагательные. Эмоционы-глаголы единичны.

Толкование эмодзи нельзя рассматривать вне их структуры. Структуру можно рассматривать в узком и широком значении. В узком – это совокупность знаков, из которых они состоят. В соответствии с этим, в состав эмоционов входят:

- знаки реального языка (буквы латинской графики, буквы кириллического письма, параграфемные знаки, которые делятся на пунктуационные и диакритические).
- знаки искусственных языков, которые являются подязыками реальных языков (математические; знаки компьютерной графики, знаки нумизматики, знаки машинописи.)

В широком значении структуру невозможно рассмотреть без сочетания с семантикой. На основе этого можно выделить простые, сложные, усложненные и упрощенные идеограммы.

Все эмоции строятся на одной базовой идеограмме :-) «улыбка». Пример демонстрирует, что чаще всего эмоции состоят из трех символов. Основными из них являются «;» в значении глаз, «-» в значении носа и «)» в значении рта. Два первых символа являются основными и меняются редко. Третий символ, передающий значение рта, является семантическим центром и всегда варьируется, по сравнению с остальными знаками, в зависимости от тональности информации, которая передается от адресата к адресанту. На наш взгляд, можно говорить о метафорической проекции эмоции из реального языка в виртуальный.

Проиллюстрируем данное положение на конкретных примерах. Сравним два эмотикона: базовый :-) «улыбка» и эмоции со структурным элементом D, который участвует в придании смайлам, его содержащим, экспрессивный характер: «:-D / X-D / :'-D / >:-D / D-:». В базовом смайле символ «)», обозначающий рот, трансформируется в символ D со значением «открытый рот». Заглавная буква означает интенсивную эмоцию. Первые три эмоциона образуют синонимический ряд, носящий градационный характер: «:-D» - «громкий смех»; «X-D» - «смех с зажмурившимися глазами»; «:'-D» - «смех до слез». Эмоцион «>:-D» «злорадный смех» вступает в антонимические отношения с членами синонимического ряда. Он образован на основе синонимической доминанты :-D, однако, отрицательное значение придает символ > (закрывающая угловая скобка). Последний символ, «D-:» - ярость, противопоставлен всем синонимическому ряду эмоционов и не вступает с ними в парадигматические отношения. Таким образом, 4 эмоциона «:-D / X-D / :'-D / >:-D» вступают в антонимо-синонимические отношения. Как видим, в базовом смайле символ «)», обозначающий рот, трансформируется в символ D со значением «открытый рот». Заглавная буква означает интенсивную эмоцию. В силу своей схожести с широкой улыбкой, эта буква несет в основном положительную семантику. Однако при помощи дополнительных символов либо смены местоположения может приобретать отрицательное значение, хотя оно не является основным. Этот символ в своей семантике совмещает два диаметрально противоположных значения, в силу этого основной смысл вычленяется из всей структуры смайлика в целом.

Также продуктивным является символ «O/o». Рассмотрим смайлики «8-o и =-O». Как заглавная, так и прописная буквы несут значение «открытый рот», тождественное рассмотренному выше знаком «D». Однако отличие данных символов в том, что символ «O/o» используется в выражении эмоции «удивление, сильное удивление». Заглавная буква указывает на интенсивность эмоций. Из этого следует, что символы «O» и o противопоставлены друг другу в категории интенсивности, хоть и указывают,

по сути, на одну и ту же эмоцию. В таком случае формируется синонимический ряд на основе градации: «8-о - =-О - :-[ ]». В последней идеограмме квадратные скобки означают «отвисшая челюсть». Определенные трудности вызывает вычленение доминанты ряда. В эмоцион «8-о» «о» передает неинтенсивное значение удивления, в то время как символ 8 в данном образовании усиливает общее значение. В эмотикине «=-О» символ, передающий значение глаз, не придает дополнительных оттенков значения в данном контексте. Однако мы рассматриваем семантику буквы «О», и на этой основе полагаем, что доминантой является идеограмма 8-о, так как строчная буква не передает интенсивного значения.

Данный символ участвует в образовании не только эмоций. Также он с его помощью образуются смайлики, изображающие персонажей:

- \*:O) – клоун
- o-)/O-) – циклоп
- O:-) – ангел

Очевидно, что в данном контексте буква «О» приобретает значение в первом случае конкретного предмета – атрибута грима артистов; во втором – приобретает статус одного из основных структурных элементов, иначе говоря, является вариантом символа:, передающего значение глаз; а в третьем случае – означает «нимб», то есть становится дополнительным структурным символом.

Таким образом, данный знак является полифункциональным символом, может быть как основным так добавочным структурным элементом, передавать как основные, так и добавочные оттенки значений.

Рассмотренные нами положения и примеры позволяют нам говорить о том, что каждый из элементов графики и пунктуации передает одно или несколько значений, может вступать в синонимические, антонимические, омонимические отношения, а также иметь различные полисемантические значения. При этом элемент структуры эмоциона приобретает статус символа в значении «знак, имеющий скрытый смысл, обладающий семантикой».

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Символ// Википедия [Электронный ресурс]. – 2009. – Режим доступа: <http://ru.wikipedia.org>. – Дата доступа: 28.10.2010.
- 2 Словоново – словарь современной лексики, жаргона и сленга [Электронный ресурс]. – 2009. – Режим доступа: <http://www.slovonovo.ru>. – Дата доступа: 28.10.2010.
- 3 Смайл// Википедия [Электронный ресурс]. – 2009. – Режим доступа: <http://ru.wikipedia.org>. – Дата доступа: 28.10.2010.
- 4 Язык сети Интернет/Стилистический энциклопедический словарь русского языка /под ред. М. Н. Кожинной: члены редколлегии: Е. А. Боженова и др. – 2-е изд. – М. : Флинта: Наука, 2006. – 696 с.

## **ИНТЕГРАЦИЯ ГЕОИНФОРМАЦИОННЫХ ДАННЫХ В ИНТЕЛЛЕКТУАЛЬНЫХ СПРАВОЧНЫХ СИСТЕМАХ**

Общение пользователей в привычных для человека формах является залогом успешного использования созданного программного продукта. В интеллектуальных системах достаточно большое количество информации (фактографических данных и знаний) имеет пространственную ориентацию, что предполагает использование картографических материалов и, следовательно, требует проектирования картографических интерфейсов для такого класса систем, что ведет к расширению функциональности интеллектуальной системы. Таким образом, первоочередной задачей является интеграция картографических материалов и в целом геоинформационных данных в интеллектуальные системы для организации картографического интерфейса.

Представление знаний в базах знаний интеллектуальных систем предполагает первичное семантическое осмысление информации для заданной предметной области. С этой целью на первом этапе проектирования интеллектуальных систем создается онтология предметной области, которая включает формальную запись ключевых объектов и понятий, а также установление отношений между ними. Особо отметим, что для использования картографического интерфейса необходима онтология карт, включающая: формальную запись всех типов используемых картографических объектов и установление отношений между ними, а также кодирование координат, совместимое с кодированием знаний в интеллектуальных системах. Создание такой онтологии непосредственно и обеспечивает интеграцию геоинформационных данных в интеллектуальных справочных системах.

Для разработки интеллектуальных справочных систем используется семантическая технология проектирования интеллектуальных систем (проект OSTIS.net – Open Semantic Technology for Intelligent Systems) [1]. В соответствии с предложенной технологией: 1) знания, используемые в интеллектуальных системах представляются в виде семантической сети специального вида; 2) проектирование интеллектуальных систем основано на принципе компонентного проектирования, что позволяет многократно использовать уже спроектированные компоненты и баз знаний, и операций обработки знаний (машин обработки знаний), и интерфейсов; 3) используется принцип открытого кодирования.

Возможность использования картографической информации в рамках технологии OSTIS предполагает разработку совместимых компонентов, которые обеспечивают интеграцию карт в базу знаний интеллектуальной системы. Существует множество форматов хранения картографической информации. Их можно разделить на два класса: внутренние (закрытые) и обменные (открытые). Внутренние форматы – это форматы, поддерживаемые определенными геоинструментальными средствами

(например, MAP Info, Arc Info и др.). Такие форматы имеют внутреннюю защиту и могут быть использованы только в совокупности с соответствующими инструментальными средствами. Обменные форматы, напротив, могут быть использованы в любой системе без необходимости применения каких-либо дополнительных средств и их основная задача – обмен информацией между системами или подсистемами [2].

В качестве исходных данных предложено использование карт, подготовленных в открытом обменном формате mid/mif [2], что соответствует идеологии проектирования интеллектуальных систем по технологии OSTIS.

Карта в формате mid/mif представляется как совокупность слоев, каждый из которых описывает набор родственных, схожих, принадлежащих одному классу объектов, но слои карты имеют связность за счет одной системы координат. Mif-файл содержит служебную информацию, задает структуру объектов слоя, их атрибуты, содержит координаты объектов. В mid-файле определены атрибуты объектов, которые объявлены в mif-файле.

Все объекты карты, представленной в данном формате должны соответствовать ОКРБ 012-2007 [3]. Данное требование обеспечивает совместимость всех слоев карты и позволяет рассматривать карту как источник новых знаний, т. е. в базе знаний добавятся знания не только о координатах географических объектов, но и фактографические данные – характеристики объекта, а также знания о типе географического объекта. На основе данного классификатора [3] строится онтология карт и картографических материалов, удовлетворяющие принципам технологии OSTIS. На рисунке 1 приведен фрагмент семантической сети, соответствующий следующей классификации: класс 3 – гидрография и гидротехнические сооружения, подкласс 31 – гидрография, группа 311 – водоемы.

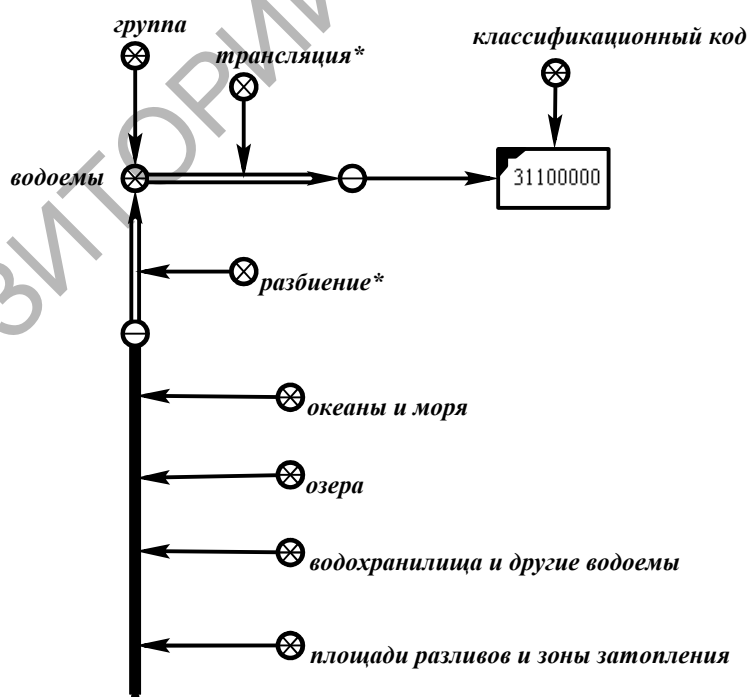


Рисунок 1 – Фрагмент семантической сети объекта классификации

Для того, чтобы интеллектуальная система могла работать с картой во внешнем формате, разработан программный модуль конвертации карты во внутренний формат интеллектуальной системы – семантическую сеть. В результате его работы и использования семантической сети классификатора топографической информации объекты преобразуются в структуры, с которыми может работать система и записывается в файл \*.scs. В итоге вся полученная геоинформация интегрируется в базу знаний системы. На рисунке 2 приведен фрагмент семантической сети, соответствующий географическому положению объекта гидрографии.

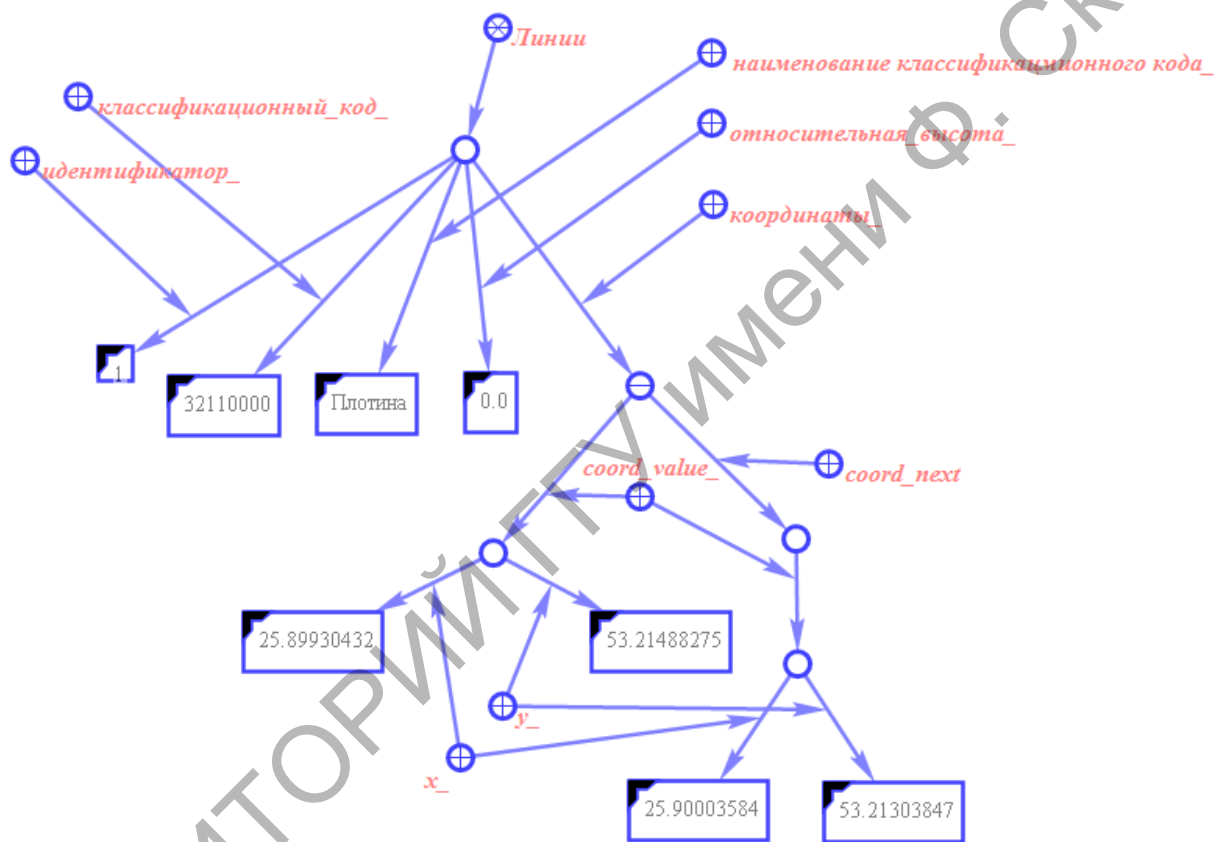


Рисунок 2 – Фрагмент семантической сети, соответствующий географическому положению объекта гидрографии (плотина)

Представление координат объектов осуществляется с использованием иерархической структуры – каждый узел, обозначающий координату, хранит ее текущие координаты и указывает на узел следующей пары. Для площадных и полилинейных объектов характерно то, что они могут состоять из нескольких частей. Это предусматривается в конвертере и реализуется за счет введения отношения «состоять из\*».



После конвертации в памяти интеллектуальной системы формируется семантическая сеть, в которой интегрирована картографическая информация, а с помощью пользовательского интерфейса отображается в привычной для человека форме, как показано на рисунке 3.

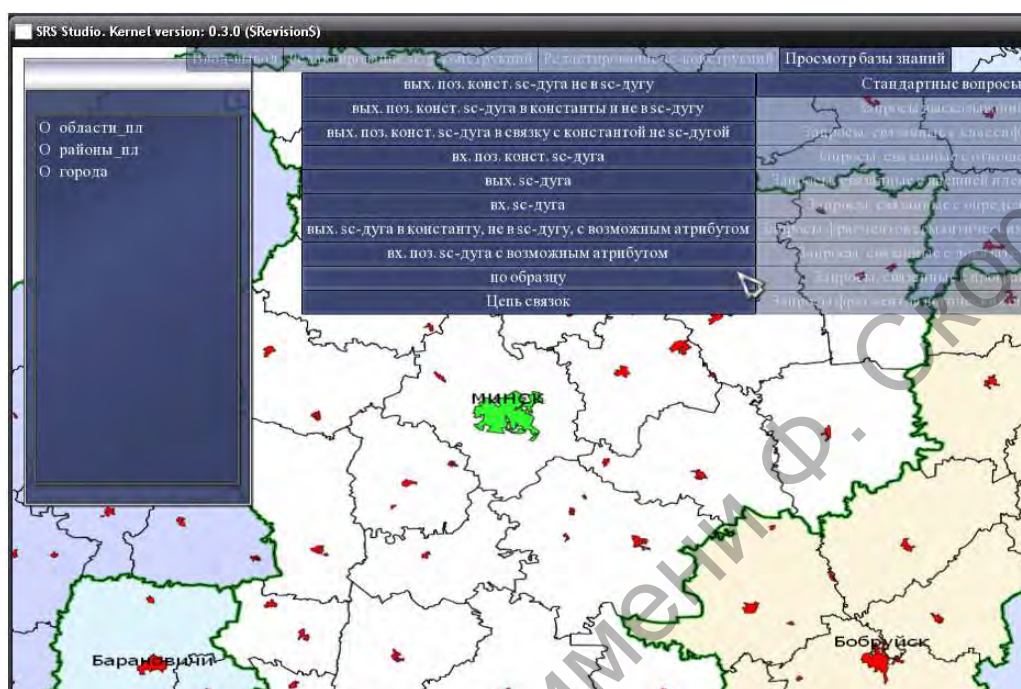


Рисунок 3 – Отображение информации карты в системе

Приведенный в работе подход к интеграции геоинформационных данных в память интеллектуальных справочных систем, реализованных в соответствии с технологией OSTIS, позволяет использовать разработанные компоненты как непосредственно для проектирования геоинформационных систем с интеллектуальным выводом, так и использовать на уровне многократно используемых компонентов карты местности. Скомпонованная система протестирована на примере местности – Республика Беларусь, а разработанные инструментальные средства позволяют загружать в память интеллектуальных систем карты, соответствующие классификатору [3] в обменном формате mid/mif.

Работа выполнена при поддержке молодежного гранта БГУИР «Научно-методическое обеспечение дисциплин специализации «Интеллектуальные геоинформационные системы».

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Open Semantic Technology for Intelligent Systems [Electronic resource]. – 2010. – Mode of access: <http://www.ostis.net>. – Date of access: 25.10.2010.
- 2 Интеллектуальные технологии в геоинформационных системах /А. Н. Крючков и др.; под науч. ред. В. В. Голенкова. – Минск: БГУИР, 2006. – 203с.
- 3 Цифровые карты местности. Топографическая информация, отображаемая на топографических картах и планах городов / ОКРБ 012-2007.

А. Ю. Станкевич

## СВЯЗНОСТЬ СОЧЕТАНИЙ КВАЗИЛЕКСЕМ: ИНСТРУМЕНТ ОПРЕДЕЛЕНИЯ

Под квазилексемой понимаем именованную коллекцию подстрок, наличие любой из которых на некотором фрагменте текста считаем входом квазилексемы в текст. Множество именованных текстовых фрагментов называем телом квазилексемы, имя этого множества (может быть произвольным, но обычно совпадает с одним из тел) называем именем квазилексемы. Текстовые фрагменты могут соотноситься как с одной словоформой, так и с некоторым участком текста, содержащим более одной словоформы. Означаемое квазилексемой определяется запросами пользователя, но обычно одной квазилексемой рационально обозначать единицы, объединенные общностью лексического и/или грамматического значения. Тело квазилексемы может быть задано регулярным выражением: например, тело квазилексемы СТАНОВИТЬСЯ КАКИМ, определяющей конструкцию вида «спрягаемые и неспрягаемые формы глагола *становиться* + зависимое прилагательное / местоименное прилагательное в Тв. п.», можно задать так:

$\{\backslash\text{станов}[a-я\text{ё}]+\backslash s+[a-я\text{ё}]+(?:\text{им}|\text{ими}|\text{ым}|\text{ыми}|\text{ей}|\text{ой})\backslash b\}$ .

Под словарем квазилексем понимаем некоторое непустое неупорядоченное множество квазилексем. Словарь квазилексем, единицы которого объединены общностью лексического значения, называем лексически ориентированным. В качестве примера приведем фрагмент словаря цветообозначений; для удобства восприятия начальные и конечные пробелы обозначаем через «\_»: {ЗЕЛЕНЬИЙ: \_зелен, \_позелен; КРАСНЬИЙ: \_красн; \_красен; покрасн; ОГНЕНЬИЙ: огненн, цвета огня, цвет огня, цветом огня}. Словарь квазилексем, единицы которого объединены общностью грамматического значения, называем грамматически ориентированным. Таким словарем может стать набор квазилексем, описывающих конструкции деривационных сочетаний (пример тела квазилексемы СТАНОВИТЬСЯ КАКИМ приведен выше), набор квазилексем, описывающий строевые слова научного текста.

Традиционно определение устойчивости сочетаний слов производится с обращением к функциям, принимающим следующие значения: объем корпуса в словоупотреблениях (N), значение частоты встречаемости биграммы  $f(n, c)$ , значения частот независимых употреблений компонентов биграммы  $f(n)$  и  $f(c)$ .

Вычислить критерий связности для лексем русского языка можно, обратившись к сервисам: Querying Internet corpora С.А. Шарова [1], «Поиск биграмм» А.Н. Аверина (ресурс сайта группы АОТ) [2], Sketch Engine от Lexical Computing Ltd. [3]. В первом случае вычисляются значения MI-score, Ti-score, Loglikelihood-score; во втором – значения MI-score и  $f(n, c)$ ;

статистический аппарат сервиса Sketch Engine описан в [4]<sup>1</sup>. Поиск по пользовательской коллекции текстов и переопределение лексем не поддерживаются.

Среди открытых программных продуктов, позволяющих определение пользовательского шаблона для ключевого слова и работу с пользовательской коллекцией текстов, можно назвать корпус-менеджер AntConc [5]. Программа позволяет вычислять значения MI-score и T-score, однако работа с коллекцией шаблонов для кириллического текста не поддерживается. Ввиду этого возникла необходимость разработки утилиты, отвечающей следующим требованиям: 1) поддержка работы с пользовательским словарем квазилексем по пользовательской коллекции текстов; 2) расчет нескольких критериев связности, предположительно пригодных для оценки связности компонентов лексически и грамматически ориентированных словарей квазилексем.

Ниже опишем возможности разработанной нами утилиты расчета критериев связности квазилексем.

На вход поступают файл словаря и файл текста. Файл текста автоматически разбивается на предложения; для текстов с большим числом сокращений используется предобработка для определения точки как конца и не-конца предложения.

Файл словаря представляет собой строки формата тело квазилексемы; имя квазилексемы. Например, представленная выше квазилексема ЗЕЛЕНый собрана по следующему фрагменту входного файла:

\_зелен; зеленый  
\_позелен; зеленый

Избыточность представления имен, удобная при наборе входного файла словаря, снимается при сборке рабочего словаря. Сборка рабочего словаря квазилексем по входному файлу организована так, что в одну квазилексему объединяются все тела, сопоставленные одному имени, даже если они рассредоточены по всему входному файлу; дублирование тел порождает исключение (что является своеобразной проверкой корректности входного файла).

Критерии связности рассчитываются отдельно для левого и для правого окружения квазилексемы, что достаточно традиционно.

Левым/правым окружением считается набор словоупотреблений в непересеченном знаками препинания окне, ширина которого по умолчанию равна 3 и может быть переопределена (обычно не рекомендуется использовать окна шире 5 (см., например [6]). Под словоупотреблением понимается ограниченный пробелами (на границах окна – пробелом и/или знаком препинания) набор символов латиницы, кириллицы, арабских

---

<sup>1</sup> Отличительной особенностью сервиса Sketch Engine, для русского языка определяющего коллокат к лексеме по Web-корпусу С.А. Шарова, является дифференциация выдачи по грамматическим категориям, с оценкой связности леммы и категории; сервис платный, его trial-версия доступна в течение 30 дней после регистрации.

цифр; точка в сокращении и дефис не считаются границей словоупотребления.

При таком определении одним словоупотреблением считаются следующие группы единицы с дефисным написанием: *Миллера-Ингве, ЕЯ-процессор, head-banding, Log-log, l-форме, иммуцилин-h, Т-лимфоцитов-киллеров, XIX-XX* (группа с дефисом/дефисами в последовательностях букв кириллицы, латиницы, кириллицы и латиницы), *схсr4-тронных, e2f-2, 7-го, sb-18-подобный* (группа с дефисом/дефисами в последовательностях букв и цифр), *18-24, 274-25-78* (группа с дефисом/дефисами в последовательности цифр).

Результаты расчетов критериев связности для левого и правого окружения сохраняются в два файла. Приведем фрагмент результирующей таблицы для правого окружения (словарь строевых слов научного текста):

Таблица 1

QUAZILEX	COLLOC	MI-score	OE-fMI	MI-q	T-score	fqs-sqr	Z-score	N-chain
<...>	<...>	<...>	<...>	<...>	<...>	<...>	<...>	<...>
Метод	ph_анализ	4.8094	28.0395	4.8084	<b>5.623</b>	5.831	29.7751	34
Метод	технологии	7.1855	145.5627	7.1845	<b>4.7629</b>	4.7958	57.4639	23
Метод	машинного	8.8006	445.9177	8.7996	<b>4.3491</b>	4.3589	91.8394	19
Метод	обучения	7.9772	251.9867	7.9762	<b>4.1067</b>	4.1231	65.1908	17
Метод	исчисления	9.0381	525.7135	9.0371	<b>3.7345</b>	3.7417	85.6272	14
<...>	<...>	<...>	<...>	<...>	<...>	<...>	<...>	<...>

Мера связности выдается для пар квазилексема – словоформа и квазилексема – квазилексема (в этом случае имя квазилексеммы-коллоката снабжено префиксом "ph\_", т. е. меткой для фильтра списка коллокатов).

Поле **QUAZILEX** содержит квазилексемму, поле **COLLOC** – коллокат к ней.

Утилита рассчитывает и записывает в соответствующие поля следующие критерии: MI-score, T-score, Z-score. Также вычисляем MI-q, вариант MI-score, где учитывается накопленная частота тел квазилексем.

Расчет ведется по приведенным ниже формулам:

$$MI - score = \log_2 \frac{f(q,c) * N}{f(q)f(c)},$$

$$MI - q = \log_2 \frac{f(q,c) * N^2}{(N + NQ) * f(q) * f(c)},$$

$$TI - score = \frac{f(n,c) - \frac{f(n) * f(c)}{N}}{\sqrt{f(n,c)}},$$

$$Z\text{-score} = \frac{f(n,c) - \frac{f(n)*f(c)}{N}}{\sqrt{\frac{f(n)*f(c)}{N}}},$$

где  $f(q,c)$  – частота коллокации элемента  $c$  (словоформы или квазилексемы) с квазилексемой  $q$ ,  $f(q)$  – частота квазилексемы,  $f(c)$  – частота коллоката к квазилексеме,  $N$  – объем текста в словоупотреблениях;  $NQ$  – накопленная частота тел квазилексем некоторого словаря.

О применимости этих критериев для оценки связности русскоязычных коллокаций разного функционального статуса см., например, [6–8].

Применительно к расчету MI-score в работе [9] отмечается некоторая искусственность оценки связности единиц ЕЯ через логарифмирование по основанию 2 отношения частоты наблюдаемой ( $O = \frac{f(n,c)}{N}$ ) к частоте ожидаемой ( $E = \frac{f(n)*f(c)}{N^2}$ ).

Автор работы, М. Стаббс (M. Stubbs), отмечает, что MI-score и  $O/E$  (равно как и T-score и  $\sqrt{f(n,c)}$ ) являются вариантами; статистические преобразования, по его мнению, иногда скрывают первоначальные значения и тем самым затрудняют интерпретацию данных о связности. Аргументируя этим, включаем в результирующую таблицу следующие поля:  $O/E$ , со значением отношения наблюдаемой частоты к частоте ожидаемой; поле  $fqs\text{-}sq$ , со значением квадратного корня из  $f(q, c)$ ; поле  $N\text{-}chain$ , со значением накопленной частоты коллокации.

Предполагаем, что набор критериев, вычисляемых утилитой, позволит работать как с лексически, так и с грамматически ориентированными словарями квазилексем.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Sharoff, S. Querying Internet corpora / S. Sharoff // [Electronic resource]. – University of Leeds. – Mode of access: <http://corpus1.leeds.ac.uk/internet.html>. – Date of access: 20.10.2010.
- 2 Аверин, А. Н. Поиск биграмм / А.Н. Аверин // [Электронный ресурс]. – АОТ. – Режим доступа: <http://www.aot.ru/demo/bigrams.html>. – Дата доступа: 20.10.2010.
- 3 Sketch Engine // Lexical Computing Ltd [Electronic resource]. – Mode of access: <http://www.sketchengine.co.uk>. – Date of access: 20.10.2010.
- 4 Statistics used in the Sketch Engine // Lexical Computing Ltd [Electronic resource]. – Mode of access: <http://trac.sketchengine.co.uk/attachment/wiki/SkE/DocsIndex/ske-stat.pdf>. – Date of access: 20.10.2010.
- 5 AntConc // L. Antony [Electronic resource]. – Mode of access: [http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html). – Date of access: 20.10.2010.
- 6 Захаров, В. П. Статистический метод выявления коллокаций / В. П. Захаров, М. В. Хохлова // Языковая инженерия: в поисках смыслов / Доклады семинара «Лингвистические информационные технологии в Интернете»: XI Всероссийская объединенная конференция интернет и современное общество. Под ред. В. Ш. Рубашкина и В. П. Захарова. – СПб., 2008 [[http://www.infosoc.ru/2008/pdf\\_CL/Zakharov&Hohlova.pdf](http://www.infosoc.ru/2008/pdf_CL/Zakharov&Hohlova.pdf)].

7 Ягунова Е. В. Природа коллокаций в русском языке. Опыт автоматического извлечения и классификации на материале новостных текстов / Е. В. Ягунова, Л. М. Пивоварова // НТИ. – Сер.2. – 2010. – № 6. – С. 30–40 [[http://www.webground.su/services.php?param=priroda\\_collac&part=priroda\\_collac.htm](http://www.webground.su/services.php?param=priroda_collac&part=priroda_collac.htm)].

8 Браславский, П. П. Сравнение четырех методов автоматического извлечения двухсловных терминов из текста / П. П. Браславский, Е. Соколов // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» / под ред. Н. И. Лауфер, А. С. Нариньяни, В. П. Селегея.– М., 2006. – С. 88–94.

9 Stubbs, M. Collocations and semantic profiles: on the cause of the trouble with quantitative studies / M. Stubbs // Functions of Language. – 1995. – Vol. 2:1. – P. 23–55 [<http://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf>].

О. Ю. Шиманская

## **ИСПОЛЬЗОВАНИЕ МЕТОДА БАЗОВЫХ ОПРЕДЕЛЕНИЙ И СЕМАНТИЧЕСКИХ ОППОЗИЦИЙ ПРИ СОСТАВЛЕНИИ ЭЛЕКТРОННОГО ДВУЯЗЫЧНОГО СЛОВАРЯ МЕТАФОР**

Электронный двуязычный словарь метафор психологической сферы человека является уникальным ресурсом, представляющим собой как результат научного исследования метафорических моделей внутреннего мира человека в белорусском и английском языках, так и средство обработки лексикографической информации. На основании сопоставления метафорических и мотивирующих значений полисемантов в базе данных словаря осуществляется определение синонимов и эквивалентов, а также сортировка метафор по семантическому критерию в соответствии с запросом пользователя.

При составлении электронного словаря был разработан и апробирован метод базовых определений. Данный метод опирается на категорию семантических оппозиций, с одной стороны, и на дескрипторную теорию метафоры, с другой. Теория семантических оппозиций предполагает установление соотношений между значениями на основе наличия или отсутствия в значении определенной семы (частные оппозиции) и наличия семы 'интенсивность признака' в структуре значения парадигматически связанных метафор (градуальные оппозиции) [1].

Анализ метафор внутреннего мира человека показывает, что большинство из них являются не номинативными, а характеризующими метафорами, т. е. они используются не для наименования процессов, состояний и свойств, а для их оценки как приятных или неприятных, интенсивных или неинтенсивных, адекватных или неадекватных. В данном случае оказывается, что при систематизации метафор, обозначающих, к примеру, состояние аффекта, совершенно неважным является состав исходного значения и его тематическая соотнесенность – важным оказывается наличие или отсутствие семы

‘интенсивность’ в данном значении либо ассоциативная связь данной семы с мотивирующим значением. Сравним: «агонь / ураган / землятрес у души», «chemistry / storm / volcano of emotions» ‘химия / ураган / вулкан эмоций’.

Дескрипторная теория метафоры, разработанная в исследованиях А. Н. Баранова [2], описывает один из подходов к «расшифровке» метафоры с выделением в структуре значения сигнификативных и денотативных дескрипторов, соотносимых с областью источника и областью цели. Дескрипторы, используемые в словаре А. Н. Баранова, подобно семантическим примитивам, теория которых обосновывается в труде А. Вежбицкой [3], не могут быть использованы для систематизации данных в электронном словаре в силу чрезмерного дробления значений. Так, результаты исследования метафор в белорусском и английском языках показывают, что выделение единиц сопоставления <внезапное неприятное чувство> и <внезапное неприятное воспоминание> более обосновано, чем оперирование т. н. «элементарными смыслами» <внезапный>, <не>, <приятный>, <чувство>.

Приведенные выше единицы сопоставления и есть базовые определения, разработанные нами в процессе создания словаря. Базовые определения основываются на словарных толкованиях (семантический и коннотативный аспекты), контекстном употреблении метафор (структура метафорического высказывания, типичная сочетаемость метафор с определенной семантикой), а также определенном уровне обобщения («приравнивание» метафор, семантика и контекстное употребление которых не совпадает, но и не противоречит друг другу).

Нами были разработаны около 150 базовых определений на белорусском языке, неполный перечень которых приводится ниже. Следует отметить, что данные определения существуют только в базе данных словаря как своего рода «элементарный язык», в словарных же статьях (в электронных карточках) толкование метафор приводится более дифференцировано и приближено к толкованию словарей (за исключением случаев, где словарное толкование неполное или неудовлетворительное, например, за счет отсылки к иной метафоре). Кроме того, если денотативные определения являются именно определениями, то сигнификативные скорее обозначают тематический класс исходного понятия, лексико-семантическое поле источника.

Базовые определения области источника: *прадмет: ежа, посуд, адзенне, тканіна, асвятляльныя прыстасаванні, вупраж, механізмы, амуніцыя, зброя і прыстасаванні, збудаванні і лакальна абмежаваныя прасторы, шляхі і дарогі, часткі прадмета, рэчыва; прырода: нябесныя целы, элементы зямной паверхні і ландшафту, прыродныя ўтварэнні, расліны і іх часткі, жывёлы, часткі цела і покрыва жывёл, групы жывёл і іх жытло, гукі і дзеянні, якія ўтвараюць жывёлы; надвор’е: атмасфера і рух паветра, аблогі і ападкі, з’явы прыроды і надвор’е, агонь і гарэнне, вада і плынь; чалавек: тактыльнае ўспрыманне, тэмпературнае ўспрыманне, смакавае ўспрыманне, зрокавае ўспрыманне, слыхавое ўспрыманне, хваробы і выздараўленне, палітыка і вайна; працэсы і адносіны: працэсы, якія адбываюцца ў рэчывах ці з рэчывамі*

(прадметамі), віды энергіі і фізічныя з’явы, узаемадзеянне паміж прадметамі і яго вынікі, суб’ектныя дзеянні, прасторавыя і параметрычныя адносіны.

Базовые определения области цели: *падманлівая знешнасць; раптоўныя інтэнсіўныя пачуцці, афекты; пра тое, што абмяжоўвае свабоду, прыносіць неспакой; стрыманасць; незакончаная думка; узнікненне пачуццяў, якасцей; вяртанне сядомасці, актыўнасці; тое, што скоўвае, перашкаджае развіццю; бляск вачэй як адлюстраванне стану; энтузіязм; адзінота; прызнасць; тонкая розніца ў праяўленні пачуцця; вялікая колькасць, інтэнсіўнасць чаго-н.; прытворная ласкавасць; непрыемны ўспамін; нечаканае непрыемнае пачуццё; актыўная пазіцыя ў спрэчцы; абраза.*

Одно базовое определение используется для описания от 4 до 47 отдельных метафорических значений, например: <сум> (47 значений), <вялікая колькасць слоў, уражанняў> (41), <афекты> (38), <супярэчлівыя пачуцці> (21). Что касается дескрипторов области источника, то в тематических группах обнаруживается от 4 (например, в группе <нябесныя целы>) до 60 слов (например, в группе <вада і плынь>) [4].

Для обработки материала была создана база данных в Microsoft Office Access; каждому базовому определению был присвоен определенный числовой код (номер), использовавшийся при дальнейшей систематизации метафор в базе данных, код языка («белорусский» или «английский»), а также набор числовых кодов-дескрипторов, соотносимых с базовыми определениями (сигнификативными и денотативными).

Таблица 1 – Фрагмент основной базы данных

Word ID	WordSe nsNum	Word Name	WordDefinition	WordExample	Word Lang
697	1	TWIS T	1) an unusual feature of a person's character, typically an unhealthy one.	If in a mind so beautiful any moral twist or handiness could be found, Miss Sally Brass's nurse was alone to blame. (Ch. Dickens).	1
698	2	TWIS T	2) * a particular feeling or emotion.	He smiled, but her humility gave him a curious twist in his heart. (W. Maugham).	1

Подбор синонимов одного языка и эквивалентов другого осуществляется Запросами: при совпадении кода базового определения слова с совпадающим кодом языка определяются как синонимы, а несовпадающим – как эквиваленты. Сортировка данных осуществляется благодаря установлению связей между различными значениями (подчеркнем, что одно слово может иметь несколько различных метафорических значений).



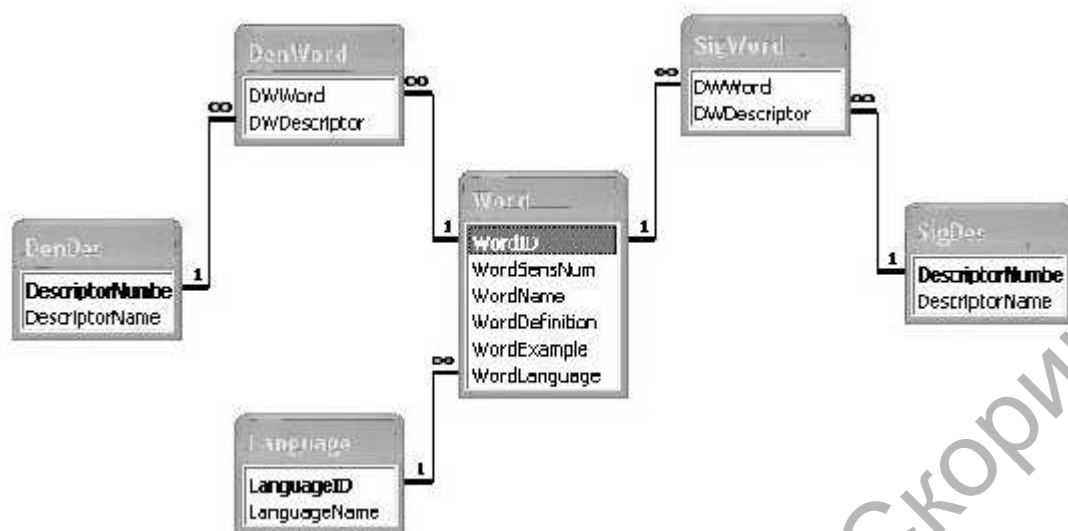


Рисунок 1 – Схема данных

Таким образом, первый в отечественной лексикографии электронный двуязычный словарь метафор психологической сферы человека, составленный с опорой на метод базовых определений и теорию семантических оппозиций, позволяет осуществлять поиск метафор относительно цели и источника метафоризации, подбирать синонимы и эквиваленты, а также дает возможность систематизации лексикографического материала в соответствии с запросом пользователя.

Электронный словарь внедрен в учебный процесс кафедры иностранных языков Международного института трудовых и социальных отношений (Минск) и используется для совершенствования навыков перевода, формирования лингвокультурных компетенций учащихся, повышения мотивации к изучению белорусского и английского языков.

Эффективность методов составления словаря свидетельствует в пользу того, что разработанный нами подход является обоснованным и может быть использован при составлении других электронных словарей, словник которых представлен лексемами, имеющими некоторые общие значения и для которых возможно установление общих базовых определений.

#### СПИСОК ЛИТЕРАТУРЫ

1 Баранов, А. Н. Дескрипторная теория метафоры и типология метафорических моделей / А. Н. Баранов // Научно-образовательный портал «Лингвистика в России»: ресурсы для исследователей [Электронный ресурс]. – 2004. – Режим доступа : [http://www.uisrussia.msu.ru/linguist/\\_A1\\_2\\_4\\_1\\_metaphor.jsp](http://www.uisrussia.msu.ru/linguist/_A1_2_4_1_metaphor.jsp). – Дата доступа : 20.03.2008.

2 Вежбицкая, А. Семантические универсалии и описание языков / А. Вежбицкая ; пер. с англ. А.Д. Шмелева ; под ред. Т. В. Булыгиной. – М. : Языки славянской культуры, 1999. – 776 с.

3 Методы сопоставительного изучения языков: сб. ст. / общ. ред. и вступ. ст. В. Г. Гака. – М. : Прогресс, 1989. – Вып. 25 : Контрастивная лингвистика. – 440 с.

4 Шиманская, О. Ю. Электронный двуязычный словарь метафор психологической сферы человека / О. Ю. Шиманская // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции “Диалог” (Бекасово, 26-30 мая 2010 г.) / Ин-т проблем информатики РАН, РГГУ ; редкол.: А.Е. Кибрик [глав. ред.] и др. – Вып. 9 (16). – М.: Изд-во РГГУ, 2010. – С. 584–589.

Е. А. Юрина

## **АКТУАЛЬНЫЕ ЗАДАЧИ КОМПЬЮТЕРИЗАЦИИ ДИАЛЕКТНОГО АРХИВА ГОВОРОВ СРЕДНЕГО ПРИОБЬЯ И ПУТИ ИХ РЕШЕНИЯ**

Диалектный архив, хранящийся в Лаборатории общей и сибирской лексикографии при кафедре русского языка Томского государственного университета, представлен тетрадами с рукописными записями речи носителей русских старожильческих говоров, расположенных на территории Томской области (частично – Новосибирской и Кемеровской областей), а также словарными картотеками. Начало его формирования датируется 1946 годом. Именно в этом году состоялась первая послевоенная диалектологическая экспедиция, организатором и руководителем которой была заведующая кафедрой русского языка ТГУ доцент А. А. Скворцова. Так томские лингвисты включились в работу Института русского языка АН СССР по сбору материалов к «Диалектологическому атласу русского языка». В числе первых были В. В. Палагина, возглавлявшая затем кафедру русского языка в периоды 1957–1969, 1971–1989 гг., П. Г. Черемисин, Е. П. Молчанова и др. Для обследования были выбраны старожильческие сёла, основанные первопоселенцами: Томский район – сёла Вершинино и Батурино; Туганский район – сёла Халдеево, Турунтаево, Подломск и др. Источником изучения была живая разговорная речь сельских жителей.

Со временем определились типичные фонетические, лексические и грамматические черты речи сибирских старожил: аканье (*па начам*), иканье (*гричи ха*), долгие твердые шипящие (*та`шшшт*), стяжённые формы местоимений, прилагательных и глаголов (*така` харо`ша*, ) и др. О. И. Блиновой была предложена трактовка термина «русский старожильческий говор Сибири» с опорой на совокупность фонетико-грамматических и лексических черт [1]. Полная характеристика исследуемых говоров, представляющих собой особый тип вторичного сибирского старожильческого говора с северной диалектной основой, изложена в коллективной монографии «Русские говоры Среднего Приобья» [2, 3]. В книге описаны типологические черты среднеобского диалекта и дана языковая характеристика каждой группы говоров в его составе: Нарымской, Прикетской, Приобской, Притомской, Причулымской [2, с. 30–33].

Первые тетради записаны в транскрипции: «*на фс'о`j но`чи гул'а`jut / д'ефк'и да роб'а`ты...*» (1946 г.). Впоследствии речь информанта записывалась на магнитофон, а затем расшифровывалась в так называемой «полутранскрипции». В этих записях употреблялись буквы я, ю, е, ё, не обозначалась мягкость согласных перед гласным переднего ряда, не ставилось ударение, если оно не отличалось от литературного, ставились знаки препинания, писались заглавные буквы. С 1962 года была введена в учебный план существующая и по сей день диалектологическая практика для студентов-филологов.

В результате за 66 лет работы коллективом томских диалектологов было обследовано свыше 400 населенных пунктов Томской и Кемеровской области. На сегодняшний день архив представлен 1500 тетрадями с записями диалектной речи, отражающей русские старожильческие говоры Среднего Приобья. На базе данного материала были созданы картотеки диалектных словарей более чем на 1 000 000 единиц: картотеки Среднеобского диалектного дифференциального словаря [4, 5], Словаря диалектно-просторечной лексики [6], Полного словаря говора села Вершинино [7], Полного словаря диалектной личности [8] и целого ряда других словарей разного типа. Библиотека среднеобских словарей насчитывает на сегодняшний день более 30 томов.

Томский диалектный архив является уникальным памятником народной речевой культуры, бесценным источником исследования истории и современного состояния русского национального языка. Собранные диалектные материалы позволяют проводить широкий круг разноаспектных исследований (см. подробнее в [9]).

Задача создания электронного варианта текстового архива говоров Среднего Приобья, а также компьютерного словаря среднеобского диалекта назрела очень давно. Её актуальность была осознана уже двадцать лет назад, в период становления отечественной компьютерной лингвистики. Именно в конце 80-х годов Г.А. Раков, заведовавший кафедрой русского языка ТГУ с 1990 по 1994 годы, начал исследовательскую деятельность, связанную с созданием машинного идеографического словаря среднеобских говоров [10–12]. Им была организована работа исследовательской группы по компьютерному набору диалектных текстов, автоматическому созданию конкорданса, семантической разметке словарного массива. К сожалению, разработки были прерваны в связи с трагическим уходом из жизни этого замечательного человека и ученого, а созданные в то время электронные материалы утрачены.

За последние десятилетия компьютерные технологии шагнули далеко вперед. Сегодня мы наблюдаем расцвет корпусной лингвистики на базе Интернет-технологий. В 2004 году начал функционировать Национальный корпус русского языка [13]. Работа над созданием электронных корпусов русской диалектной речи ведется во многих научных центрах России: в Институте русского языка РАН в рамках проекта «Национальный корпус русского

языка» [14], в Санкт-Петербургском, Саратовском [15, 16] и других государственных университетах.

Создание диалектного корпуса на материале говоров Среднего Приобья крайне необходимо. Уже имеется некоторый задел в виде базы электронных текстов (около 10 % от имеющегося рукописного архива), компьютерный набор которых осуществлен диалектологами-исследователями и студентами, проходившими диалектологическую практику. И если 20 лет назад, в период становления компьютерной лингвистики, актуальность исследований была связана с поиском и разработкой методов компьютерной обработки текстовых массивов, то сегодня перед Томской диалектологической школой стоят не менее актуальные задачи прикладного характера. Они состоят в эффективном использовании уже имеющихся в арсенале корпусной лингвистики технологий, позволяющих сохранить архив, наиболее ранние записи которого находятся в аварийном состоянии; создать максимально полные текстовую и словарную базы данных и осуществлять их ведение; оптимизировать поиск данных, их статистическую обработку и анализ при помощи современной программной платформы (составление конкордансов, словников, подкорпусов, соответствующих исследовательским задачам). Применение имеющихся в арсенале корпусной лингвистики методов позволит вывести исследования Томской диалектологической школы на новый уровень, соответствующий современным техническим возможностям.

Для решения поставленной задачи в 2010 году группой томских диалектологов и программистов (Е. А. Юрина, В. В. Поддубный, С. В. Волошина, М. А. Толстова, О. Г. Шевелев) разработан проект создания диалектного Корпуса говоров Среднего Приобья. В результате работы планируется полный перевод рукописных текстов на электронные носители, проведение лингвистической разметки текстов с целью последующей машинной обработки, создание программ по автоматизированной разметке текстов, машинному поиску, анализу результатов поиска, создание лингвистических баз данных и программной оболочки корпуса. Основным результатом проекта станет действующий, доступный для широкого круга исследователей диалектный корпус, который в последствии может быть интегрирован в качестве подкорпуса в более крупные национальные проекты.

Предполагается поэтапно осуществить следующие виды работ:

1. Перевод рукописных записей диалектных текстов в электронный формат методом компьютерного набора и создания электронных текстовых файлов. Для решения этой задачи проводится оцифровка рукописных тетрадей и сохранение изображений в формате DJVU. Более 50 % диалектных тетрадей уже оцифровано. Проводится набор текстов и редактирование ранее набранных текстов для унификации графических символов. С этой целью разработана единая система графической передачи особенностей диалектной фонетики и морфологии в тексте с использованием

символов юникода. Общерусские слова и словоформы, не имеющие отличий от литературного языка, передаются в орфографической записи, а диалектные, диалектно-просторечные единицы и диалектные варианты общерусских слов записываются с передачей фонетических и морфологических особенностей. В них проставляется ударение. Например: «Игры вся́ки ра́зны были. Пляса́ля ши́тко. У нас был Введе́нье [праздник]. Отку́пишь дом на три-четыре дня. Деньги соберёшь с молодёжъя. Отдашь хозяину. Поу́жинашь вечером и на вече́рку. В дома́-то таки́. Для молодёжу» (с. Вершинино Том. обл., Вершинин И. В., 1909 г.р.).

2. Проведение метаразметки массива диалектных текстов, задающей параметры паспортизации текстов, квалификации тематики и жанровой специфики. Паспортная часть включает сведения об информантах, о времени и месте записи, собирателе. Жанрово-тематическая характеристика текста включает указание на время описываемых событий (двадцатые 20 в; тридцатые 20 в; сороковые 20 в и т. д.); тематику (Война, Семья, Огород, Колхоз, Дом, Односельчане, История села, Болезни и лечение и т. д.); жанр (Биографический рассказ, Сюжетный рассказ, Описание, Рассуждение, Интервью, Сказка, Песня, Частушка, Пословица). Тематика задается через указание на ключевой концепт, определяющий смысловое единство текста: Сватовство, Свадьба, Похороны, Крестины, Приметы, Праздники, Развлечения, Чужие, Пьянство... (всего 34 единицы, маркирующие тему).

Паспортная часть задается перед текстом в следующем виде:

#МестоЗаписи с. Вершинино, Томский район, Томская область

#МестоЗаписиКоротк Верш

#ДатаЗаписи 1991

#Информант Вершинин Иван Васильевич

#ИнформантКоротк ВершининИВ

#ИнформантПол муж

#ИнформантГодРожд 1906

#ИнформантМестоРожд Неизв\_информантместорожд

#ИнформантОбразов Неграмотн

#Собиратель МихалеваЕВ

#Тетрадь 1168

#Тематика Семья, Война, Свадьба, Праздники, Развлечения, Кедровый промысел, Труд

#ВремяСобытий двадцатые 20в, сороковые 20в

#Жанр Биографический рассказ

Далее за паспортной частью следует запись диалектного текста.

Приведем в качестве примера короткий фрагмент записи речи указанного диалектоносителя:

*Я-то на Курской Дуге воевал. Орёл – Ку`рска Дуга. У-у, деточки мое`, всего там насмотрелись. Чего только не ви`дывали. Взя`тый в сорок втором году был. А так оставляли все по семейному положению. Дети были. Чё. Пришел в ноябре сорок пятого. Везде были, все видели. В Литве*

*был, в Германии был, в Норвегии был. В лагере, в плену. Ехали через Швецию, Финляндию. Старуха не ждала уж. Всё, grit, по`мер, наверно. ...Пой там ей-ей. Было. Самолеты, орудия, минометы. Дело было горя`че. Орёл... Ку`рска дуга.*

3. Создание сводного электронного словаря говоров Среднего Приобья. Решение этой задачи предполагает формирование базы электронных версий всех диалектных среднеобских словарей, а также создание и разметку сводного словника. Сводный словарь будет включать список всех лексем и их вариантов (акцентологических, фонематических, словообразовательных, морфологических); список морфологических форм лексем; указание постоянных и изменяемых морфологических признаков; толкование лексических значений; квалификацию единицы с точки зрения соотношения с формами национального языка (общерусское, диалектное, диалектно-просторечное, диалектный вариант общерусского слова); квалификацию семантических свойств слова (мотивированное/немотивированное, экспрессивное/нейтральное, образное, оценочное, интенсивное и т. д.); стилистическую маркированность (народно-поэтическое, сниженное, бранное).

4. Разработка и создание программной платформы корпуса, позволяющей осуществлять автоматизированную разметку текстов, хранить лингвистические базы данных, осуществлять многоаспектный поиск по заданным параметрам языковых единиц, проводить статистическую обработку материала, сохранение результатов поиска.

Решение намеченных в начале исследовательского пути задач связано с осуществлением большого объема практической работы. И тот факт, что к её осуществлению с большим энтузиазмом подключаются студенты и аспиранты филологического факультета ТГУ, внушает разработчикам чувство оптимизма и уверенность в успехе.

#### СПИСОК ЛИТЕРАТУРЫ

1 Блинова, О. И. О термине «старожильческий говор Сибири» // Вопросы языкознания и сибирской диалектологии. – Томск, 1971. Вып. 2. – С. 3–8.

2 Русские говоры Среднего Приобья / О.И. Блинова и др.; под ред. В. В. Палагина. Ч. 1. – Томск, Изд-во Том. ун-та, 1985. – 205 с.

3 Русские говоры Среднего Приобья / О. И. Блинова и др.; под ред. В. В. Палагина. Ч. 2. – Томск, Изд-во Том. ун-та, 1989. – 324 с.

4 Словарь русских старожильческих говоров средней части бассейна р. Оби / сост. О. И. Блинова [и др.]; под ред. В.В. Палагина. – Томск: Изд-во Том. ун-та, 1964. – Т. 1. А – Е. – 143 с.; 1965. Т. Ж – О. – 233 с.; 1967. Т. 3. П – Я. – 249 с.

5 Среднеобский словарь: (Дополнение) / сост. Л. Г. Гынгазова [и др.]; под ред. В. В. Палагина. – Томск: Изд-во Том. ун-та, 1983. – Ч. 1. А – К. – 180 с.; 1986. – Ч. 2. П – Я. – 212 с.

6 Словарь просторечий русских говоров Среднего Приобья / сост. О.И. Блинова [и др.]; под ред. О. И. Блинова. – Томск: Изд-во Том. ун-та, 1977. – 183 с.

7 Вершининский словарь / сост. Т. Б. Банкова [и др.]; под ред. О. И. Блинова. – Томск : Изд-во Том. ун-та, 1998. – Т.1. А – В. – 308 с.; 1999. Т. 2. Г – З. – 309 с.; 2000. Т. 3. И – М. – 318 с.; 2001. Т. 4. Н – О. – 368 с.; Т. 5. П. – 512 с.; Т. 6. Р – С. – 454 с.; Т. 7. Т – Я. 524 с.

8 Полный словарь диалектной языковой личности / сост. О.И. Гордеева [и др.]; под ред. Е.В. Иванцовой. – Томск: Томск: Изд-во Том. ун-та, 2006. – Т. 1. А – З. – 358 с.; Т. 2. И – О. – 338 с.; Т. 3. П – Р. – 324 с.

9 Томская диалектологическая школа: Историографический очерк / под ред. О.И. Блиновой. – Томск: Изд-во Том. ун-та, 2006. – 392 с.

10 Раков, Г. А. Принципы отбора языковых данных при составлении словаря народного говора на машинном носителе / Г. А. Раков // II Всесоюзн. конф. по созданию Машинного фонда русского языка. – М., 1987.

11 Раков, Г. А. Автоматический словарь народного говора как фрагмент Машинного фонда русского языка / Г. А. Раков // III Всесоюзн. конф. по созданию Машинного фонда русского языка. – М., 1989.

12 Раков, Г. А. Диалектные словари и ЭВМ: Подготовка данных и некоторые результаты сравнения / Г. А. Раков // Русские говоры Сибири: Семантика. – Томск, 1995. – С. 53–62.

13 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru>. – Дата доступа: 27.10.2010.

14 Летучий, А. Б. Корпус диалектных текстов: задачи и проблемы / А. Б. Летучий // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. – М., 2005. – С. 215.

15 Гольдин, В. Е. К проекту текстового диалектологического подфонда Машинного фонда русского языка / В. Е. Гольдин // Доклады III Всесоюзной конференции по созданию машинного фонда русского языка. – М., 1990.

16 Гольдин, В. Е. Тематическая разметка и тематический анализ диалектного текстового корпуса / В. Е. Гольдин, Крючкова О. Ю. // Языковая личность – текст – дискурс: теоретические и прикладные аспекты исследования: материалы междунар. научн. конф.: в 2-х ч. Ч.1. – Самара, 2006.

# ИСПОЛЬЗОВАНИЕ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В УЧЕБНО-МЕТОДИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

Е. Ф. Асенчик

## ПОИСК ИНФОРМАЦИИ В INTERNET: НЕКОТОРЫЕ АСПЕКТЫ ОБУЧЕНИЯ

В настоящее время всемирная сеть Internet хранит множество разнообразных информационных объектов. Очевидно, что эффективность использования информационного ресурса такой сложности требует от пользователя Сети определенных знаний, навыков и умений в области организации и проведения поиска информации в Internet.

Именно поэтому большинством программ обучения предусмотрена тема "Поиск информации в Internet".

Поиск информации - задача, которую человечество решает уже многие столетия. По мере роста объема информационных ресурсов, потенциально доступных одному человеку (например, посетителю библиотеки), были выработаны все более изощренные и совершенные поисковые средства и приемы, позволяющие найти необходимый документ.

Поиск информации представляет собой процесс выявления в некотором множестве документов (текстов) всех таких, которые посвящены указанной теме (предмету), удовлетворяют заранее определенному условию поиска (запросу) или содержат необходимые (соответствующие информационной потребности) факты, сведения, данные.

Процесс поиска включает последовательность операций, направленных на сбор, обработку и предоставление необходимой информации заинтересованным лицам.

В общем случае поиск информации состоит из четырех этапов:

- определение (уточнение) информационной потребности и формулировка информационного запроса;
- определение совокупности возможных держателей информационных массивов (источников);
- извлечение информации из выявленных информационных массивов;
- ознакомление с полученной информацией и оценка результатов поиска.

Рассмотрим основные методы поиска информации в Internet, используемые по отдельности или в комбинации друг с другом.

Прежде всего, это использование поисковых машин. В настоящее время этот метод – один из основных и фактически единственный при проведении предварительного поиска. В результате поиска появится список ресурсов Internet, который требуется детально рассмотреть.



Применение поисковых машин основано на использовании ключевых слов, передающихся поисковым серверам в качестве аргументов поиска. В качестве методов рассматривается и непосредственный поиск с использованием гипертекстовых ссылок. Так как все сайты в пространстве Всемирной паутины фактически связаны между собой, то поиск информации может быть произведен путем последовательного просмотра связанных страниц с помощью браузера.

На первый взгляд, этот полностью ручной метод поиска выглядит полным анахронизмом, в то же время этот способ Web-страниц часто оказывается единственно возможным на заключительных этапах информационного поиска, когда механическое исследование уступает место более глубокому анализу.

При этом использование каталогов, классифицированных и тематических списков и всевозможных справочников также относится к этому виду поиска.

Поиск с применением специальных средств – это полностью автоматизированный метод, весьма эффективный для проведения первичного поиска.

Сущность этого метода заключается в применении специализированных программ-спайдеров, которые в автоматическом режиме просматривают Web-страницы, отыскивая на них искомую информацию.

То есть, можно говорить об автоматизированном варианте просмотра с помощью гипертекстовых ссылок (поисковые машины используют похожие методы).

Этот метод является особо эффективным в том случае, если использование поисковых машин не дает необходимых результатов в силу нестандартности запроса, либо других причин.

Еще один метод – анализ новых ресурсов, поиск по которым может оказаться необходимым при проведении повторных циклов поиска, поиска наиболее свежей информации либо для анализа тенденций развития объекта исследования в динамике.

Еще одна причина применения такого вида поиска заключается в том, что большинство поисковых машин обновляет свои индексы со значительной задержкой, вызванной гигантскими объемами обрабатываемых данных, и это упущение обычно тем больше, чем менее популярна интересующая тема. Такое соображение является весьма существенным при проведении поиска в узкоспециальной предметной области.

В процессе реализации информационного поиска следует руководствоваться рядом общих методических рекомендаций.

Несомненно, что характер поиска всецело определяется содержанием поставленной задачи. Формулирование задачи поиска позволяет определить область и особенности используемых источников, облегчить составление программы поиска.

Вначале поиск привязан к теме и к цели детальной, всесторонней разработки ее плана, когда особенно необходим широкий охват источников.

Затем информационный поиск конкретизируется, ограничивается непосредственным содержанием задачи.

В самом общем виде процесс информационного поиска можно разделить на этапы формулирования задачи поиска, разработки рабочей программы поиска, реализации поиска и оформления его результатов.

На занятиях студенты получают следующие задания:

1. Открыть браузер Internet Explorer и запустить поисковую систему Ya.ru
2. Найти в Internet адреса других поисковых систем. В отчете указать ссылки на них.
3. С помощью поисковой системы найти в Internet ответы на приведенные вопросы (указывается перечень вопросов).

В отчете указать ссылки на сайты содержащие ответ по каждому вопросу.

4. Используя любой из поисковых серверов, попытайтесь найти информацию по Вашей специальности или по другой тематике, представляющей для Вас особый интерес.

Оформление отчета

Отчет оформить рукописным или печатным способом, в него включить описания действий и результатов по пп. 1–3 задания.

Оформите в виде таблицы характеристику результатов поиска по п. 4:

Таблица 1

Тематика поиска	Адрес в Сети	Описание и характеристики ресурса	Степень релевантности найденной информации	Ваша оценка ресурса
-----------------	--------------	-----------------------------------	--	---------------------

Затем студентам предлагается осуществить поиск с учётом морфологии, с этими целями для ознакомления им предлагается следующая информация:

Что такое учет морфологии при поиске?

В русском языке слова могут иметь разные формы (падежи, единственное и множественное число и т. д.) В какой бы форме ни стояло слово в тексте страницы (даже отличной от формы слова в запросе), Яндекс найдет его. Например, если задан запрос «идти», то в результате поиска будут найдены документы, содержащие слова «идти», «идет», «шел», «шла» и т. д. На запрос «окно» будет выдана информация, содержащая и слово «окон», а на запрос «отзывали» – документы, содержащие слово «отозвали».

Обратите внимание, что при выборе форм слова для поиска Яндекс различает части речи. Например, если вы задали запрос «компьютер» (существительное), то будут найдены страницы со словами «компьютера» или «компьютерами», но не «компьютерный» (прилагательное).

Яндекс находит документы, в которых слова запроса могут встречаться в разных предложениях. Я лингвист, и для моих научных целей нужно полностью контролировать поведение поисковой машины. Как мне быть?

Чтобы контролировать поведение Яндекса, вы можете воспользоваться расширенным поиском или языком запросов.

Как отказаться от поиска всех форм слова? Мне нужна именно такая форма, которую я ввел в строке запроса.

Для поиска точной формы возьмите нужное слово в кавычки или поставьте перед ним восклицательный знак.

Как сделать так, чтобы по моему запросу находились только определенные формы слова?

Иногда при поиске требуется сделать так, чтобы по запросу «день» находились только формы существительного «день» («день», «дня», «дню» и т. д.), но не глагола «деть» («деньте», «денешь», «денут» и т. д.), а по запросу «Иванов» только формы фамилии «Иванов» («Иванов», «Ивановым», «Ивановы» и т. д.), но не имени «Иван» («Иван», «Ивану», «Иванам» и т. д.)?

Если вы уверены в своем чувстве языка и в том, что указали в запросе именно нормальную форму слова, то для поиска только его форм укажите перед словом два восклицательных знака. Например, [!!день] или [!!Иванов].

Нормальной формой слова считается именительный падеж единственное число. Так, с точки зрения Яндекса, «кот», «коты», «кота» и т. д. - формы одного и того же слова, среди которых нормальной является форма «кот».

Я задал запрос [на дне] и заметил, что Яндекс не ищет слово «на». Что делать?

Предлог «на» является стоп-словом, и по умолчанию не ищется. Чтобы находить его, достаточно перед стоп-словом поставить знак плюса (без пробела между словом и плюсом): [+на дне].

В любом языке есть слова, которые встречаются намного чаще других. Слова, которые встречаются слишком часто, чтобы быть значимыми для поиска, называются стоп-словами. Как правило, это местоимения, предлоги и частицы. Поисковые машины обычно игнорируют стоп-слова, даже если они встречаются в тексте запроса.

Что будет искать Яндекс на запрос [Петр+Барабашкин]? Почему в таких случаях не работает знак «плюс»?

Оператор «плюс» задан в запросе [Петр+Барабашкин] неправильно, как и в запросе [Петр + Барабашкин]. Корректное написание – перед плюсом должен идти пробел, но само слово с плюсом должно писаться слитно: [Петр +Барабашкин]. По запросу найдутся документы, где слова «Петр» и «Барабашкин» расположены не очень далеко, причем слово «Барабашкин» должно присутствовать во чтобы то ни стало, а слово «Петр» – по возможности.

Если вы хотите, чтобы в найденных документах обязательно встречались оба слова запроса, задайте [+Петр +Барабашкин].

При контроле знаний очень важно отделять технические умения и умение работать с информацией. Именно поэтому и вводится еще один этап поиска - этап анализа результатов. Полученные результаты анализируются студентами и оформляются отдельным файлом.

И в заключение укажем, что, в любом случае, только при систематическом и целенаправленном изучении, знании и умении рационально применять систему информационного поиска современный специалист может быть уверен в эффективности и качестве своей работы.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Комер, Д. Принципы функционирования Интернета. Учебный курс / Д. Комер – СПб., 2002.
- 2 Кузнецов, И. Н. Информация: поиск, анализ, защита / И. Н. Кузнецов. – М., 2004.
- 3 Ландэ, Д. В. Поиск знаний в Internet. Профессиональная работа/ Д. В. Ландэ. – М., 2005.
- 4 Прохоров, А. Эффективный поиск в Интернете / А. Прохоров // Компьютер-Пресс. - 2006. – № 2.
- 5 Тактаев, С. Задача поиска информации, существующие подходы и проблемы / С. Тактаев // [Электронный ресурс]. – Режим доступа: <http://www.searchengines.ru/articles/004603.html>. – Дата доступа: 25.10.2010.
- 6 Талантов, М. Профессиональный поиск в Интернете: планирование поисковой процедуры / М. Талантов // Компьютер-пресс. – 1999. – № 8.
- 7 Татарников, О. Новые технологии поиска в Интернет / О. Татарников // КомпьютерПресс. – 2005. – № 10.

Т. Н. Богоедова

### **ИСПОЛЬЗОВАНИЕ РЕСУРСОВ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА В КУРСЕ РУССКОЙ ДИАЛЕКТОЛОГИИ**

В составе национального корпуса русского языка имеется отдельный диалектный корпус. Как известно, диалектный корпус ещё в начале своего формирования. По данным на январь 2008 г., он был представлен 144 099 словоупотреблениями (122 текстами). Для сравнения, основной корпус на тот момент содержал 147 577 522 словоупотребления (42 387 текстов). Поиск в диалектном корпусе пока ещё нуждается в совершенствовании: по семантическим признакам он не осуществляется, по таким грамматическим приметам диалектной словоформы, как «нестандартная флексия», «другая флексия», «нестандартная форма -ся», «наличие/отсутствие -ся», «склоняемая частица», «стяженная форма прилагательного», «формы числа», «отличие в роде» и под. поиск оказывается невозможным. Часто поисковик выдаёт неточную или ошибочную информацию. Яркий, на наш взгляд,

пример: на запрос о прилагательных наряду с нужными словоформами предлагается большое число местоимений и наречий на -о/-е.

Тем не менее, уже сейчас использование материалов корпуса обогатит учебную работу по русской диалектологии новыми приёмами и свежим фактическим материалом.

Прежде всего, данные корпуса могут быть использованы для наблюдений над диалектной лексикой. В частности, студенты могут получить задание выявить диалектные слова различных типов либо в материале всего корпуса, либо (задав подкорпус) в текстах, записанных в определённой местности. Попутно можно будет сделать вывод о частотности диалектизмов того или иного типа в диалектах вообще или в диалектах определённой местности. Например, поиск в говорах Забайкалья по записям 2006 г. даёт следующий результат: преимущественно обнаруживаются собственно лексические диалектизмы: *Она отшатилась* ‘отошла в сторону’, *но пошли там с подружкой или с кем ли, она отшатилась, она ей встретилась. На крыге* ‘глыба льда’, *сидел, а его выбросило. – А, выброси крыгу на берег? – Крыгу на берег! Лёд ломат, а крыги же. Есть больши крыги, есть маленьки крыги. Пошто* ‘почему’ *сѣдни не дала? А завтра ты придѣшь, она даст? Веселушку* ‘балалайку’ *с престола убери! Подходит человек. Весь вот таким, вот я как щас на его гляжу, во всем бусом* ‘неопределённого серого цвета’ *у чем-то. Вот поперѣд* ‘ранее’ *ещѣ у нас было тоже там же эта говорил в Еловке это чѣ та кажется. Я то сечасного* ‘нынешний’ *не знаю, кто был и др. Несколько менее частотными оказываются деривационные диалектизмы: Каково, Мать, спала? Каково спочивала? – Не столь я спала, не столь спочевала, а сколь про Иисуса Христа сны видала. Так эт, че, он болел... Оны гыт, потом, назавтра лучше стало ему, оздоровел. Мы ходили – Уличка, я. Шибко к Афоне ходил. Уж шибко мы ходили! Каженный день. На курганчике огонек горит, Покрай* *этого удал добрый молодец лежит. Я может спужалась. Одна была да спужалася* и др. Приблизительно такая же частотность свойственна и фонематическим диалектизмам: *Прибежали в окошко постукали и дунули, а скрозь стекла та зимой, аж дали ветер прошѣл. Чѣренька та рабынька ничѣ не скажет, а беленька то не надейся. А вы слышите чѣ ѣн говорит? Часты мелкие раночки були* *каленой стрелы. Широкие глубокие були* *копья вострого...* и др. Гораздо реже встречаются диалектизмы семантического типа: *Веселушку с престола ‘со стола’ убери! Чѣренька та рабынька ‘служанка’ ничѣ не скажет, а беленька то не надейся* и др. Из числа этнографизмов удалось выявить лишь глагол *окачивать* ‘в народной медицине – обливать больного водой из ведра, в котором лежали камешки, принесѣнные с поля’: *И я потом мясяц окачивала, болела.*

Материалы корпуса будут исключительно полезны в процессе изучения диалектных грамматических явлений. Данные корпуса позволяют организовать выполнение традиционных заданий типа «охарактеризовать диалектное явление». Например, явление, представленное в корпусе следующим материалом: *а вот доколь могут они строить, вот вокруг этак*

*всё закладывали закладывали...* (Архангельская область, 2007); *а не знаю, можно ли нет найти. Давно не **искаем**, дак* (Архангельская область, 2007); *житники, он **пекётся** просто...* (Архангельская область, 2001-2003); *не, туда она не убредёт, потому что надо перьвым переходить, там Хопёр **текёт** в этот в новый Хопёр* (Волгоградская область, 1999); *там тина / вода **текёт*** (село Белогорное Вольского района Саратовской области, 1999); *бык-то большущий / и так **могёт*** (село Белогорное Вольского района Саратовской области, 1999); *а так как-то сряживались тут... так возьмут эти брюквенны зубы сделают, да в рот складут, да срядятся во всё бело, вот и **бежат**, пугают людей* (Архангельская область, 1996); *меня Господь Истиной Христос, наверно, **берегёт*** (Архангельская область, 1994) и т. п.

Выбрав в поле «грамматические признаки» искомую морфологическую характеристику, можно получить материал для самых разнообразных заданий. К примеру: «сравните систему причастных форм в диалектах и в литературном языке: причастия какого типа в диалектных текстах встречаются часто, какие причастия практически отсутствуют», «охарактеризуйте особенности формообразования деепричастий в диалектах: от каких основ они образуются и с помощью каких суффиксов», «какую синтаксическую функцию чаще всего выполняют в говорах деепричастия совершенного вида» и т. д. Конечно, фактический материал для заданий подобного рода преподаватель может собрать и традиционным способом – путём сплошного обследования диалектных текстов, помещённых в хрестоматии и сборники. Однако на это потребуются многие часы, а не считанные минуты, как в случае обращения к корпусу. Более того, с помощью корпуса студент и сам с интересом и без лишних усилий соберёт нужный фактический материал.

Метаразметка текстов позволяет сузить поиск, инвертировав выбор места записи, года записи, пола информанта. Также может быть задана тематика текстов. Это позволяет проводить исследования, сопоставляя факты, встречающиеся на разных территориях, характерные для разных диалектных систем; быстро находить диалектные слова различных тематических групп и т. д. К сожалению, здесь механизм работы корпуса пока не безупречен: в блоке «Параметры текста» в разделе «Место записи» предлагается выбрать название города (Архангельск, Брянск, Владимир и т. д.), тогда как в разметке текстов название города соответствует региону записи, место записи фиксируется иначе, – например, «село Белогорное Вольского района Саратовской области» или «Архангельская область». Возникающее несоответствие делает в большинстве случаев невозможным отбор текстов по месту записи. Исключение составляют только Забайкалье и Карелия, поскольку здесь регион записи совпадает с местом записи текстов.

Создатели корпуса не ставили перед собой задачу передать фонетическую информацию, однако диалектный материал, размещённый в корпусе, оказывается ценным и в этом отношении. Он позволяет наблюдать яркие проявления многих фонетических диалектных явлений, таких, например,

как преобразование заударных сочетаний гласных после утраты интервокального *j*: *я сама и косила, за короушку, что на нихней земле гуляэт* (Тверская область, 1995–2005), *пораньше, а как уж...* *Вот быват Микола праздник. Весенний Микола, двадцать первого мая* (Архангельская область, 2001); ассимиляция в группе [бв]: *три струенки вот эти, ну житинки, три житинки вот стоя е, три, гоорит, срежь, вот по этим житним ты так прогладь, и, гоорит, кругом тела овяжи. Голого. И нё тронь. Я так сделала, со словами что нать-та, потом этого-да овязала, и болше ни спина болеть не стала* (Архангельская область, 2001); замена звука [ф] сочетанием хв: *пить-ть её все пьютъ, её хто ни пьеть, только один телехвонный столб, а то все пьютъ...* (Тамбовская область, 2001), *один дядя был на хвинской и на польской...* (с. Куськово, Мантуровский район Курской области, 2000); переход [e] в [‘o] в заударной позиции: *вот это хорошоё, вот это я помню хорошо* (Архангельская область, 1999); прогрессивное смягчение заднеязычных: *а уж старенькей-то будешь / ещё хуже / вон как у меня одни жилы // там ведь этот / пчельник был / их / кто / сколько пенькёв* (Саратовская область, 1999); гиперкорректное оканье: *девки в отласники, да и в отласные платки* (Архангельская область, 1996), *стоит сторичок с коклёвочкой. Стоит за коклёвку держится* (Архангельская область, 1994) и многое другое.

Конечно, достаточно полный фонетический или грамматический портрет говора с помощью корпуса получен быть не может, но для демонстрации и анализа конкретных диалектных явлений использование корпуса может оказаться весьма продуктивным.

#### СПИСОК ЛИТЕРАТУРЫ

1 Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru> – Дата доступа: 21.10.2010.

А. В. Браздзіна

### АДМЕТНАСЦІ ВЫКАРЫСТАННЯ МУЛЬТЫМЕДЫА-ТЭХНАЛОГІЙ ПРЫ ВЫКЛАДАННІ КУРСА “ГІСТОРЫЯ БЕЛАРУСКАЙ ЛІТАРАТУРЫ 19 СТАГОДДЗЯ” ЗАМЕЖНЫМ СТУДЭНТАМ

За апошнія дзесяцігоддзе мультымедыа-тэхналогіі трывала ўвайшлі ў практыку навучальнай дзейнасці як сярэдніх, так і вышэйшых устаноў адукацыі. У параўнанні з традыцыйнымі метадамі выкладання мультымедыа-тэхналогіі маюць шэраг відавочных пераваг: дазваляюць узмацніць эфектыўнасць засваення студэнтамі матэрыялу, ствараюць умовы для ўстойлівай матывацыі да вывучэння пэўнай дысцыпліны, дапамагаюць павысіць тэмп падачы інфармацыі. Самым распаўсюджаным у сучаснай практыцы выкладання відам мультымедыа-заняткаў з’яўляецца

слайд-прэзентацыя, якую нескладана стварыць з дапамогай камп'ютарнай праграмы Power Point. Паспяховае працэсу навучання, якае засваення студэнтамі тэмы забяспечваецца за кошт таго, што выкарыстанне прэзентацыі дае магчымасць успрымаць інфармацыю – вербальную, візуальную, гукавую – адначасова некалькімі органамі пачуццяў, а не паслядоўна, як гэта адбываецца пры традыцыйных формах арганізацыі навучальнай дзейнасці (згаданае спалучэнне тут магчыма толькі ў самым простым выглядзе – ілюстраваным каментары). Іншымі словамі, прадстаўленне інфармацыі праз мультымедыя максімальна адпавядае натуральнаму ўспрымання рэалій навакольнага свету чалавекам. Несумненным плюсам прэзентацый з'яўляецца і тое, што яны не патрабуюць кардынальных зменаў у структуры заняткаў (лекцыйных, практычных), а таму метадыка правядзення апошніх можа як захоўваць традыцыйныя рысы і акадэмічныя формы выкладання, так і быць падпарадкаванай любой з асобна арыентаваных тэхналогій.

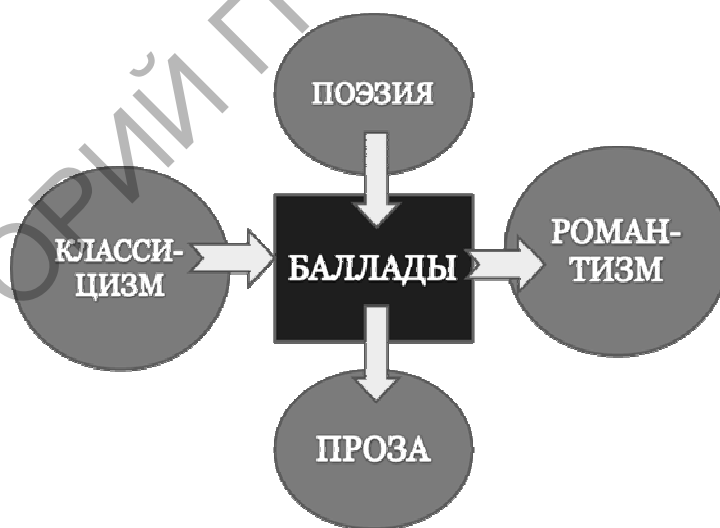
Асаблівую значнасць набываюць мультымедыя-тэхналогіі пры выкладанні курса “Гісторыя беларускай літаратуры 19 стагоддзя” (зрэшты, як і іншых перыядаў мінулага) замежным студэнтам. Па-першае, выкладанне такой дысцыпліны, як літаратура, сёння патрабуе культуралагічнага падыходу, паўнаважнасць рэалізацыя якога ўяўляецца немагчымай без прыцягнення багатага нагляднага матэрыялу. Па-другое, тлумачэнне спецыфікі грамадска-культурнай сітуацыі эпохі вымагае сістэматычнага прымянення метадаў ілюстрацыі і дэманстрацыі. І калі для беларускіх студэнтаў, якія маюць пэўнае кола ведаў па гісторыі айчыннай культуры 19 стагоддзя, прэзентацыя можа з'яўляцца толькі прыватным прыёмам у межах лекцый ці семінараў, што не адмаўляе, аднак, яго метадычнай каштоўнасці, то для замежных – мультымедыя-тэхналогія бачыцца неабходнай стратэгіяй арганізацыі заняткаў па вывучэнні літаратуры згаданага перыяду. Руплівы падбор самых разнастайных слайдаў дазваляе пераадолець аб'ектыўныя цяжкасці выкладання: адносна невысокі для спасціжэння тэкставых арыгіналаў моўны ўзровень студэнтаў (дысцыпліна выкладаецца на рускай мове), невялікі багаж іх ведаў у галіне еўрапейскай культуры, слабая арыентацыя ў спецыфіцы нацыянальнай гісторыі, беларускай міфалогіі і вуснай народнай творчасці і інш. Скажам, для адэкватнага разумення ідэйна-мастацкага зместу травесційна-бурлескнай літаратуры (паэм “Энеіда навыварат” і “Тарас на Парнасе”) неабходна мець уяўленні пра пантэон антычных бостваў; для спасціжэння спецыфікі твора “Шляхціц Завальня, або Беларусь у фантастычных апавяданнях” патрабуецца веданне істот беларускай ніжэйшай міфалогіі; асэнсаванне публіцыстыкі К. Каліноўскага вымагае непасрэднага звароту да гістарычных падзей сярэдзіны 19 стагоддзя і г.д. Урэшце, фарміраванне такіх базавых паняццяў, як “класіцызм”, “рамантызм”, “крытычны рэалізм”, немагчымае без увядзення еўрапейскага гісторыка-культурнага



кантэксту і тыпалагічных паралелей з іншымі відамі мастацтва (найперш жывапісам і архітэктурай).

Такім чынам, адной з аптымальных форм правядзення заняткаў з замежнымі студэнтамі з’яўляецца лекцыя, спалучаная з прэзентацыяй і наступным тэставым кантролем. Імкненне ўлічыць названья вышэй цяжкасці і спецыфічныя адметнасці курса “Гісторыя беларускай літаратуры 19 стагоддзя” выступае асноўным крытэрыем падбору разнастайных паводле свайго зместу слайдаў, сярод якіх можна вылучыць некалькі груп. Першую групу складаюць слайды, якія ўмоўна можна назваць стандартнымі. Гэта назва тэмы і план, гады жыцця і партрэт пісьменніка, спіс прапанаванай літаратуры. Змяшчэнне гэтай інфармацыі дазваляе не толькі якасна арганізаваць пачатак заняткаў, але і пазбегнуць магчымых памылак у напісанні нязвыклых для слыху замежных студэнтаў уласных назоўнікаў пры фіксацыі гэтых звестак у канспектах. Да стандартных адносяцца і вокладкі кніг пісьменніка, выданні розных гадоў і пераклады яго твораў на іншыя мовы. Асабліва актывізуюць увагу студэнтаў пераклады тэкстаў беларускай літаратуры на іх родную мову. Да прыкладу, заўсёды эмацыянальна ўспрымаецца замежнымі (кітайскімі) студэнтамі слайд з выявай кнігі “Пан Тадэвуш” А. Міцкевіча на кітайскай мове.

Другую групу слайдаў утвараюць табліцы і схемы, якія спрыяюць упарадкаванасці і сістэматызацыі матэрыялу, дазваляюць наглядна і лаканічна прадставіць тэкставую інфармацыю. Да прыкладу, слайд “Значэнне балад Яна Баршчэўскага”.



Малюнак 1

Прадстаўленая схема дэманструе ключавое значэнне балад у станаўленні таленту Баршчэўскага-празаіка, маркіруе этап канчатковага пераходу аўтара ад класіцыстычных да рамантычных пазіцый і ад лірыкі да эпаса – вяршыні творчасці пісьменніка.

Наступную групу складаюць тэкставыя слайды. Гэта можа быць пералік асноўных рыс мастацкай канцэпцыі твора ці індывідуальнага стылю аўтара; гістарычных падзей, якія дазваляюць больш глыбока зразумець моўна-культурную сітуацыю перыяду. Скажам, слайд “Исторические события середины 19 века” ўтрымлівае наступныя тэзісы:

- 1832 г. – закрытие Виленского университета;
- 1836 г. – преподавание в школах переведено с польского языка на русский;
- 1839 г. – запрещение униатской конфессии;
- 1840 г. – замена названия «Беларусь» на «Северо-Западный край»;
- 1840 г. – полное введение Российского законодательства.

Адной з асноўных цяжкасцей засваення курса “Гісторыя беларускай літаратуры 19 стагоддзя” з’яўляецца азнаямленне замежных студэнтаў з тэкстамі-першакрыніцамі, напісанымі на беларускай ці польскай мове. І калі, дзякуючы інтэрнэт-рэсурсам, пра недаступнасць рускамоўных перакладаў сёння гаварыць не даводзіцца, то адсутнасць тэкстаў-адаптацый, адрасаваных людзям, якія толькі некалькі гадоў (у нашым выпадку два) вывучаюць рускую мову, – даволі балючая праблема. Так, пераклад твора “Шляхціц Завальня” Я.Баршчэўскага, прадстаўлены на сайце [www.barszczewski.spb.ru](http://www.barszczewski.spb.ru), утрымлівае даволі складаныя для разумення замежнымі студэнтамі лексічныя і фразеалагічныя адзінкі. Ужо на першай старонцы сустракаем выразы “каменицы имений”, “промысел Создателя”, “речистые критики”, “ополночь летала чума”, “возвестил кару свою святотатцам” і інш. Таму ўяўляецца мэтазгодным уключаць у прэзентацыю слайды з адаптаванымі ўрыўкамі твораў, што дае магчымасць іх прачытання і абмеркавання на занятках у выглядзе рэпрадуктыўнай і / ці эўрыстычнай гутаркі.

Яшчэ адной разнавіднасцю тэкставых слайдаў з’яўляецца разнастайная інфармацыя, аб’яднаная пад рубрыкай “Слоўнік”. Найперш гэта дэфініцыі неабходных для разумення мастацкіх асаблівасцей твора літаратуразнаўчых тэрмінаў (санет, балада, гавэнда, прадмова, фарс, вадэвіль і інш.) і азначэнні спецыфічных з’яў нацыянальнай гісторыі. Да прыкладу, той факт, што героем літаратуры першай паловы 19 стагоддзя з’яўляецца шляхціц, вымагае падрабязнага тлумачэння паняцця “шляхта”.

Да чацвёртай, самай шматлікай, групы належаць слайды-выявы. Яны таксама даволі не аднародныя паводле зместу і мэт выкарыстання. Гэта рэпрадукцыі карцін мастакоў 19 стагоддзя і сучасныя фотаздымкі архітэктурных помнікаў, звязаных з біяграфіяй ці апетых у творчасці таго ці іншага пісьменніка (выгляд некаторых пабудоў, да прыкладу, можна ўявіць толькі па малюнках Н. Орды, бо да нашага часу яны не захаваліся); ілюстрацыі да мастацкіх тэкстаў; сцэны спектакляў і фільмаў, пастаўленых па творы; партрэты сваякоў і людзей, якія аказалі пэўны ўплыў на жыццё аўтара; краявіды, якія яго натхнялі (азёры Свіцязь, Нешчарда, рака

Мышанка і інш.); выявы славутых гістарычных асоб (Радзівіла Пане Каханку, Вітаўта, Міндоўга, Каспара Бекеша і інш.), якія нярэдка з'яўляюцца ў літаратуры рамантычнага характару з яе ўвагай да тэмы мінулага; родавыя гербы пісьменнікаў (а ўсе яны па паходжанні былі шляхціцамі); слайды “этнаграфічнага” тыпу, што дапамагаюць адчуць дух і каларыт эпохі; і г. д.

У адрозненне ад пералічаных выяў, якія дэманструюцца як замежным, так і беларускім студэнтам, толькі на першых разлічаны паказ ілюстрацый, якія па сваіх функцыях набліжаюцца да слоўніка. Сюды ўваходзяць “экзатычныя” для замежных студэнтаў прадстаўнікі беларускай фаўны і флоры (шчупак – балада “Калдычэўскі шчупак” Я.Чачота, пралеска – аднайменная балада А.Міцкевіча, кветка “Ясь-ды-Марыля” – балада “Дзве бярозы” Я.Баршчэўскага); беларускія народныя музычныя інструменты (дуда, дудка, скрыпка са смыкам – творчасць Ф.Багушэвіча); гістарызмы (“ямшук” – гавэнда “Паштальян” У.Сыракомлі); міфалагічныя істоты (цмок, ваўкалак, русалка – балада “Русалка-спакусніца”, твор “Шляхціц Завальня” Я.Баршчэўскага). Іншымі словамі, гэта паняцці, тэкставыя азначэнні якіх будуць не такімі эфектыўнымі ў параўнанні з выявамі.

Такім чынам, мультымедыя-тэхналогіі адкрываюць перад выкладчыкам шырокія перспектывы ў рабоце з замежнымі студэнтамі. Выкарыстанне прэзентацыі пры вывучэнні літаратуры дазваляе абмінуць шэраг праблем, невырашальных у межах традыцыйных форм арганізацыі адукацыйнага працэсу.

Я. Галло

## К ВОПРОСУ ОБ ЭЛЕКТРОННОМ ОБРАЗОВАНИИ (E-LEARNING)

Среди современных тенденций высшего образования часто встречаются понятия on-line образование, дистанционное и комбинированное образование и, естественно, e-learning. Если перевести с английского языка общепринятую и столько раз повторяющуюся аббревиатуру ИКТ как информационные и коммуникативные технологии, имея в виду методы, приёмы и способы сбора, сохранения, оценивания, выбора, распределения и современной поставки нужной информации согласно требуемой формы и качества, то можно воспользоваться известным высказыванием Я. А. Коменского из Великой дидактики, «... что касается лучшего будущего, мы не можем остановиться в поисках средств, которые это могут позволить». Это, на первый взгляд, смелое сравнение с e-learning сбывается.

В настоящее время существует обучение и в традиционном понимании. В последние годы, хотя и заметно модернизируется введение новых

технологических средств, но, в сущности, способ обучения не меняется. Существует педагог, который читает лекции студентам. Студенты в качестве основного источника знаний пользуются пособиями и другой рекомендуемой литературой. Вместо доски и мела используются современные интерактивные доски на основе компьютерных технологий. Введение современных электронных и коммуникативных средств в процесс обучения приводит к более комплексному решению проблемы.

Развитие и возможности ИКТ способствуют обучению иностранным языкам в электронной форме. Использование многих способов представления учебного материала вместе с аудио- и видео записями, анимациями и т. д. делает возможным дистанционное обучение и в этой области, где, в общем, роль преподавателя считают трудоёмкой.

Как во всех современных (иногда и модных) тенденциях обучения, можно встретиться с терминологической несогласованностью понятий и в области e-learning. По этой причине G. Švejda и Z. Horváthová [7, с.17] приводят дефиницию e-learning в более широком смысле слова и определяют e-learning как «применение новых мультимедийных технологий и Интернета в образовании с целью повышения его качества и открытия доступа к реальным данным, услугам, обмену информацией, сотрудничеству». В более узком смысле слова G. Švejda и Z. Horváthová [там же, с.17] понимают e-learning как «образование, поддерживаемое современными технологиями и осуществляемое посредством компьютерных сетей – Интранета и Интернета». Общепринятую дефиницию e-learning формулируют G. Švejda и Z. Horváthová [там же, с.17] следующим образом: «E-learning понимаем как мультимедийную поддержку процесса обучения с использованием современных информационных средств и коммуникативных технологий, который, как правило, осуществляется посредством компьютерных сетей. Его основной задачей является свободный доступ к образованию во времени и в пространстве».

И. Турек [8, с.413] делает вывод, что «термин *e-learning* (слово, происходящее из английского языка, буквальный перевод которого означает электронное обучение) является очень широким и неустойчивым понятием (введён только в 1999 году), который не имеет даже общепризнанного словацкого эквивалента. И в словацких текстах, в связи с образованием при помощи ИКТ (электронным образованием), наиболее часто используется термин *e-learning* и только иногда – термины *электронное образование*, *электронное учение* или *электронное обучение*».

И. Турек [8, с.413] указывает, что «пока нет единого мнения, что, собственно, есть e-learning. Согласно одному мнению, e-learning является любым образованием, при котором используются ИКТ (это может быть только на определенном этапе процесса обучения, например, когда преподаватель составляет программу предмета при помощи компьютера и делает его доступным для учащихся, или если учащихся экзаменуют при помощи дидактических тестов, представленных посредством компьютера,

или когда учащийся учится по тексту, который находится на компакт-диске (CD) или в Интернете». Другую крайность, по мнению И. Турека, «представляют собой мнения, которые под e-learning понимают только обучение, выполняемое посредством компьютера. Кроме того, хаос вносят мобильные телефоны, которые дают возможность полноценно общаться посредством Интернета и, если они используются для обучения, которое называется m-learning (мобильное обучение)».

Интенсивное развитие e-learning началось только после 1993 года, когда очень быстро и массово расширилось использование Интернета. Пока e-learning используется во многих случаях только в самом простом виде, например, учащийся обучается на учебном тексте, который преподаватель записал на CD-ROM, передал посредством электронной почты (e-mail), веба, или учащийся находит учебный текст на веб-страницах при помощи некоторых поисковиков (напр. GOOGLE). Таким же способом учащийся может дойти до дидактических тестов (автотестов), при помощи которых проверяется степень усвоения учебного материала. Учащиеся также ищут на веб-страницах значения физических, статистических и других величин, проводят трудоёмкие расчёты, повторяют (закрепляют) учебный материал, диагностируют себя (выясняют, например, свой учебный стиль) и т. п. E-learning включает не только мультимедийные курсы обучения (содержащие и анимации, и симуляции), но также и многие дополнительные средства – электронную почту, электронные демонстрации и документы, видео, чат, электронные конференции, виртуальные классы и дискуссионные группы. E-learning можно осуществлять не только посредством компьютера (Интернет, Интранет, CD ROM, DVD и т. п.), но и при помощи телевидения (кабельного, спутникового), аудио- и видео записей, даже мобильных телефонов.

Как правило, формы *e-learning* G. Švejda и Z. Horváthová разделяют на off-line e-learning и on-line e-learning [7, с.17]. *Off-line e-learning* не требует того, чтобы компьютер был подключён к другому компьютеру посредством компьютерной сети. Учебный материал студенты получают посредством CD-ROM или же DVD-ROM. Это очень удобный способ для домашней подготовки, так как в Европе постоянно используют форму обучения, названную Blended learning (смешанное обучение), когда происходит соединение традиционного обучения с мультимедийной поддержкой e-learning.

*On-line e-learning* известен в двух формах. *Синхронная* форма предполагает постоянное подключение и общение в реальном времени. Все учащиеся находятся в контакте с учителем, который управляет учебным процессом, или между собой в точно определённое время посредством электронных конференций: видеоконференций, телеконференций, интерактивной коммуникации или чатов. Однако могут также общаться из разных мест. В настоящее время сюда входит и одна из самых развитых форм e-learning – виртуальный класс, размещение в виртуальном пространстве –

в Интернете, которое даёт возможность учащимся встречаться, говорить, общаться и сотрудничать. При *асинхронной* форме участники общаются не в реальном времени, а посредством дискуссионных форумов. В сущности, дело касается самостоятельного обучения посредством ИКТ. Учащиеся могут не заниматься одновременно, а сами выбирают время учёбы, занимаются посредством веб-страниц, CD-ROM, DVD, общаются с помощью электронной почты, веб-блогов.

Несмотря на выше приведённые преимущества обучения посредством e-learning, можно выделить и недостатки. G. Švejda и Z. Horváthová [7, с. 17] приводят следующие:

- зависимость от подходящего технологического обеспечения;
- недостаточная стандартизация как технологической системы, так, главным образом, системы образования;
- требования образовательных стандартов (историческая связь с программированным обучением);
- его использование без рационального обдумывания пропорциональности или, в отдельных случаях, подходящих условий использования.

Согласно И. Туреку «если обучение посредством e-learning должно быть действительно эффективным, нужно им управлять: планировать ход обучения; подготовить учебный материал; регистрировать учащихся и их учебные результаты; представить учебный материал таким образом, чтобы он был интересным и подходил разным учебным стилям учащихся (напр., использование графики, анимации, аудио- и видео техники); обеспечить общение между учащимся и учителем и между самими учащимися; мотивировать и активизировать учащихся (например, давать им задания и оценить их решения), наблюдать за деятельностью учащихся; обеспечивать обратную связь и оценивать учащихся; установить уровень прав доступа для разных категорий потребителей и гарантировать безопасность; сдавать в архив курсы и т. п.» [8, с. 415].

Таким образом, мы живём во время постоянных и быстрых изменений, и прогнозирование будущего развития (в области образования в том числе) является в этих условиях очень сложным, противоречивым и спорным. Однако, специалисты, интересующиеся прогнозированием и футурологией, сходятся в одном: информационные и коммуникационные технологии (ИКТ) играют решающую роль в развитии общества. Уже и сейчас ИКТ влияют на нашу повседневную жизнь. Эти технологии могут совершить в образовании революцию, сравнимую с революцией образования, вызванной изобретением книгопечатания. ИКТ развиваются так быстро и турбулентно, что их современное использование в образовании представляет собой только маленький фрагмент их потенциальных возможностей. В прошлом люди приходили за знаниями, которые были сосредоточены, главным образом, в вузах. В будущем знания сами придут к людям, независимо от того, где они находятся. Учебные заведения, которые не смогут с этим справиться, будут иметь большие проблемы с целью выживания. Будет

существовать высокая конкуренция между учебными заведениями и разными коммерческими фирмами, обеспечивающими высококачественное, аккредитованное образование посредством ИКТ – e-learning.

Для качественного составления программ e-learning нужно иметь как компьютерную компетенцию (умение работать с компьютером), так и методическую компетенцию (владеть методикой преподавания). Это соединение компетенций не является частым. E-learning не является хобби. Многие информатики (а также инженеры, математики, естествоиспытатели и т.п.) являются настоящими специалистами, даже фанатами компьютеров, но у них нет требуемого педагогического образования. Сотрудничество этих двух групп специалистов также недостаточно, и в итоге, качество материалов e-learning оставляет желать лучшего. В настоящее время мало кто знает, что в рамках e-learning могла бы быть очень хорошо использована теория программного обучения, которая в Словакии, уже практически не включается в преподавательские учебные программы.

#### СПИСОК ЛИТЕРАТУРЫ

1 Игнатова, И. Б. Обучение русскому языку иностранных студентов-русистов в русле концепта «языковая личность»/ И. Б. Игнатова, Л. Г. Петрова, Т. В. Самосенкова. – Белгород: Изд-во БелГУ, 2003. – 304 с.

2 Kalhous, Z. OBST. O. a kol.: Školní didaktika/ Z. Kalhous. – Praha: Portál, 2009. – 447 s.

3 Крючкова, Л. С., Мощинская, Н. В.: Практическая методика обучения русскому языку как иностранному/ Л. С. Крючкова, Н. В. Мощинская. – Москва: Флинта, Наука, 2009. – 480 с.

4 Кукушин, В. С.: Дидактика (теория обучения). Учебное пособие/ В. С. Кукушин. – Изд. 2-е, перераб. и доп. – Ростов н/Д: Издательский центр «МарТ»; Феникс, 2010. – 366 с.

5 Petlák, E. Všeobecná didaktika/ E. Petlák. – Bratislava: Iris, 2004. – 311 s.

6 Skalková, J.: Obecná didaktika. 2. rozšířené a aktualizované vydání. Praha: Grada, 2007. – 328 s.

7 Švejda, G. a kol.: Vybrané kapitoly z tvorby e-learningových kurzov. – 1. vydanie. – Nitra: UKF, 2006. – S. 17.

8 Turek, I.: Didaktika/ I. Turek. – Bratislava: Iura Edition, 2008. – S. 413, 415.

А. И. Головня

#### **КОМПЬЮТЕРНЫЕ ЭКСПЕРТНЫЕ СИСТЕМЫ В ОБУЧЕНИИ РКИ**

В современном мире развитие информационных технологий, их стремительное внедрение во все сферы жизни общества оказывает все большее влияние на образование, определяет необходимость модернизации учебного процесса в школах и в вузах, особенно при преподавании РКИ. Все большая роль отводится контролируемой самостоятельной работе студентов дневной формы обучения. Для того чтобы самостоятельная работа студентов была более продуктивной, важно создавать хорошие компьютерные учебники, в которых бы сочетались краткие комментарии

по теории и приводились бы задания в виде тестов или тестовых заданий. С помощью таких электронных учебников, в которых существовала бы шкала оценки знаний относительно каждого выполненного упражнения для каждого студента, преподавателю легко было бы определить, что студент освоил хорошо и на что надо обратить особое внимание при чтении лекций и проведении практических занятий по РКИ. Над созданием таких комплексных электронных учебно-методических пособий, прежде всего, должны работать сами преподаватели при тесном взаимодействии с хорошим программистом и коллективом, который может грамотно обеспечить техническую часть обучающей экспертной системы.

В БГУ на филологическом факультете создана такая обучающая и контролирующая экспертная система в виде компьютерного учебника по курсу «Современный русский язык». Она представляет собой электронную версию пособия «Русский язык в тестах и комментариях», созданного на кафедре русского языка филологического факультета БГУ под руководством доктора филологических наук, профессора И. С. Ровдо. Программное обеспечение предоставлено инженером-программистом Я. В. Карповым. Техническое исполнение и корректировка теоретического и практического материала подготовлены коллективом бывшей НИЛ ТиПЛ под руководством доктора филологических наук, профессора В. А. Карпова.

В процессе выполнения данной работы были решены следующие задачи:

1) определены проблемы взаимосвязей языковых объектов разных уровней, что позволит на занятиях с помощью компьютерного пособия усваивать именно целостную систему, а не получать разрозненные знания по отдельным разделам языка, что с методической точки зрения очень важно для получения глубоких знаний;

2) создан компьютерный учебно-методический комплекс по всем курсам современного русского языка, кроме словообразования.

Основные выводы, относящиеся к проделанной работе, касаются двух важных моментов любого исследования: научной и практической значимости.

Научное значение работы заключается в подходе, который обеспечил создание действующей системы, благодаря теоретической полноте и непротиворечивости, на базе которой был создан компьютерный учебно-методический комплекс с использованием в упражнениях тестовых заданий, направленных на интенсификацию процесса обучения студентов. В ходе разработки проекта были подтверждены результаты, гарантируемые разработанной в рамках теоретических и прикладных задач системно-языковой методологией.

Данная разработка базируется на системном подходе, предполагающем симметрично-асимметричное представление языка, позволяющее определить как теоретическую полноту проблемы, так и ее формализацию для создания компьютерной обучающей системы. В основе этого метода



лежит математическое доказательство системности объектов любого рода. Базирующееся на последовательном приложении к языку общей теории систем или ОТСУ (вариант ОТС, развиваемый академиком Ю. А. Урманцевым), а в области лингвистики ОТСУ использовалась профессором В. А. Карповым для подтверждения системности языка, в котором принципы симметрии-асимметрии используются в их теоретической полноте. Эти принципы являются неспецифическими для объектов любого рода, поэтому применимы при описании систем разных предметных областей.

Практическая значимость обучающей экспертной системы связывается с тем, что данная разработка выполнена таким образом, что ее результаты возможно использовать при разработке и внедрении аналогичных пособий в системе высшей школы республики. Предполагается, что с минимальными отклонениями-поправками полученные результаты будут пригодны в качестве системных образцов для создания аналогичных электронных обучающих учебно-методических пособий по болгарскому, польскому и другим языкам.

Большой плюс при использовании обучающих электронных комплексов состоит в том, что они позволяют иностранному студенту реально оценить уровень своих знаний, овладеть практическими навыками в освоении курса любого предмета и закрепить полученные знания. Работа преподавателя над новыми вариантами тестовых заданий будет содействовать высокой степени оптимизации знаний, экономии учебного времени, эффективному усвоению изучаемого материала. Грамотно составленные тестовые задания помогут проводить практические занятия, одновременно проверяя степень подготовленности большой группы студентов, что сэкономит время преподавателя для создания новых учебных пособий.

Разработка может быть внедрена в учебный процесс, распространена в методкомиссиях по созданию учебников русского языка и других дисциплин.

Полученные результаты являются достаточным основанием для начала разработки подобных пособий в БГУ и в других вузах республики Беларусь. Целостных разработок таких курсов (объем и связность) в республике Беларусь пока мало.

Главным результатом данной работы стало создание компьютерных учебно-методических комплексов почти по всем курсам современного русского языка. Результаты работы в виде компьютерного учебного пособия по современному русскому языку реализованы на машинных носителях в компьютерных классах филологического факультета БГУ.

Само компьютерное учебное пособие представляет собой четыре независимых друг от друга программных продукта («Фонетика»; «Морфология»; «Лексика», «Синтаксис»), которые устанавливаются и работают автономно друг от друга. Оно предназначено для выработки умений и навыков, закрепления знаний по всем разделам (фонетика, лексика, морфология, синтаксис) вузовского курса «Современный русский язык».

Продолжительность одного урока составляет 80 минут и может меняться в зависимости от уровня подготовки студентов. Правильность ответов на предложенные задания проверяется автоматически сразу после их выполнения, результаты проверки отображаются на экране сразу с правильными ответами, к которым при необходимости даются пояснения или развернутый комментарий.

Оценка уровня подготовленности студента осуществляется с помощью тестовых заданий. Каждое упражнение содержит 25 тестовых заданий. Прежде чем приступить к квалификационному экзамену, студент может реально оценить свой уровень подготовки, а также практически проработать наиболее трудные для него темы.

Использование комплекса электронных тестовых заданий позволит студентам самостоятельно овладеть практическими навыками в освоении курса современного русского языка и закрепить полученные знания, а также будет содействовать высокой степени оптимизации занятий, экономии учебного времени, эффективному усвоению изучаемого материала. Преподаватель может проводить практические занятия, одновременно проверяя степень подготовленности целой группы студентов, так как работа оценивается в процентном отношении к числу допущенных ошибок.

Тестовый характер заданий, предполагающий ситуацию выбора, позволяет, во-первых, установить место языкового явления в системном ряду. Эта системность является важной предпосылкой повторения и обобщения материала. Во-вторых, подбор дидактического материала осуществлялся авторами с учетом предсказуемости лингвистических трудностей, увидеть и преодолеть которые помогут тестовые задания.

Каждая тема (блок упражнений или отдельное упражнение) предваряется комментариями: минимум теоретических сведений по данной теме, образец рассуждения, алгоритм действий, анализ наиболее трудных случаев. В разделах представлены следующие тестовые задания: обучающие, обобщающие и контролирующие.

Тестовая система заданий обеспечивает разнообразие форм работы: задания могут выполняться как в режиме тренировки, так и в режиме контроля, под руководством преподавателя и самостоятельно, коллективно и индивидуально, как при изучении соответствующих тем, так и при повторении материала.

Ответы к заданиям позволяют использовать пособие и для самоподготовки и самоконтроля. Компьютерное учебное пособие разбито на уроки. Каждый урок состоит из двух частей: 1) теоретического минимума по данной теме; 2) упражнений или тестовых заданий.

Процесс изучения языковых единиц и категорий предполагает обязательное движение от осмысления теоретического материала к его практическому использованию. Каждая тема (блок упражнений или отдельное упражнение) предваряется комментариями: минимум теоретических сведений по данной теме, образец рассуждения, алгоритм действий, анализ

наиболее трудных случаев, кроме того обращается внимание на спорные и сложные случаи квалификации языковых явлений. Теоретическая часть предназначена для того, чтобы освежить в памяти материал, если он был ранее прослушан на лекции. В том случае, если материал является новым для студента, то наличие теории позволяет ему освоить основные моменты изучаемой темы самостоятельно.

Упражнения предназначены для контроля и закрепления полученных знаний. Система тестовых заданий составлена в соответствии с основным дидактическим принципом (от простого к сложному) и нацелена на поступательное развитие умений и навыков грамматического анализа, на сознательное усвоение и закрепление правил и норм правописания, закрепление фонетических, лексических, морфологических и синтаксических языковых норм.

Уроки желательно проходить по порядку, но можно возвращаться к предыдущим урокам для закрепления и проверки знаний и даже забегать вперед. Программа обеспечивает свободный доступ к любой теме.

Для подготовки студентов по русскому языку в основном используются два типа упражнений: 1) ввести при помощи клавиатуры нужное слово или нужную букву, а также вписать соответствующую вашему выбору цифру в синий квадрат; 2) выбрать правильный вариант ответа и вписать соответствующую цифру в синий квадрат. Тренировочные упражнения (тестовые задания) снабжены ответами к заданиям для самоконтроля (ключами-ответами).

Созданные компьютерные учебно-методические комплексы по всем курсам современного русского языка могут быть использованы при преподавании русского языка в школе и вузе, а также при преподавании русского языка как иностранного.

В настоящее время разработаны и внедрены в учебный процесс учебно-методические комплексы по курсу фонетики белорусского языка, компьютерный учебник по украинскому и английскому языкам. Подобного рода исследования на основе созданной методики следует проводить и на базе других языков. В дальнейшем предполагается разрабатывать конкретные компьютерные учебно-методические комплексы для обучения французскому, итальянскому, испанскому языкам.

Многие преподаватели филологического факультета БГУ используют электронные тесты, созданные в eUniversity, для промежуточных зачетов по определенной теме, а также тесты, в которых есть вопросы по всему пройденному материалу лекционного курса.

Активно стали использоваться презентации при чтении лекционных курсов.

Все, что касалось плюсов, уже указано. Большим минусом при использовании комплексных компьютерных учебных пособий с использованием тестовых заданий при подготовке управляемой контролируемой самостоятельной работой (УКСР) является недостаточная материальная база. На факультете работает четыре компьютерных класса. Нагрузка на

эти классы огромна. Там проводятся практические занятия по информационным технологиям и тестирование по другим дисциплинам, и заниматься самостоятельной работой с использованием компьютерных учебников студентам фактически невозможно. Необходимо увеличить техническую базу компьютерных классов и создать хотя бы два класса для управляемой контролируемой работы студентов.

О. Е. Елисеева

## **СЕТЬ ИНТЕРНЕТ КАК ИНСТРУМЕНТ И СРЕДСТВО ИЗУЧЕНИЯ ПРИКЛАДНОЙ И КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ**

Развитие и повсеместное внедрение современных информационных технологий требует по-новому рассматривать практически любую профессию. Профессия филолога – не исключение. Очевидная на сегодняшний день потребность использования информационных технологий в работе филолога, в свою очередь, предполагает поиск новых подходов к процессу подготовки молодых специалистов в данной отрасли.

Следует также заметить, что современный уровень развития общества и производственных отношений требует уточнения понятия «филология» и связанного с ним наименования специальности «филолог». Так, например, в сети Интернет можно найти различные трактовки к пониманию филологии и профессии филолога в сравнении, например, с лингвистикой [1, 2]. В работе [3] в качестве филологических предлагаются следующие актуальные для современного общества профессии: переводчик; специалист по связям с общественностью; референт; пресс-секретарь; имиджмейкер; копирайтер; спичрайтер. Там же указано, что «К этим профессиям примыкает и профессия менеджера, от которого требуется именно умение эффективно общаться». Все перечисленные профессии, в свою очередь, немислимы без хороших знаний современных информационных технологий.

В данном сообщении предлагается рассмотрение сети Интернет как инструмента и средства изучения прикладной и компьютерной лингвистики, а также объекта проведения всевозможных лингвистических исследований. При этом следует уточнить, что автор придерживается той точки зрения, что филология является более общим и широким понятием по отношению к лингвистике (языкознанию). Лингвистика, таким образом, рассматривается как часть (раздел) филологии. Такой подход был избран в результате анализа определений указанных понятий, взятых из разных словарей: Большого энциклопедического словаря, Толкового словаря русского языка под ред. Д. Н. Ушакова, Малого энциклопедического словаря Ф. А. Брокгауза и И. А. Ефрона, Большой советской энциклопедии, Психологической энциклопедии, Социологического словаря, Толкового словаря

живого великорусского языка В. И. Даля, Википедии – свободной энциклопедии, а также на основании изучения «терминологических споров», размещенных в сети Интернет.

В работе не проводится чёткой границы между терминами «прикладная лингвистика» и «компьютерная лингвистика», но принимается во внимание тот факт, что по существу компьютерная лингвистика является разделом прикладной лингвистики [4, 5]. Исследование, представленное в данном сообщении, базируется на обоих этих понятиях, т. к. по мнению автора, овладение профессией филолога обязательно должно предусматривать изучение прикладной и компьютерной лингвистики.

Ниже перечислены и охарактеризованы ресурсы сети Интернет, которые предлагается использовать в учебном процессе и научных исследованиях.

Словари и энциклопедии. Сеть Интернет в настоящее время содержит электронные версии практически всех наиболее популярных словарей и энциклопедий, неизбежно требующихся в работе филолога. К преимуществам таких словарей по сравнению с их бумажными аналогами относятся: открытый коллективный доступ; расширенные возможности быстрого поиска по словарю; возможность оперативного пополнения и исправления ошибок и др. В качестве примеров таких словарей можно привести [4], а также такие ресурсы, как Мир словарей (<http://mirslovari.com>), Яндекс. Словари (<http://slovari.yandex.ru>), многоязычные словари для переводчиков и многие другие [6]. В качестве учебно-методических задач использования словарей в сети Интернет следует выделить формирование у студентов:

- знаний классификации словарей;
- знаний адресов основных словарей, а также особенностей этих словарей и сфер их практического использования;
- умений искать словари и определения терминов в словарях в словарях;
- умений создавать и пополнять электронные словари;
- умений формулировать задачи поиска и анализа словарных статей для реализации соответствующих программных решений с целью развития электронных словарей и расширения их возможностей.

Электронные библиотеки, аудио- и видеоматериалы. Наличие в сети Интернет достаточно больших объёмов текстов литературных произведений открывает дополнительные возможности их изучения и исследования. Например, на основе электронного текста произведения намного быстрее и эффективнее можно осуществить статистический анализ текста, произвести разметку и подготовить его для включения в языковой корпус и пр. Очевидно, что помимо собственно ресурсов сети Интернет в данном случае от специалиста требуется овладение другими компьютерными средствами. Доступ к аудиокнигам может помочь молодому специалисту не только сэкономить время на прочтение того или иного произведения, но и сформировать у него навыки художественного чтения, выработки дикции

и артистичности на примерах работы профессиональных дикторов. Кроме того, аудиокнига является незаменимым средством при изучении иностранного языка, а также способствует повышению интереса к процессу ознакомления с литературными произведениями. Эти же задачи решают и экранизации литературных произведений. Аудиоинформация может иметь и историческую ценность, если это, например, запись голоса известного автора, читающего собственное произведение. Видео-запись (а также фотоснимки) может содержать также оттиски старинных книг, недоступных массовому чтению ввиду особенностей их хранения в мировых библиотеках. Так, в 2009 году был запущен международный проект по созданию Всемирной виртуальной библиотеки, одной из задач которой является сбор таких аудио-, видео- и фотоматериалов. В качестве учебно-методических задач работы с электронными библиотеками можно выделить задачи, аналогичные перечисленным выше.

Специализированные ресурсы по филологии, лингвистике, компьютерной лингвистике, искусственному интеллекту и пр. Постоянное развитие и упрощение инструментов разработки веб-ресурсов даёт возможность специалистам различной квалификации создавать свои специализированные ресурсы. Их несомненным преимуществом перед «бумажными» аналогами является актуальность представленной в них информации и возможность участвовать в различных научно-практических проектах и обсуждениях, проводимых на базе таких ресурсов. Кроме того, такие ресурсы, как правило, содержат примеры работающих программных приложений, используемых для решения разного рода лингвистических задач. С помощью каталога Яндекс, в специальном разделе «Филология» можно найти 24 примера специализированных ресурсов (по состоянию на 9 сентября 2010 года) – <http://yasa.yandex.ru/yasa/cat/Science/Sciences/Humanities/Philology/>. В учебно-методическом плане специализированные ресурсы позволяют формировать у студентов:

- более компетентные знания в соответствующих областях, включающие самые последние открытия и результаты исследований;
- знания современных тенденций развития того или иного направления;
- навыки обсуждения различных актуальных вопросов специальности;
- навыки использования прикладных программ лингвистического назначения.

Гипертекстовая технология. Изучение филологами языка разметки гипертекста позволяет на практике усвоить технологию разметки текстов вообще, а также приобрести навыки создания специализированных ресурсов, пригодных для размещения в сети Интернет. На базе гипертекстовой технологии в настоящее время активно развиваются проекты в области семантического веб (semantic web). Проекты указанной тематики логичным образом входят в зону рассмотрения филологов, несмотря на их изначальную

техническую ориентацию и участие в этих проектах преимущественно ИТ-специалистов.

Интернет является огромным хранилищем ценной информации, а также удобной и доступной площадкой для обмена информацией, проведения различных мероприятий (интернет-конференций) и совместных исследований. Освоение будущими филологами всех средств коммуникации сети Интернет (электронной почты, чатов, IP-телефонии, блогостроения и блоговодства, социальных сетей и др.) позволяет им стать специалистами, востребованными в самых разных сферах профессиональной деятельности.

Системы автоматического перевода. Наличие в интернете специализированных сервисов автоматического перевода в значительной степени облегчает работу специалистов разных отраслей. Будущим филологам следует знать и уметь использовать такие ресурсы. Кроме того, в учебно-методическом плане такие ресурсы являются очень хорошей площадкой для исследования на практике возможностей современных систем машинного перевода (МП), анализа лингвистических проблем МП, формулировки задач МП и пр.

Поисковые системы. Сеть Интернет постоянно меняется, появляются новые ресурсы, сервисы и возможности. Каждому пользователю Сети известны основные поисковые системы и их возможности. Однако не каждому известно то, что существуют специализированные поисковые системы, например, поиск книг, научной информации [6], глубинный поиск, семантический поиск. В реализации поисковых систем участвуют лингвисты, которые обеспечивают возможности учёта морфологии, синтаксиса в процессе анализа поисковых запросов. Развитые поисковые системы содержат также специализированные языки запросов, умелое использование которых обеспечивает лучшие результаты поиска, с одной стороны, и приобретение навыков формирования языковых запросов – с другой.

Системы дистанционного обучения, компьютерные обучающие системы и т.п. Использование и изучение систем указанного класса, освоение технологии создания таких систем на базе специальных инструментальных средств, доступных в сети Интернет, способствует приобретению молодыми специалистами навыков, которые в перспективе позволят им стать современными педагогами.

В завершение следует отметить, что в данном сообщении рассмотрены возможности лишь одной современной информационной технологии – сети Интернет. Аналогичного анализа заслуживают и многие другие компьютерные технологии. Если рассматривать вопрос грамотного использования современных компьютерных технологий в работе филолога, то имеет смысл отдельно рассматривать компьютерное направление в филологии и говорить о появлении нового термина «компьютерная филология». Именно этот термин может включить в себя все аспекты использования

вычислительной техники в различных сферах деятельности филолога и таким образом преодолеть проблему некорректного использования термина «компьютерная лингвистика», что в настоящее время происходит достаточно часто.

#### СПИСОК ЛИТЕРАТУРЫ

1 Спор лингвиста и филолога [Электронный ресурс]. – Режим доступа: – <http://my.mail.ru/community/philology/3103F1FB96557B52.html> – Дата доступа: 27.10.2010.

2 Чем отличаются филолог и лингвист? [Электронный ресурс]. – Режим доступа: <http://otvety.google.ru/otvety/thread?tid=730800dd483f877b> – Дата доступа: 27.10.2010.

3 Ким, И. Е. Зачем обществу нужны филологи // Новая университетская жизнь. Периодическое издание Сибирского федерального университета [Электронный ресурс]. – Режим доступа: <http://gazeta.sfu-kras.ru/node/303> – Дата доступа: 27.10.2010.

4 Онлайн Энциклопедия Кругосвет [Электронный ресурс]. – Режим доступа: <http://www.krugosvet.ru> – Дата доступа: 27.10.2010.

5 Соснина, Е. П. О проблеме определения границ предметной области компьютерной лингвистики / Е. П. Соснина // [Электронный ресурс]. – Режим доступа: [http://ling.ulstu.ru/linguistics/resources/literature/articles/delimitation\\_computer\\_linguistics](http://ling.ulstu.ru/linguistics/resources/literature/articles/delimitation_computer_linguistics) – Дата доступа: 27.10.2010.

6 Елисеева, О. Е. Использование ресурсов сети Интернет в обучении филологов // IV Чтения, посвященные 70-летию со дня рождения профессора В. А. Карпова (Минск, БГУ, 19–20 марта 2010 г.): сб. науч. ст.: в 2 ч. Ч. 1 / редкол.: А. И. Головня [и др.]. – Минск: РИВШ, 2010. – С. 80–86.

С. Б. Кураш, Н. Н. Егоров

### **СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ «КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА» В СИСТЕМЕ ПОДГОТОВКИ БУДУЩИХ УЧИТЕЛЕЙ РУССКОГО ЯЗЫКА И ЛИТЕРАТУРЫ**

С 2008–2009 учебного года в типовой учебный план педагогической специальности 1-02 03 04-03 «Русский язык и литература» (с дополнительной специальностью «История») введена новая дисциплина – «Компьютерная лингвистика» (3-й курс, 5-й семестр), что является одним из наиболее примечательных нововведений нового учебного плана. Несмотря на то, что данная учебная дисциплина включена в блок факультативных дисциплин, её значение и роль в системе подготовки будущих учителей-русистов в педагогическом университете на данном этапе весьма велика. В жизни современного общества трудно переоценить ту роль, которую играют автоматизированные информационные технологии. С течением времени их значение непрерывно будет возрастать. Естественный язык, как известно, является инструментом мышления. Он также универсальное средство общения между людьми – средство восприятия, накопления, хранения, обработки и передачи информации. Проблемы использования естественного языка и функционирования естественно-языковых текстов в системах



автоматической обработки информации – основное проблемное поле компьютерной лингвистики. Эта наука возникла сравнительно недавно – на рубеже 50–60-х годов прошлого столетия. За прошедшие полвека в области компьютерной лингвистики были получены значительные научные и практические результаты: созданы системы машинного перевода текстов, системы автоматизированного поиска информации в текстах, системы автоматического анализа и синтеза устной речи и многое другое.

Данная дисциплина к настоящему моменту достаточно прочно обосновалась в учебных планах ряда специальностей филологического профиля. Думается, что для подготовки современного педагога-филолога (в частности, по указанной выше специальности) её введение также выглядит достаточно обоснованным и своевременным, учитывая обозначенные выше тенденции в современном мире.

Согласно типовому учебному плану, разработанному в БГПУ им. М. Танка, на изучение дисциплины отводится 26 аудиторных академических часов, из которых (уже на основании рабочего учебного плана филологического факультета Мозырского государственного педагогического университета им. И. П. Шамякина) 8 часов отведено на лекционные занятия, 18 часов – на лабораторные занятия, проводимые в специализированной аудитории, оснащенной современными персональными компьютерами и программным обеспечением в соответствии с тематикой изучаемого материала. Число рабочих мест в аудитории должно быть таким, чтобы обеспечивалась индивидуальная работа студента на отдельном персональном компьютере.

Цель дисциплины – сформировать у студентов системное представление о приложимости и развитии компьютерных технологий в лингвистических исследованиях.

Для достижения поставленной цели в процессе изучения дисциплины предполагается решить следующие задачи: определить наиболее существенные лингвистические направления, в которых активно задействуются компьютерные технологии; выявить значимые аспекты автоматической обработки естественного языка естественно-языкового текста в программах информационного поиска, машинного перевода, атрибуции текста, в лингвистических базах данных и т. п.; дать студентам представление о работе с компьютерными программами, которые так или иначе задействуют лингвистическую информацию.

В результате изучения курса студенты должны знать состав и структуру лингвистического обеспечения современных ПЭВМ; основные характеристики коммуникативного взаимодействия человека и компьютера; состав и структуру современных информационных и поисковых систем; методы автоматического индексирования и реферирования естественно-языковой информации; уметь использовать компоненты лингвистического обеспечения ПЭВМ, программные и информационные средства компьютерной лингвистики в практической деятельности; использовать информационные и поисковые системы; иметь навыки

использования естественного языка для формирования (выражения) информационных потребностей для работы в сети Интернет.

Содержание разделов курса базируется и в свою очередь уточняет и дополняет разделы и темы таких дисциплин, как «Основы высшей математики», «Основы информационных технологий», «Введение в языкознание», «Основы интеллектуальной собственности», готовит студентов к изучению дисциплины «Общее языкознание». Параллельно с изучением компьютерной лингвистики, в дополнение к данной дисциплине, в рабочий учебный план факультета (на той же специальности, в том же семестре) введён также спецкурс «Информационные технологии в гуманитарном образовании», имеющий преимущественно методический характер.

На лекциях по компьютерной лингвистике нами освещаются теоретические проблемы, связанные с лингвистическим обеспечением взаимодействия человека и компьютера, а также с компьютерной обработкой естественного языка и естественно-языковых текстов.

На лабораторных занятиях закрепляются проработанные теоретические проблемы и вырабатываются практические навыки в области компьютерной обработки естественно-языковых данных. При подготовке к лабораторным занятиям студенты работают над лекционным материалом, а также читают учебные пособия по данной проблеме.

Примерный тематический план лекционных занятий и содержание лабораторного практикума представлены в нижеследующих таблицах.

В ходе преподавания данной дисциплины акцент делается на приобретении студентами конкретных практических навыков в области компьютерной обработки естественного языка и естественно-языковых текстов. Есть основания полагать, что в результате изучения курса «Компьютерная лингвистика» студенты, будущие учителя русского языка и литературы, будут вооружены необходимыми базовыми умениями в области применения компьютерного программного обеспечения для целей лингвистических исследований, будут способны применять полученные навыки в дальнейшей исследовательской и творческой работе.

В ходе преподавания данной дисциплины акцент делается на приобретении студентами конкретных практических навыков в области компьютерной обработки естественного языка и естественно-языковых текстов. Есть основания полагать, что в результате изучения курса «Компьютерная лингвистика» студенты, будущие учителя русского языка и литературы, будут вооружены необходимыми базовыми умениями в области применения компьютерного программного обеспечения для целей лингвистических исследований, будут способны применять полученные навыки в дальнейшей исследовательской и творческой работе.

Таблица 1

№ п/п	Наименование тем лекционного курса	Часы (8)	Содержание лекций
1	Компьютерная лингвистика и её связь с другими науками	2	1.1. Вопрос о новом месте лингвистики в современном мире. Становление и развитие прикладного (компьютерного) направления в лингвистике. 1.2. Компьютерная лингвистика на современном этапе. 1.3. Перспективы развития компьютерных технологий в лингвистике.
2	Программно-лингвистическое обеспечение диалога человека с компьютером	2	2.1. Понятие информации. Виды информации. Способы ее передачи и накопления. 2.2. Понятия трансляции, компиляции и интерпретации. 2.3. Пути решения проблемы коммуникации: создание диалоговых систем и систем "обработки естественного языка". 2.4. Информационно-поисковые системы. Поиск информации как лингвистическая проблема. Современные ИПС (Google, Яндекс, Rambler и др.). Возможности расширенного поиска в ИПС. Синтаксис запросов. 2.5. Общие принципы индексации и ранжирования документов. 2.6. Проблемы машинного перевода. Перевод как прикладная лингвистическая дисциплина.
3	Компьютерная обработка естественно-языковых данных. Корпусная лингвистика и компьютерная лексикография	2	3.1. Корпусная лингвистика. Понятие источника материала. Типы источников (словари, тексты, экспериментальные данные). 3.2. Массивы языковых данных. Лингвистические корпусы. Конкорданс. Лингвистическая разметка. 3.3. Классификация и обзор корпусов. Национальный корпус русского языка ruscorpora.ru. 3.4. Компьютерная лексикография. Компьютерные программы поддержки словарей (базы данных, электронные картотеки, программы обработки текстов). Принципы построения машинного словаря. Машинные словари. Компьютерная поддержка электронных библиотек. 3.5. Информационно-поисковые системы. 3.6. Обзор программного обеспечения лингвистических исследований.
4	Компьютерный анализ естественно-языкового текста	2	4.1. Проблема атрибуции текста. Атрибуция текста в криминалистике. 4.2. Лингвистические принципы автоматического выделения информации из текста. Проблемы автоматического реферирования документов. 4.3. Формализация филологических моделей художественного текста. Лексическая статистика и идиостиль автора. 4.4. Гипертекстовые технологии представления текста. Понятие текста и гипертекста. Структура и компоненты гипертекста.

Таблица 2

№ п/п	Содержание лабораторного занятия	Часы (18)
1	Программно-лингвистическое обеспечение ПЭВМ. Основы кодирования.	2
2-3	Лингвистические компоненты текстовых процессоров (встроенные словари; программы автоматического переноса слов; орфографические редакторы; синтаксические и стилистические редакторы).	4
4	Программы оптического распознавания печатных и рукописных текстов.	2
5	Гипертекст. Основы HTML.	2
6	Интернет как гипертекстовая технология. Поисковые системы. Лингвистические корпусы.	2
7	Системы автоматического перевода.	2
8-9	Компьютерная лексикография. Обучающие лингвистические системы.	4

Е. А. Кривонос

### **ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ ИССЛЕДОВАТЕЛЬСКОГО КОРПУСА РУССКИХ ТЕКСТОВ ПО ЭКОНОМИКЕ В ЛИНГВО-МЕТОДИЧЕСКИХ ЦЕЛЯХ**

Корпусная лингвистика является одним из динамично развивающихся направлений прикладной лингвистики. А. Н. Баранов определяет корпус текстов как «вид корпуса данных, единицами которого являются тексты или их достаточно значительные фрагменты, включающие, например, какие-то полные фрагменты макроструктуры текстов данной проблемной области» [1, с. 115]. Создание корпусов текстов имеет огромное значение, так как позволяет, во-первых, представить языковой материал определенным образом, что способствует как расширению круга исследовательских задач, так и автоматизации анализа лингвистических данных, и, во-вторых, повысить уровень абстрагирования от субъективности исследователя.

Теоретической основой нашего диссертационного исследования стали следующие положения: профессиональный подязык – это фрагмент национального языка, имеющий ярко выраженную профессиональную направленность и приспособленный для специальной коммуникации (А. С. Герд, А. А. Реформатский, А. В. Суперанская); терминополье и терминосистема являются основой базы знаний предметной области (Г. О. Винокур, М. Н. Володина, А. С. Герд, Б. Н. Головин, Р. Г. Пиотровский, С. Д. Шелов).

Одна из основных целей нашего диссертационного исследования заключалась в создании «Тезауруса по экономике для иностранных граждан,

изучающих русский язык (этап довузовской подготовки)» (ТЭИ-0) на базе исследовательского корпуса текстов (ИКТ-0). Также целью является описание структурных типов терминологических сочетаний, включенных в ТЭИ-0.

Материалом для исследования явились учебные тексты по экономике на русском языке из учебно-методического комплекса «Экономика» для иностранных граждан [2], активно применяемого в учебном процессе на факультете доуниверситетского образования БГУ. Всего в ИКТ-0 был включен 41 учебный текст.

Работа по разработке ТЭИ-0 была разбита на несколько этапов. На первом этапе работы над ТЭИ-0 тексты из ИКТ-0 были обработаны. В результате содержание каждого текста было представлено в виде семантической сети, и для каждого текста создан минитезаурус. На втором этапе полученная информация была обобщена, что позволило определить круг терминологических единиц, подлежащих упорядочиванию. Следующим этапом в работе явилось структурирование информации в соответствии с моделью тезауруса. Опытная модель построения двуязычного словаря-тезауруса по экономике, предназначенного в первую очередь для целевого использования в учебном процессе, подробно описана в монографии З. И. Комаровой и А. А. Прошиной [3]. Авторы отмечают, что моделируемый словарь сочетает в себе типологические характеристики идеографического, толкового и переводного словаря и выполняет все основные словарные функции: систематизирующую, справочную, нормативную и учебную. Наличие в словаре дефиниций необходимо, потому что экономическая терминосфера находится в стадии формирования, что влечет за собой некоторую неупорядоченность. Кроме того, созданный нами ТЭИ-0 ориентирован на иностранных студентов, поэтому в учебных целях наличие полной дефиниции, готовой для воспроизведения мы считаем необходимым. Модель тезауруса. Всего в ТЭИ-0 содержится 328 элементов: 70 терминов и 258 терминологических словосочетаний, что в процентах составляет 21,34 % и 78,66 % соответственно.

Как мы видим, количество терминологических сочетаний значительно превышает количество однокомпонентных терминов. С учетом того, что ТЭИ-0 ориентирован на иностранных студентов, представление структурных типов является весьма актуальным, так как эксплицирует важную информацию для подготовки учебных пособий на начальном этапе, ориентируя авторов относительно приоритетности презентации грамматической информации.

Двухкомпонентную структуру, которая далее представлена двумя типами сочетаний, имеют 177 элементов, что соответствует 54 % от общего числа: типу **2.1** соответствует структура **Adj<sup>N</sup>S<sup>N</sup>** (*аграрное богатство*), **2.2** – **S<sup>N</sup>S<sup>G</sup>** (*биржа труда*). Количественные показатели свидетельствуют, что именно двухкомпонентные структуры являются самыми частотными в ТЭИ-0: тип 2.1 – 125 элементов, из которых опорное существительное

женского рода – 41, мужского – 30, среднего – 18, множественного числа – 36; тип 2.2 – 52 элемента.

Трехкомпонентную структуру имеют 64 элемента, что составляет 19,51 %. В данной структуре мы выделили 7 типов: **3.1** –  $S^N Adj^G S^G$  (*агрегирование экономических показателей*), 13 эл.; **3.2** –  $Adj^N S^N S^G$  (*валовые издержки производства*), 13 эл.; **3.3** –  $Adj^N Adj^N S^N$  (*валовый внутренний продукт*), 14 элементов; **3.4** –  $S^N S^G S^G$  (*виды издержек производства*), 19 эл.; **3.5** –  $S^N S^N S^G$  (*график «Кривая предложения»*), 2 эл.; **3.6** –  $S^N Pr^{+Acc} S^{Acc}$  (*налог на наследство*), 2 эл.; **3.7** –  $S^N Pr^{+G} S^G$  (*налог с продаж*), 1 эл. После анализа типов трехкомпонентных структур мы можем утверждать, что в своем большинстве базой для их образования являются двухкомпонентные.

Четырехкомпонентные структуры (11 элементов, 3,35%) представлены 6 типами: **4.1** –  $Adj^N S^N Adj^G S^G$  (*внешние факторы экономических циклов*), 1 эл.; **4.2** –  $S^N Adj^G Adj^G S^G$  (*закон возрастающих вмененных издержек*), 1 эл.; **4.3** –  $S^N S^G Adj^G S^G$  (*предмет изучения экономической теории*), 4 эл.; **4.4** –  $S^N Adj^G S^G S^G$  (*факторы ценовой эластичности спроса*), 1 эл.; **4.5** –  $S^N S^N Adj^G S^G$  (*график «Кривая совокупного предложения»*), 2 эл.; **4.6** –  $S^N Pr^{+Acc} Adj^{Acc} S^{Acc}$  (*налог на добавленную стоимость*). На основе данных схем мы можем сделать вывод, что четырехкомпонентные элементы в основной массе образуются через расширение трехкомпонентных, реже – двухкомпонентных структур.

Пятикомпонентные структуры, представленные 3 типами, образуют группу из 6 элементов, что соответствует 1,8%: **5.1** –  $Adj^N S^N S^G S^G S^G$  (*неценовые факторы изменения величины предложения*), 3 эл.; **5.2** –  $Adj^N S^N S^G Adj^G S^G$  (*ценовые факторы изменения совокупного предложения*), 2 эл.; **5.3** –  $S^N S^G Pr^{+Acc} Adj^{Acc} S^{Acc}$  (*способы воздействия на денежное предложение*), 1 эл. Данные структуры, как и предыдущие, развиваются посредством включения лексических единиц четырех-, трехкомпонентные структуры.

Описание структурных типов терминологических сочетаний указывает необходимость подробного изучения в первую очередь двухкомпонентных структур, которые можно назвать ядерными, в силу того что они изначально выступают базой, на которой происходит дальнейшее структурное и семантическое развитие сочетаний. Самыми частотными являются формы родительного падежа существительных и прилагательных, что влечет за собой необходимость знания студентом-инофоном способов образования форм данного падежа, а также значений этих форм (*часть целого, атрибут* и т. д.). Кроме того, схемы позволяют проследить пути развития терминологического сочетания и презентовать данный материал обучаемым, с целью формирования умения самостоятельно опознавать терминологические сочетания в научных текстах.

Исходя из вышеизложенного мы можем сделать следующие выводы. Во-первых, достоинством специального корпуса текстов является то, что

он представляет собой многоуровневую систему многократного использования, которая позволяет совершать различные операции в процессе решения конкретных исследовательских задач. Во-вторых, содержание и организация частей созданного ТЭИ-0 в полном объеме позволяют реализовать систематизирующую, справочную, нормативную и учебную функции. ТЭИ-0 необходим как для облегчения процесса усвоения основных понятий экономической терминологии на русском языке, так и для осознания связей и отношений между элементами терминологической системы. В-третьих, описание структурных типов терминологических сочетаний имеет важное значения для определения приоритетности презентации грамматической информации.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 Баранов, А. Н. Введение в прикладную лингвистику / А. Н. Баранов. – МГУ им. М. В. Ломоносова. Филол. фак. – 2-е изд., испр. – М. : Эдиториал УРСС, 2003. – 350 с.
- 2 Родионов, В. Н. Экономика: учеб.-метод. комплекс для иностранных слушателей фак. доунив. образования БГУ / В. Н. Родионов, А. В. Воробьев, Т. З. Сафронова. – Минск. : БГУ, 2006. – 240 с.
- 3 Комарова, З. И. Моделирование двуязычного словаря-тезауруса по экономике: Монография / З. И. Комарова, А. А. Прошина. – Екатеринбург: Форум-книга, 2009. – 276 с.

Н. И. Лапицкая

### **ПРЕЗЕНТАЦИЯ КАК ВИД НАГЛЯДНОСТИ НА УРОКАХ РУССКОГО ЯЗЫКА В ШКОЛЕ**

Компьютерная презентация на сегодняшний день является довольно распространённым средством наглядности на уроках русского языка в школе. Многие учителя активно используют их в своей практической деятельности. На использование компьютерных презентаций в течение педагогической практики на 4 и 5 курсах нацеливаем студентов на занятиях по методике преподавания русского языка.

Попробуем проанализировать некоторые моменты построения и применения компьютерных презентаций.

От чего зависит значимость презентации, на каких этапах урока можно её использовать? Что следует включать в презентацию? На наш взгляд, это те вопросы, на которые ещё не даны чёткие ответы, хотя решаются они каждым педагогом индивидуально.

По мнению С. В. Гегеле, эффективность учебной компьютерной презентации во многом зависит от качества реализации первого этапа процесса подготовки – формулировки цели её использования. «Цели компьютерной презентации определяются с одной стороны её назначением, то есть конкретной ситуацией использования презентации, потребностью процесса в её

создании, с другой стороны целями процесса обучения в целом и целями конкретного урока в частности» (1, с. 384).

На наш взгляд, первой из названных целей является конкретная цель по объяснению того или иного материала (дать представление об основных признаках текста, учить определять спряжение глагола по неопределённой форме, повторить правописание личных окончаний глагола и т. д.), реже закрепление какого-то материала (формировать навыки различения омонимичных частей речи и др.).

Как видно из поставленных целей, компьютерную презентацию целесообразно использовать в первую очередь на этапе объяснения нового материала, для того чтобы дать представление о том или ином материале. Нам представляется, что с помощью презентации должны излагаться большие по объёму темы, довольно сложные, насыщенные большим количеством теоретического материала (н-р, «Функциональные стили речи», «Качества хорошей речи» и др.). Вот как может выглядеть начало одной из таких презентаций:

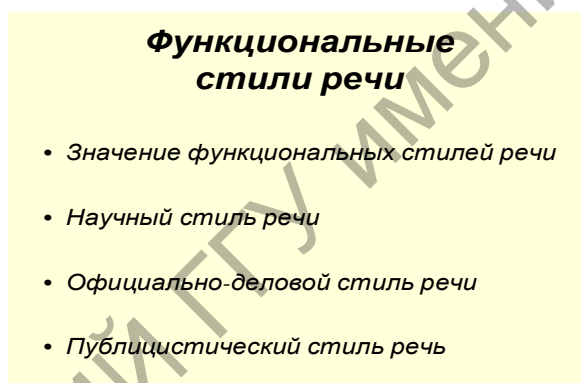


Рисунок 1

Доказательством сказанного выше является тот факт, что студенты, которым на занятиях по «Методике преподавания русского языка» предлагается сделать компьютерную презентацию, выбирают теоретические темы гораздо чаще («Имя существительное как часть речи», «Жанры речи» и т. д.).

Кроме того, презентации активно используются при объяснении материала по орфографическим темам, когда нужно объяснить конкретное написание (н-р, «Правописание Н и НН в суффиксах прилагательных»). В данном случае компьютерная презентация «работает» как таблица, схема, алгоритм, т. к. позволяет большой и сложный материал изложить в компактной форме, наглядно, чётко, а алгоритм использовать в качестве руководства к действию. Вот пример такого использования:





Рисунок 2

На наш взгляд, менее эффективно использование презентаций на этапе закрепления материала. Это связано с тем, что многие упражнения представляют собой продублированные упражнения из учебника. Например:

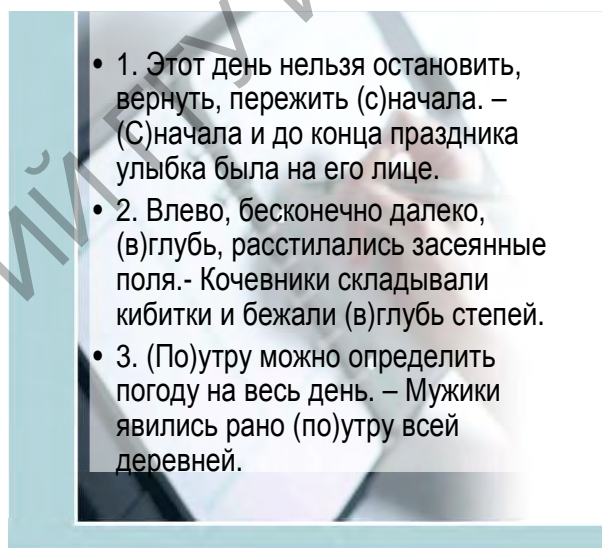


Рисунок 3

Но ведь рассматривать и выполнять такое упражнение по учебнику (следовательно, по напечатанному тексту) гораздо легче. Это не требует напряжения зрения. Материал находится рядом с ребёнком, что способствует большей сосредоточенности внимания.

Целесообразно в такой ситуации использовать упражнения в тестовой форме либо упражнения, после которых даётся правильный вариант выполнения для самопроверки.

По-разному можно относиться и к большому количеству эффектов, возможность использования которых предоставляет презентация. Нам представляется, что презентация, выполненная в «строгом» стиле, в обычном чёрно-белом исполнении даёт гораздо больше возможности сосредоточиться, чётко воспринимать материал. Не следует делать упор и на возрастные различия школьников. Анимационные эффекты отвлекают внимание учащихся и в среднем и в старшем звене.

Таким образом, учебная компьютерная презентация является новым эффективным средством наглядности на уроках русского языка в школе. Она, как и другие виды наглядности, требует серьёзного продумывания при подготовке.

#### СПИСОК ЛИТЕРАТУРЫ

1 Гегеле, С. В. Анализ этапов подготовки учебной компьютерной презентации. С. В. Гегеле // Российско-белорусско-украинское пограничье: аспекты взаимодействия в контексте единого социокультурного пространства – история и перспективы: Материалы международной научно-практической конференции. – Брянск: РИО БГУ, 2010. – 432 с. – С. 383–385.

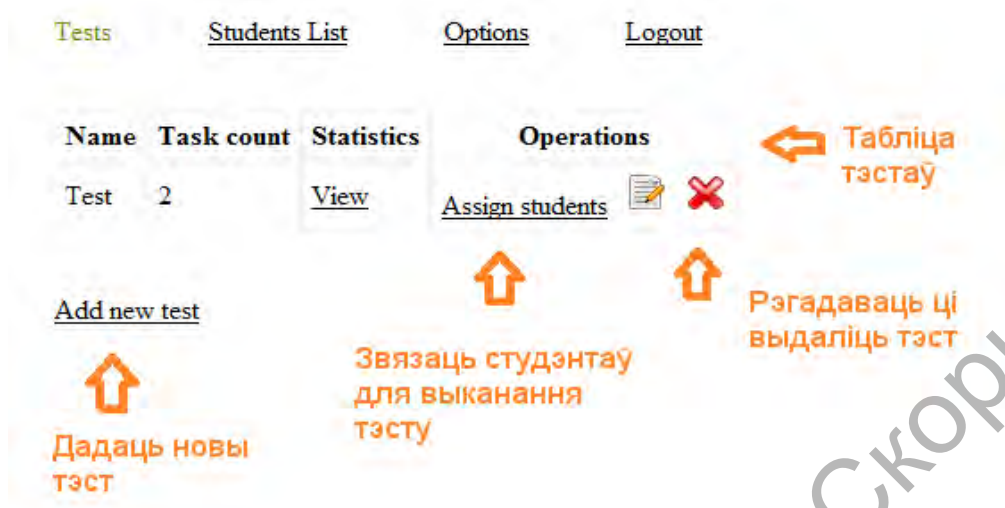
А. М. Палуян, С. С. Палуян

#### **ВЫКАРЫСТАННЕ ІНФАРМАЦЫЙНЫХ ТЭХНАЛОГІЙ ПРЫ ПРАВЕРЦЫ ВЕДАЎ СТУДЭНТАЎ**

У апошнія гады ў сферы адукацыі асаблівае значэнне набывае кампетэнтнасны падыход у падрыхтоўцы спецыялістаў і расце цікавасць да аўтаматызацыі працежкавага і вынікавага кантролю іх навучання. Адным з самых папулярных відаў такога кантролю з'яўляецца тэсціраванне, заснаванае на дыялогу вылічальнай сістэмы з карыстальнікам. Укараненне інфармацыйных тэхналогій у адукацыйнае асяроддзе, з'яўленне эфектыўных сістэм праграмавання, якія прапануюць разнастайныя формы ўзаемадзеяння, дазваляе аб'ектыўна, хутка і надзейна ацаніць веды студэнтаў.

Сістэма тэст-кантролю ведаў студэнтаў па тых ці іншых дысцыплінах вучэбнага плана ўмоўна дзеліцца на дзве часткі.

Першая – стварэнне ці рэдагаванне тэстаў, другая – праходжанне тэста студэнтамі і прагляд вынікаў праверкі ведаў. Першая частка даступна для выкладчыка, які ўваходзіць у сістэму з дапамогай імені карыстальніка і пароля. Выкладчыку даступны заданні тэстаў, спісы студэнтаў і спасылкі на старонкі для рэдагавання.



Малюнак 1 – Табліца тэстаў у раздзеле для выкладчыка

Пры падрыхтоўцы сістэмы тэст-кантролю неабходна вызначыць: 1) тыпы тэстаў (колькасць варыянтаў заданняў, колькасць правільных адказаў, колькасць заданняў у тэсце); 2) спосабы ацэнкі адказаў (балы за адказ, ніжняе і верхняе значэнне балаў); 3) элементы інтэрфейса (кіраванне, увод – вывад адказаў-вынікаў); 4) адміністрацыйныя функцыі сістэмы.

У спецыялізаванай літаратуры педагагічны тэст разглядаецца як сістэма кароткіх, узаемазвязаных паміж сабой агульнай логікай і агульным зместам заданняў, якія адпавядаюць пэўным навукова абгрунтаваным крытэрыям якасці інфармацыі. Патрабаванні да складання тэстаў наступныя:

- змястоўная валіднасць: заданні тэсту павінны быць пабудаваны на аснове тых вучэбных элементаў, якія з’яўляюцца абавязковымі для вывучэння, г. зн. адпавядаюць адукацыйнаму стандарту;
- функцыянальная валіднасць: адпаведнасць заданняў тэсту ўзроўню засваення ведаў, вызначанаму стандартам;
- рэпрэзентатыўнасць: неабходнасць максімальнага адлюстравання студэнтамі матэрыялу;
- надзейнасць: устойлівасць вынікаў праведзенага тэсціравання пры неаднаразовым выкарыстанні тэсту, г. зн. надзейны тэст павінен даваць пастаянныя вынікі ацэнкі пры паўторным прад’яўленні.

Распрацоўваць тэсты неабходна па прынцыпу ўзрастання цяжкасці. Суаднясенне заданняў нізкай, сярэдняй і высокай ступені цяжкасці па кожным ключавым моманце вывучаемага матэрыялу вызначае структуру тэсту.

Пры рэдагаванні тэстаў мы можам дадаваць, змяняць ці выключаць са спісу прапанаваных заданні. Даступны тры віды заданняў: з некалькімі варыянтамі адказу, дзе толькі адзін будзе правільным; з некалькімі варыянтамі адказу і некалькі з іх будуць правільнымі; з полем для ўвядзення адказу.

## Рэдагаванне тэсту

Test name:

**Заданні** **Tasks** **Answers** **Адказы**

Задание 1 – выберите 1 вариант

Задание 2 – выберите 1 или несколько вариантов

Задание 3 – впишите ответ

Вариант 1  
Вариант 2  
Вариант 3  
Вариант 4

Вариант 1  
Вариант 2  
Вариант 3

Ответ

Адзін правільны варыянт адказу

Некалькі правільных варыянтаў

Студэнт сам уведзіць адказ

[Add task with one correct answer](#) [Add task with some correct answers](#) [Add task with input field](#)

Малюнак 2 – Старонка рэдагавання тэсту

Пасля стварэння тэсту выкладчык “звязвае” з ім спіс студэнтаў. Пасля выбару студэнтаў для кожнага з іх ствараецца ўнікальны код, з дапамогай якога студэнт можа ўвайсці ў сістэму для праходжання тэсціравання. Выкладчык можа праглядаць спіс студэнтаў з кодамі і захаваць ці раздрукаваць яго. Рэдагаванне спісу студэнтаў з’яўляецца максімальна простым і зручным: паказана агульная табліца студэнтаў і груп, у якіх яны вучацца, даступныя кнопкі “Рэдагаваць” і “Выдаліць”, якія дазваляюць выканаць выбранае намі дзеянне прамі ў табліцы без перазагрузкі старонкі. Акрамя таго, у канцы табліцы ёсць два палі – для ўводу імені студэнта і яго групы, а таксама кнопка “Дадаць”, якая адразу ж дадае студэнта ў спіс.

Tests **Students List** Options Logout

Surname and name	Group	Operations
Иван Петров	БА-43	
Алексей Иванов	РФ-31	
<input type="text" value="Василий Васильев"/>	<input type="text"/>	

Спіс студэнтаў

Даданне, рэдагаванне ці выдаленне студэнта

Малюнак 3 – Старонка рэдагавання спісу студэнтаў

Акрамя пералічанага вышэй, выкладчыку даступны шэраг іншых магчымасцей: дадаць новага карыстальніка-выкладчыка, які таксама можа рэдагаваць тэсты і спісы студэнтаў; выбраць мову інтэрфейса праграмы (беларуская, руская, англійская); прагляд статыстыкі па тэстах (колькі студэнтаў прайшлі; сярэдні бал) і прагляд статыстыкі для канкрэтнага студэнта (якія тэсты праходзіў; колькі набраў па іх балаў і як адказаў на пытанні).

Праходжанне тэсту студэнтам адбываецца наступным чынам: студэнт уваходзіць у сістэму пры дапамозе кода, прысвоенага яму выкладчыкам. Далей студэнту даступна старонка са спісам пытанняў па тэсту і варыянтамі адказу. Адказаўшы на ўсе пытанні, студэнт націскае кнопку “Адправіць”. Сістэма правярае яго вынік і паказвае, на якія пытанні ён адказаў правільна і колькі набраў балаў. Гэтыя вынікі даступны і выкладчыку.

### Проверка знаний

#### Задание 1

Задание 1 - выберите 1 вариант

- Вариант 1  
 Вариант 2  
 Вариант 3  
 Вариант 4

#### Задание 2

Задание 2 - выберите 1 или несколько вариантов

- Вариант 1  
 Вариант 2  
 Вариант 3

#### Задание 3

Задание 3 - впишите ответ

Submit Query

Праходжанне тэста студэнтам

Малюнак 4 – Праходжанне тэсту студэнтам

Такім чынам, прадстаўленая намі тэст-кантрольная сістэма мае лёгкі і зразумелы інтэрфейс, у ёй зручна рэдагаваць заданні, праглядаць падрабязную статыстыку па заданнях і па выніках праходжання тэсту студэнтамі. Уваход у сістэму ажыццяўляецца для студэнтаў – па аднаразовых кодах, для выкладчыка – па імені і паролі.

Выкладчык можа: 1) рэдагаваць заданні і спісы студэнтаў; 2) дадаваць новых карыстальнікаў-выкладчыкаў; 3) праглядаць статыстыку па тэстах і выніках студэнтаў.

Сістэма дазваляе ствараць заданні розных тыпаў:

- з некалькімі варыянтамі адказу;
- з адным варыянтам адказу;
- з полем уводу адказу.

Праграма вядзе статыстыку па 1) колькасці студэнтаў, якія прайшлі тэст; 2) колькасці і спісу студэнтаў, якія павінны будучь прайсці; 3) сярэднім балам.

Праграме ўласцівы таксама такія магчымасці, як шматмоўны інтэрфейс (беларуская мова, руская і англійская), экспарт статыстыкі, вынікаў, кодаў для студэнтаў у зручным для раздрукоўкі фармаце; работа пры павольным злучэнні і на ўстройствах з невялікім экранам.

М. В. Половец

## **ФУНКЦИОНАЛЬНЫЙ АСПЕКТ ЛИНГВИСТИЧЕСКИХ СРЕДСТВ ДИАЛОГА В ИНТЕРНЕТ-ДИДАКТИКЕ ПРИ ОБУЧЕНИИ РУССКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ**

Эффективность использования компьютерных технологий и Интернета как средства обучения является неоспоримым фактом. Интернет-сайты, предназначенные для изучения русского языка как иностранного, все чаще используются преподавателями для самых различных дидактических целей. Немаловажным в процессе обучения является способ ведения диалога с пользователем, в данном случае учащимся. Средства подобного диалога имеют свою специфику. Это связано, в первую очередь, с особенностями предмета обучения, то есть языка. Как результат следует предполагать, например, использование метаязыковых средств, а также национально и лингвистически специфичных средств диалога.

Рассмотрение диалоговых средств в Интернет-продуктах, направленных на изучение русского языка как иностранного, позволило выделить три основных аспекта их описания:

- 1) функциональный;
- 2) формальный;
- 3) аспект типа использованных языковых средств.

Функциональный аспект лингвистических средств человеко-машинного диалога непосредственно связан со стилистической принадлежностью того или иного диалогового средства. Каждый функциональный стиль обладает своими особенностями использования общелитературной нормы. Чаще всего стили сопоставляются на основе присущего им словоупотребления, так как именно в словоупотреблении наиболее ярко проявляется различие сфер коммуникации. Однако и грамматические



характеристики здесь немаловажны. Каждый стиль характеризуется следующими признаками: целью общения, набором языковых средств и формами (жанрами), в которых он существует.

В исследованных лингводидактических ресурсах [1, 2, 3, 4, 5] преобладают нейтральные стилистически лексические единицы и грамматические конструкции. Однако наблюдается использование элементов научного стиля в формулировке грамматических правил либо при семантизации лексики. Также присутствуют и элементы публицистического стиля. Публицистический стиль русского языка представлен, в основном, в языковом материале для чтения, так как многие сайты предлагают газетные статьи и новости для развития навыков чтения, а иногда и аудирования как рецептивных видов речевой деятельности. Это обусловлено, по нашему мнению, особенностями среды ведения диалога. Диалог с компьютером ведется в специфической технически обусловленной среде. В отличие от естественного общения, где возможно использование различных, характерных для естественного общения, невербальных диалоговых средств, технически опосредованное общение имеет свой инструментарий диалога. К тому же, Интернет рассматривается как одно из средств массовой информации, а публицистический стиль является преобладающим в этой сфере.

В диалоговой структуре сайта «Master Russian» достаточно частотным является использование Интернет-сленга, то есть разговорной речи, а именно условных сокращений (Рисунок 1).



Рисунок 1 - [http://masterrussian.com/russian\\_alphabet.shtml](http://masterrussian.com/russian_alphabet.shtml)

В данном случае акроним «а. к. а», обозначающий «also known as» или *‘также известный как’* является элементом разговорной речи, принятым в электронной коммуникации.

Еще одним лексическим маркером разговорной речи может служить лексема «Ads», являющаяся сокращением слова «Advertisements» (*‘Реклама’*). Данное диалоговое средство широко используется почти всеми рассмотренными Интернет-продуктами.

Использование элементов разговорной речи в Интернет-среде вполне оправдано законом «экономии речевых усилий», это является общепризнанным фактом. Но следует принять во внимание цель существования рассмотренных Интернет-продуктов – изучение языка. Поскольку использование разговорной речи в дидактическом диалоге неуместно, то становится очевидным вывод о том, что Интернет-среда в данном случае играет большую роль в выборе диалоговых средств, чем принятые в обществе нормы ведения дидактического диалога.

Примером грамматической характеристики публицистического стиля является использование неполных предложений в названиях статей и подзаголовках. Это же грамматическое средство прослеживается и в структуре диалога сайта «Transparent Language», например «The better choice for language learning» (*‘Лучший выбор для изучения языка’*). Такого рода предложения называются слоганами и часто используются в диалоге на разных сайтах. В Интернет-общение они пришли из рекламы в печатных и других масс медиа.

Необходимо отметить также использование терминологии в структуре диалога выбранных сайтов. Не стоит забывать, что терминологические единицы могут быть рассмотрены в двух различных аспектах: как элементы метаязыка и как обозначения понятий предмета обучения. Использование терминологии, безусловно, характерно научному стилю речи.

Таким образом, функциональный аспект лингвистических средств диалога в Интернет-дидактике при обучении русскому языку как иностранному определяется взаимодействием научного, публицистического и разговорного стилей с элементами разговорной речи.

#### СПИСОК ЛИТЕРАТУРЫ

- 1 An Interactive Online Reference Grammar [Electronic resource]. – Mode of access: <http://www.alphadictionary.com/rusgrammar>. – Date of access: 22.09.2010.
- 2 Elanguage school. [Electronic resource]. – Mode of access: <http://learnrussian.elanguageschool.net>. – Date of access: 22.09.2010.
- 3 Master Russian. [Electronic resource]. – Mode of access: <http://masterrussian.com>. – Date of access: 22.09.2010.
- 4 Russian language lessons. [Electronic resource]. – Mode of access: <http://russianlessons.net>. – Date of access: 22.09.2010.
- 5 Transparent language. [Electronic resource]. – Mode of access: <http://www.transparent.com>. – Date of access: 22.09.2010.



## ПОСТРОЕНИЕ ТЕРМИНОЛОГИЧЕСКИХ ИНФОРМАЦИОННЫХ РЕСУРСОВ ПО ДОКУМЕНТАЛЬНЫМ ПРОЦЕССАМ, ТЕРМИНОВЕДЕНИЮ И ПОИСКУ ИНФОРМАЦИИ

Проводимые исследования в области формирования, систематизации, описания норм (кодификации) и стандартизации терминологии документно-коммуникационной и других сфер являются важной научной проблемой, т.к. результаты происходящих в этих сферах трансформаций напрямую отражаются в их терминосистемах. При этом важной задачей является «системное изучение различных профессиональных лексиконов» [1, с. 7].

Терминология каждой науки, сферы деятельности отображает богатство профессиональных знаний [2, с. 31]. Внедрение информационных технологий (ИТ) во все сферы жизни обществ активизирует работу международных организаций по принятию стандартов (законов) по регламентации управления документальными процессами (УДП) и по унификации терминологии. Обогащение национальных терминосистем иностранной лексикой серьезно испытывает действующий терминофонд на системность: из-за примирения с алогизмом заимствованных иноязычных терминов можно попросту оказаться в Вавилонской башне [3, р. 158]. Не меньшую опасность представляет также пренебрежение терминологией.

Сложившаяся экстралингвистическая ситуация диктует необходимость создания интеллектуальных терминологических информационных ресурсов (ИТИР), которые: 1) стали бы легитимным средством фиксации терминологии, накопления научных знаний и общения профессионалов; 2) предоставляли бы полную информацию для выбора и правильного использования нужного термина: сведения о терминах и допустимости их использования, правовом статусе терминов и документах, его устанавливающих, и др.; 3) смогли бы стать научной основой стандартизации терминосистем.

Созданный автором статьи ИТИР «InfoLio&Rec» зарегистрирован в Республике Беларусь как информационно-поисковая и справочная система по рубрике «Государство и право. Юридические науки» [4]. «InfoLio&Rec» сформирован из 8900 терминов с соответствующими определениями на основе анализа состояния науки и ИТ, нормативных правовых актов (НПА) и технических НПА (законов, директив, стандартов, руководящих документов и т. д.) международных организаций, Республики Беларусь и других стран, иных источников по документоведению и УДП; пространствам Мира хранения; менеджменту знаний; информационному праву, технологиям и безопасности в госуправлении (электронному правительству); смежным областям знаний. Документация, одновременно относящаяся к профессиональной деятельности всех видов отраслей, в настоящее время представляет собой определенную сложность для перевода в связи с отсутствием отдельных единых словарей, которые смогли бы адекватно отразить

изменившийся и пополнившийся понятийный аппарат предметной области [1, с. 7]. Сейчас в «InfoLio&Rec» включена информация на русском, белорусском, английском языках, проводится работа над включением туда терминологии на немецком, украинском и французском языках.

Актуальность создания «InfoLio&Rec» диктуется: 1) многообразием терминологических аппаратов указанных областей; 2) необходимостью их упорядочения и нормализации, отсутствием полных терминологических словарей в этих областях, а также единой информационной системы, предоставляющей полную информацию о терминосистемах указанных областей; 3) существенными расхождениями в толковании наукой и практикой ряда терминов; 4) необходимостью гармонизации национальных терминосистем с терминологией международных сообществ, закрепления новых терминов и очищения терминосистем от устаревших терминов.

Основная цель «InfoLio&Rec» – постоянная системная поддержка процессов нормализации профессиональной терминологии для повышения эффективности и устойчивости социальных коммуникаций и управления, т.е. научно-методическая помощь в установлении непротиворечивой и однозначно понимаемой терминологии по направлениям формирования ИТИР. Целью его является также возможность найти значение определенного термина, а также нормативную и справочную информацию, необходимую пользователю, связанную с объектами и процессами, стоящими за терминологией. На основе «InfoLio&Rec» возможна разработка и (или) корректировка специализированных БД, классификаторов, справочников, но ни в коем случае не полная замена уже существующих. Достижение основной цели разработки «InfoLio&Rec» предполагает: 1) выявление всего многообразия терминов указанных предметных областей; 2) отделение терминов определенной предметной области от терминов смежных областей знания; 3) выявление факторов воздействия и результатов влияния на терминосистему; 4) определение состава необходимых и достаточных терминов для развития предметной области; 5) установление иерархий понятий; 6) четкую дефиницию терминов; 7) проверку обоснованности полисемии и синонимии терминов; 8) устранение нарушения правил общего и специального языка.

Основными источниками формирования «InfoLio&Rec» в сфере информации и документации стали международные стандарты ISO, ISO/IEC, различные законы, указы, директивы, постановления руководящих органов ООН, Евросоюза, Республики Беларусь, России и других стран, регулирующие вопросы управления документами в органах государственной власти и управления. Но ни один из таких источников не может претендовать на полный и последовательный охват терминосистем по всем направлениям формирования ИТИР. К тому же, большинство средств фиксации терминосистем в указанных областях неполны и непоследовательны.

Специалисты, работающие в современном научном пространстве, должны иметь возможность выбора конкретной информационной системы

для перехода от применения «простых» терминологических ресурсов (учебные пособия, фонетические и обучающие системы, словари и глоссарии, находящиеся в памяти компьютера и сетях), к более сложным системам поиска и обработки информации, машинного перевода и т.д. [5, с. 27-28]. Каждый ИТИР должен позволять осуществлять поиск терминов с применением ключевых слов, тематических категорий, сущностей, документов, а также помогает в выборе термина на основе представления полной информации о нем. В «InfoLio&Rec», если пользователь не смог найти необходимой ему терминологии, то он может выйти на рекомендованные автором информационные ресурсы (ИР) Интернет или на работы ключевых специалистов по терминоведению и направлениям формирования ресурса.

Взаимодействие профессионального лексикона с общеязыковой системой отражается в процессах терминологизации, ретерминологизации, детерминологизации [1, с. 7]. Поэтому возникает необходимость внедрения ИТИР в учебные процессы. Сложность в запоминании формулировки в данном случае будет компенсироваться наличием четких образов (моделей) определений и обновлением их в памяти. Система обучения терминологии по направлениям формирования «InfoLio&Rec» включает терминологию на различных языках, перечень сокращений и специальных обозначений. Предлагаемая технология обучения терминологии основана на выделении для каждого термина словаря группы терминов, которые составят тезаурус-микрословарь, поддерживающий иерархию терминов, связи между ними и (или) определенные операции – поиск более общих терминов, более узких, взаимосвязанных, и т. д. В настоящее время повышается роль словарей специальной лексики «как средства регистрации и представления знаний. Словари тезаурусного типа вполне могут стать ядром компьютерных систем искусственного интеллекта. Словари могут быть использованы и как средство реконструкции и исследования процесса развития человеческих знаний» [6, с. 153–154].

Создаваемый автором второй ИТИР – «Terminus» – охватывает терминологию документоведения и поиска информации. Сейчас проходит этап государственной регистрации ИТИР «Terminus» в Государственном регистре информационных ресурсов РБ. ИТИР «Terminus» охватывает ИТИР «InfoLio&Rec» с точки зрения ранее включенной в него терминологии, поэтому в «Terminus» особое внимание уделяется работе над терминологией базовых понятий: *termin* (термин), *terminology* (терминология), *terminosystem* (терминосистема), *terminoseries* (терминоряды), *terminosophos* (терминософия), *terminosphere* (терминосфера), и т. д.

База источников, диссертаций и словарей формировалась на основе ключевых слов, представленных следующими корневыми морфемами: *граммат\**, *дефини\**, *знак\**, *индекс\**, *ключ\**, *когит\**, *когнит\**, *лекс\**, *лингв\**, *логик\**, *логич\**, *определен\**, *поиск\**, *прагма\**, *семант\**, *семио\**, *синта\**, *систем\**, *слов\**, *словар\**, *словн\**, *текст\**, *термин\**, *фраз\**, *энциклопед\**, *язык\**.

Источники по терминоведению и смежным областям классифицированы следующим образом: 1) монографии и статьи классической и современных лингвистических школ (В. фон Гумбольдта, австрийской, франко-квебекской, женеvской, датской, пражской, польской, и др.), советской и национальной белорусской школ, стран СНГ (более 4600 ссылок); 2) диссертации советской и национальной белорусской школ, стран СНГ (более 8370 ссылок); 3) глоссарии, лексиконы, терминологические сборники, словари, словники, справочники, энциклопедии (более 8200 ссылок). Отдельными списками – по персоналиям – будут включаться ссылки на труды ключевых ученых и исследователей. Вследствие существенной междисциплинарности ИТИР также включены необходимые источники по философии науки, сознания, математики, физики, психологии, права, и др.

В рекомендуемые автором электронные ИР вошли: 1) юридический словарь терминов и соответствующих им определений на Национальном правовом портале Республики Беларусь; 2) тематический глоссарий русского языка; 3) толковые словари В. И. Даля, Д. Н. Ушакова, С. И. Ожегова; 4) Википедия; 5) Оксфордский и Вебстерский толковые словари; 6) справочная система Сети Xrefer; 7) web-портал «Мир словарей» (в том числе словари бизнеса, биографический, большой энциклопедический, исторический, политический, социологический, философский, финансовый, экономический, юридический, словарь Ефремовой, словари имен, компьютерного жаргона, логики, символов, синонимов); 8) дополнительные ИР.

По аналогии с ИТИР «InfoLio&Res» включен раздел по обучению терминологии. В «Terminus» возможна работа с информацией, конвертированной в форматы DOC, HTML, PDF. «InfoLio&Res» работает только с информацией, конвертированной в формат HTML.

#### СПИСОК ЛИТЕРАТУРЫ

1 Хайчук Р. Основы создания терминологических словарей / Р.Хайчук, А.В. Зубов. – Белосток: Trans Human. 2004. – 129 с.

2 Калишевич, З. Е. Советская архивная терминология / З. Е. Калишевич, И. С. Назин, Л. Л. Смоктунович, Л. Г. Сирченко // Советские архивы. 1969. № 1. – С. 30-35.

3 Duchein, M. A propos de la terminologie archivistique / M. Duchein // La Gazette des Archives. 1965. № 50. – P. 156-161.

4 Свидетельство о включении энциклопедического электронного web-словаря современной терминологии «InfoLio&Res» в Государственный регистр информационных ресурсов Республики Беларусь от 12.01.2009 № 1100900613

5 Беляева, Л. Н. Автоматизированное рабочее место филолога: принципы организации и компонентная структура / Л.Н. Беляева // Актуальные проблемы прикладной лингвистики. Сб. научных трудов. В 2-х ч. Ч. 2. / ред. А. В. Зубов. – Минск: МГЛУ, 2008. – 214 с.

6 Гринев-Гринеvич, С. В. Введение в терминографию / С.В. Гринев-Гринеvич. Изд. 3-е, доп. – М.: «Либроком», 2009. – 224 с.

Н. В. Супрунчук

## ИСПОЛЬЗОВАНИЕ СИСТЕМЫ «E-UNIVERSITY» В ПРЕПОДАВАНИИ ЯЗЫКОЗНАНИЯ

Сегодня в нашей стране акцентируется внимание на индивидуализацию образовательного процесса, на внимательное отношение к каждому обучаемому. В средней школе этому способствует снижение количества учеников в классе. В высшей школе, однако, количественная тенденция иная: соотношение «число преподавателей: число студентов» растёт. От советской нормы «1:8» мы движемся к западноевропейской норме «1:30». Скажем, в июне 1941 в БГУ было 1337 студентов и 60 аспирантов и 148 профессоров и преподавателей [5] (соотношение «1:9,4»). В 2007 году в БГУ было более 24 тысяч студентов и 2677 преподавателей, т. е. соотношение «1:9». В настоящее время в БГУ обучаются 11 докторантов, свыше 750 аспирантов, 28,5 тыс. студентов, профессорско-преподавательский состав насчитывает 2477 человек [3], т. е. коэффициент «1:11,8».

Диверсификация образования, рост количества знаний, подлежащих усвоению, ведёт к появлению новых дисциплин и сокращению учебного времени на каждую из этих дисциплин. Провозгласив лозунг «Обучение через всю жизнь», активно развивая систему послевузовского и дополнительного образования, следует решать задачу «научить студента учиться», т. е. в ближайшее время студент должен будет меньше знаний приобретать на лекции, а больше – самостоятельно. Оставляя в стороне целесообразность акцента на самостоятельной учёбе (он не так уж бесспорен), обратимся к мощному средству оптимизации этого направления.

Одним из средств решения указанного противоречия (рост числа студентов vs. требование индивидуализировать образовательный процесс) является широкое использование в обучении компьютерной техники. В БГУ с 2007 года внедряется сетевая образовательная платформа «e-University» (СОП). «СОП «e-University» – комплекс программных средств для организации всех форм непрерывного образования, включая дистанционное, тестирования знаний и переподготовки кадров посредством Интернет/Интранет технологий в различных учреждениях системы образования независимо от их направленности и количества обучаемых» [1].

Задачи этой системы: а) организация дистанционного образования; б) сочетание традиционных педагогических методов с новейшими коммуникационными и мультимедийными технологиями; в) оптимизация традиционного учебного процесса (эффективное взаимодействие преподавателей и учащихся в удобное для каждого время; самостоятельная подготовка студентов; тестирование и автоматическая оценка знаний; контроль организации обучения и его эффективности); г) «корпоративное образование: обучение и тестирование квалификационных знаний персонала и соискателей на работу; поддержка концепции непрерывного образования и

повышения квалификации; эффективное взаимодействие удалённых филиалов» [9].

Требования к ресурсам для установки СОП e-University: а) сервер: процессор P4 HT 3,3 GHz и выше; оперативная память не менее 1024 Mb; ёмкость НМД не менее 80 Gb; операционная система Windows NT/2000/XP или Linux и выше; Microsoft Internet Explorer 6 и выше; MySQL 2.23 и выше; Apache Jakarta Tomcat 5.5.17 и выше; JDK 1.5.0 и выше; б) рабочее место клиента: процессор PIII, 600Mhz и выше; оперативная память не менее 128Mb; операционная система Windows 95/98/NT/2000/XP и выше; Microsoft Internet Explorer 6 и выше <sup>1</sup>.

В системе предусмотрены следующие роли: администратор, разработчик, консультант, экзаменатор, обучающийся.

Таблица 1 - Функции участников СОП «e-University»

Участник	Функции
Администратор	Управляет учебным процессом и пользователями
Преподаватель-разработчик	Формирует библиотеку учебных ресурсов, формирует и наполняет структуру учебных курсов, создаёт тесты, задания, итоговые тесты и систему оценок
Преподаватель-консультант	Курирует обучение, консультирует обучаемых в режиме реального времени, ведёт регулярную переписку с обучаемыми посредством доски объявлений и по электронной почте, рекомендует учебные ресурсы для дополнительного изучения
Преподаватель-экзаменатор	Курирует прохождение итоговых тестов; может изменять время начала и окончания прохождения итогового теста и назначать коды доступа для обучаемых группы
Обучающийся	Обучается и тестируется
Наблюдатель	Может осуществлять контроль за учебным процессом и пользователями
Гостевой пользователь	Имеет доступ к общедоступной информации, включая открытые тесты для самопроверки

На сайте БГУ [euniversity.bsu.by](http://euniversity.bsu.by) представлены разнообразные материалы для учебного процесса. Электронный учебный курс имеет модульную систему. Каждый учебный модуль должен состоять из аннотации, указаний по изучению, учебных материалов, списка литературы, тестов и заданий. Доступ к учебным материалам можно предусмотреть в рамках учебного модуля и/или в рамках учебника. «Основу образовательного процесса при дистанционном обучении составляет целенаправленная и контролируемая самостоятельная работа обучаемого, который может учиться в удобном для себя месте, по индивидуальному расписанию, имея при себе комплект специальных средств обучения и согласованную возможность контакта с преподавателем» (см. «Общее руководство преподавателя» СОП [4, с. 4]). В конечном итоге учебный курс должен включать в себя следующие

<sup>1</sup> В браузерах Google Chrome, Mozilla Firefox, Opera работают почти все функции СОП.

компоненты: аннотацию курса (дисциплины), программу курса, программу дисциплины в целом, учебник, учебные модули, итоговые тесты, список литературы и систему оценок. Таким образом, данная система может решать две задачи при преподавании: тренингово-обучающую и оценивающую.

С 2008 года нами разрабатываются учебные материалы по предметам «Введение в языкознание», «Общее языкознание», «Сербский язык: практикум» и «Теоретическая грамматика сербского языка». Остановимся на первых двух дисциплинах. В настоящее время закрепление теоретических знаний и отработку лингвистических умений можно вести по разным пособиям, начиная с классического задачника И. А. Бодуэна де Куртенэ, переизданного в 2004 году (см. о нём статью В. М. Алпатова [2]). Активно применяются сборники задач и упражнений Б. Н. Головина, И. П. Мучника, Л. Р. Зиндера, С. И. Калабиной, в Минске недавно вышла книга И. Г. Шкрабы «Лингвистические задачи по белорусскому языку» [11], автор которой на белорусском материале старается добиться тех же целей, что и организаторы знаменитых московских олимпиад по лингвистике и математике.

В БГУ практикум по языкознанию строится на основе задачника профессора Б. Ю. Нормана 1989 г., переработанное издание которого вышло в Москве в 2006, 2008 и 2009 гг. [6; 7]. Работа с пособием значительно облегчилась после размещения его электронного варианта на сайте кафедры теоретического и славянского языкознания БГУ [www.kateosia.com](http://www.kateosia.com). В этой книге собраны задачи практически по всем темам вузовского языкознания, поэтому она является хорошим подспорьем при организации усвоения данного курса. Вместе с тем многие задачи и по содержанию, и по форме не годятся для применения в СОП «e-University» в силу специфики последней и теста как формы учебного задания. Недостатки формы задач и сборника Б. Ю. Нормана в том, что в рассматриваемой электронной системе возможно применение тестовых заданий на выбор, подстановку или заданий с конкретными ответами. А многие задачи из сборника требуют развёрнутого ответа, комментирования. Суть в них не в ответе, который иногда и нельзя сформулировать однозначно или который занимает много времени или места, а в самом подходе, в процессе решения. Ограничения содержательного плана связаны с тем, что многие задачи сборника, к сожалению, слишком сложны для большинства студентов. Поэтому при построении модулей мы осторожно отбирали задачи из пособия Б. Ю. Нормана, а выделенные – почти всегда видоизменяли для использования в СОП «e-University».

Для иллюстраций типов заданий в СОП «e-University» приведём некоторые подготовленные нами задачи.

1. *Задание закрытого типа (самый частый тип)*: Определите, какая функция языка реализуется в следующем высказывании:

*Ой, дай, Божа, добры час,  
Як у людзей, так і ў нас.  
Вясёлая гадзіна,  
Весьляся, радзіна*

*Весьляся, радзіна,  
Каб і жыта радзіла,  
Каб і жыта і авёс,  
Весьляся, каб вырос.*

2. *Задание на установление правильной последовательности:* Расположите, пожалуйста, системы письма в хронологическом порядке появления: шумерская клинопись, китайская иероглифика, греческий алфавит, арабское письмо, кириллица, латиница.

3. *Вопрос с фиксированным ответом:* а) Определите, пожалуйста, степень аналитичности высказывания: *Галава яго працуе: гумку жуе*; б) Запишите, пожалуйста, какие русские словоформы записаны в фонетической транскрипции [стьшкá], [прлс'блѣк]. (Запишите подряд, без пробела, например: *студелать* или *музыкальныйпереулок*.)

4. *Задание на соответствие:* Отметьте, пожалуйста, понятие и соответствующее ему определение: ??? агглютинативные языки, аналитические языки, синтетические языки: – языки, в которых, главным образом, каждый аффикс выражает одно значение и легко различимы морфемные границы; – языки, изобретённые человеком для целей международного общения; – языки, в которых ударение всегда падает на один и тот же слог в слове.

Важное требование к тесту – его целостность, внутренняя согласованность заданий в нём. Выполнить это требование помогают следующие рекомендации [10, с. 50]: «ориентированность теста и тестовых заданий на диагностично заданные параметры цели образования: объём усвоения содержания (наименование и число учебных элементов (УЭ)) и уровень усвоения каждого УЭ; конкретизация уровня усвоения – четкое выделение знаний, умений и навыков по заданному уровню усвоения изучаемого УЭ – необходимое условие для наиболее точного (по возможности) применения каждого тестового задания и компоновки заданий для теста в целом; учёт возможностей каждого вида задания, выступающего в определённой форме...; по каждому тестовому заданию должен быть составлен эталон – образец его полного и правильного решения».

При использовании СОП «e-University» следует иметь в виду следующие особенности: а) технические: возможен выход из строя компьютера, университетской сети, отсутствие электричества и пр.; особое внимание следует обратить на шрифты и кодировки; б) методические: необходимы очень точные, однозначные формулировки заданий, правила выбора алфавита при транскрибировании; иногда приходится значительно упрощать задание, предлагать только тот вопрос, на который возможен краткий и однозначный ответ. Занятия в СОП «e-University» предусматривают работу по сети, поэтому необходимо следить, чтобы во время контрольных работ обучаемые не использовали Интернет (для поиска ответа, наведения справок, для консультаций через чат или электронную почту), кроме специально предусмотренных случаев.

Применение СОП «e-University» помогает интенсифицировать и дифференцировать учебный процесс, повышая тем самым его эффективность и результативность. Данная система благоприятно сказывается на психологической атмосфере занятий. Учащийся может отвечать свободно, в удобном ему



порядке располагая вопросы, он не стесняется одноклассников. Педагог же не утомляется от проверки однообразных заданий, от одинаковых вопросов и долгого обдумывания ответов учащимися. Преподаватель избавляется от значительной части рутинной работы, которую выполняет ЭВМ, и превращается в творца и консультанта. Повышается наглядность обучения и его результатов, оценивание становится более понятным, «прозрачным».

#### СПИСОК ЛИТЕРАТУРЫ

1 ИВА приняла участие в работе семинара для специалистов по информатизации системы профессионально-технического образования [Электронный ресурс]. – Режим доступа: [http://belarus.iba.by/iba\\_web/main.nsf/news/ru.news17\\_09\\_2010.html](http://belarus.iba.by/iba_web/main.nsf/news/ru.news17_09_2010.html). – Дата доступа: 26.11.2010.

2 Алпатов, В. М. «Введение в языковедение» И. А. Бодуэна де Куртенэ [Электронный ресурс]. – Режим доступа: <http://urss.ru/cgi-bin/db.pl?lang=Ru&blang=ru&page=Book&id=109792>. – Дата доступа: 26.11.2010.

3 БГУ сегодня [Электронный ресурс]. – Режим доступа: <http://www.bsu.by/ru/main.aspx?guid=1841>. – Дата доступа: 26.11.2010.

4 Демо-сайт e-University (для учебных заведений) [Электронный ресурс]. – Режим доступа: <http://eun.iba.by>. – Дата доступа: 26.11.2010.

5 Краткая история БГУ [Электронный ресурс]. – Режим доступа: <http://www.bsu.by/ru/main.aspx?guid=1891>. – Дата доступа: 26.11.2010.

6 Норман, Б. Ю. Лингвистические задачи / Б. Ю. Норман. – М.: Флинта, 2009 (2008, 2006).

7 Норман, Б. Ю. Сборник задач по введению в языковедение / Б. Ю. Норман. – Минск: Выш. шк., 1989.

8 Руководство по разработке тестовых заданий и конструированию педагогических тестов / авт.-сост. Ю.Э. Краснов [Электронный ресурс]. – Режим доступа: <http://charko.narod.ru/tekst/metodiki/krasnov.pdf>. – Дата доступа: 26.11.2010.

9 Сетевая образовательная платформа e-University. – Режим доступа: [http://belarus.iba.by/iba\\_web/main.nsf/products/ru.software.euniversity.html](http://belarus.iba.by/iba_web/main.nsf/products/ru.software.euniversity.html) [Электронный ресурс]. – Дата доступа: 26.11.2010.

10 Трофимова, З. П. Основы методологии и методики построения педагогических тестов / З. П. Трофимова. – Минск: РИВШ, 2005.

11 Шкраба, І. Р. Лінгвістычныя задачы па беларускай мове / І. Р. Шкраба. – Мінск: ТэтраСістэмс, 2010.

С. В. Телепень

#### ЛАТИНСКИЙ ЯЗЫК В ИНТЕРНЕТЕ

Процесс информатизации охватил и старейшую область филологического знания – языковедение в области классической филологии. Почти сразу возникновение соответствующих интернет-ресурсов оказалось сопряженным с возникшим еще в середине XX в. движением за возрождение живой латыни. При всей фантастичности данной идеи она имеет существенные перспективы быть хотя бы отчасти реализованной. Действительно, историческая роль латинского языка как международного языка науки и художественного слова существенно отличает его от многочисленных искусственных языков, предлагавшихся для международного общения, – как от тех, которые получили хотя бы ограниченное распространение, так и от несравненно большей их части,

оставшейся мертворожденными проектами. Долгое время уже после возникновения движения за живую латынь не существовало такого действенного механизма реализации идеи возрождения латинского языка как, например, Интернет. В связи с этим до 80-х гг. XX в. шло по сути накопление интеллектуальных сил. К середине XX в. таким образом возникла необходимость осмысления перспектив возрождения живой латыни. Так был организован 1-й Международный конгресс живой латыни, который был проведен при поддержке французского министерства народного образования, университета Эсан-Прованс и муниципалитета г. Авиньон 3–6 сентября 1956 г. и собрал свыше 200 делегатов из 22 стран. Значительное место в работе конгресса занимали вопросы методики преподавания латинского языка. Общей чертой посвященных этим вопросам докладов и выступлений в прениях явилось стремление упростить преподавание грамматики и сосредоточить внимание на усвоение учащимися практических навыков владения латинской речью. Были намечены пути обогащения латинской лексики в соответствии с требованиями современной науки [1, р. 134–136].

Конгресс обратился к редакциям научных журналов с предложением сопровождать публикуемые ими статьи латинскими резюме. Прямым продолжением 1-го Международного конгресса живой латыни были конгрессы, состоявшиеся в 1959, 1963, 1969 и 1975 гг. (Лион, Страсбург, Авиньон, По).

Более широкие задачи были намечены по инициативе Института римских исследований Международным конгрессом содействия латинской образованности и латинскому языку (Рим, апрель 1966 г.). Важнейшим результатом конгресса было учреждение при Институте римских исследований постоянной комиссии, служащей организационным центром для дальнейшей разработки вопросов, связанных с местом латинского языка в современной культуре. В 1969 г. эта комиссия была преобразована в Международную академию содействия латинской образованности, куда вошли представители 26 стран Европы, Северной и Южной Америки, Азии и Африки, в том числе и представитель СССР. Серьезный в культурно-историческом отношении подход к живой латыни нашел отражение на следующем за римским – международном конгрессе, который состоялся в Бухаресте осенью 1970 г.

Следующий международный латинский конгресс состоялся в 1973 г. на острове Мальта. На нем были одобрены все решения предыдущих конгрессов, а также принято обращение к министерствам народного образования всех стран с рекомендацией, касающейся преподавания латинского языка: «Чтобы сохранить и наполнить действенной силой то наследие гуманизма, которое народы Европы и, более того, всего мира восприняли от Рима, просим министерства народного образования всех стран поддерживать, укреплять и развивать преподавание латинского языка в средней школе, университетах и других высших учебных заведениях и озаботиться тем, чтобы в этом преподавании учащимся сообщались сведения об истории римского народа, о его юриспруденции, политических учреждениях общественном укладе,

повседневной жизни и т. д., приводящие их к пониманию того, какое значение имели и имеют латинский язык и латинская литература для развития языков и литератур современной Европы, для всей нашей гуманитарной культуры» [2, р. 80]. В том же направлении шла работа международных конгрессов содействия латинской образованности, состоявшихся в Дакаре (1977 г.) и в Трире (1981 г.).

Практика применения латинского языка в качестве международного языка науки, а также как языка литературного творчества, продолжающего старую гуманистическую традицию характеризует динамику этого движения за живую латынь. Важным здесь является наличие журналов, публикующих научные статьи на латинском языке. Замечателен в этом смысле пример нидерландского филологического журнала «*Mnemosyne*», который до недавнего времени печатал статьи только на латинском (теперь и на немецком, английском, французском и итальянском языках). Многие журналы, как, например, варшавский «*Meander*», публикуя основной текст статей на польском языке, как правило сопровождают их латинским резюме. Оксфордский научно-популярный журнал «*Greece and Roma*», издающийся на английском языке, почти в каждом номере предоставляет место для публикации латинских стихов. Наконец, на латинском языке издаются журналы, специально посвященные живой латыни: «*Latinitas*» (Ватикан), «*Palaestra Latina*» (Барселона), «*Vox Latina*» (Саарбрюккен), «*Vita Latina*» (Авиньон), «*Orbis Latinus*» (Мендоса, Аргентина). Особо хотелось бы отметить лингвистическо-исторический журнал «*Emerita*» (Мадрид). Все названные журналы имеют страницы либо сайты в Интернете. В этих журналах систематически публикуются результаты ведущейся в разных странах работы по нормированию новолатинской лексики. Критический свод этих материалов в виде фундаментального словаря составляет одну из важнейших задач, включенных в программу деятельности Международной академии содействия латинской образованности.

Настоящим прорывом стало появление специального сайта, созданного энтузиастами из разных стран и посвященного пропаганде идеи живой латыни «*Lingua Latina Aeterna*» [3]. На страницах сайта имеется «Каталог сайтов», с помощью которого любой посетитель может сам добавить любую ссылку, каталог уже содержит 85 ссылок на латинские ресурсы мира. Самой новой частью сайта является war-версия портала (версия для мобильного телефона).

Однако не менее важен факт возникновения сайтов, представляющих собой общедоступные электронные ресурсы для занятия изысканиями в области языка и истории античности и средневековья. Наиболее важной представляется, здесь «Латинская библиотека» [4] и «Классическая страница» [5]. Здесь содержатся практически все основные тексты латинских авторов периода античности и средневековья. Кроме того здесь же имеются ссылки, позволяющие выйти на другие сайты, содержащие как латиноязычный нарратив, так и новые эпиграфические тексты. В отношении последних наиболее важным представляется немецкий сайт «Корпус латинских надписей» («*Corpus Inscriptionum Latinarum*») [6]. Здесь помещены не только все тома из-

дающегося с 1863 г. СЛ, но также публикуются новые эпиграфические материалы, но и помещаются аналитические работы по результатам аналитической работы над новым эпиграфическим материалом. Показательно, что все чаще эти статьи выходят на латыни. Таким образом происходит соединение процесса расширения доступа исследователей и энтузиастов к латиноязычным ресурсам и процесса внедрения идеи возрождения живой латыни в качестве языка науки.

#### СПИСОК ЛИТЕРАТУРЫ

1. Premier Congres International pour le Latin vivant. Avignon, 3–6 septembre 1956. – Avignon: Edouard Aubanel, 1956. – 244 p.
2. Vox Latina. – 1974. – № 35. – С. 79–88.
3. Lingua Latina Aeterna [Электронный ресурс]. – Режим доступа: <http://wap.linguaeterna.com/> – Дата доступа: 10.09.2010.
4. The latin library [Электронный ресурс]. – Режим доступа: <http://www.thelatinlibrary.com/> – Дата доступа: 12.09.2010.
5. The Classics Page [Электронный ресурс]. – Режим доступа: <http://www.thelatinlibrary.com/classics.html/> – Дата доступа: 20.09.2010.
6. Corpus Inscriptionum Latinarum [Электронный ресурс]. – Режим доступа: <http://cil.bbaw.de/> – Дата доступа: 22.09.2010.

Е. В. Убоженко

### АВТОМАТИЗИРОВАННЫЙ ПЕРЕВОД И СПОСОБЫ ЕГО ИСПОЛЬЗОВАНИЯ

В настоящее время особенно востребован поиск путей максимально автоматизировать процесс перевода, осуществляемого человеком, чтобы, с одной стороны, максимально облегчить нелегкий труд человека-переводчика, а с другой – сделать этот труд максимально эффективным. Осуществить подобное можно лишь в наибольшей степени интегрировав усилия специалистов в областях кибернетики, программирования, психологии, а главное – лингвистики. На сегодняшний день наиболее распространенными способами использования компьютеров при переводе являются работа со словарями и глоссариями, памятью переводов, содержащей примеры ранее переведенных текстов, а также использование так называемых корпусов, что дает сжатое описание того, как слова и выражения реально используются в языке в целом или в конкретной предметной области. Рассмотрим подробнее способы использования компьютеров при переводе.

1. Работа со словарями и глоссариями. Основной чертой, отличающей глоссарий от словаря, является его специализированность на определенной области знаний. Глоссарий (лат. *glossarium* – словарь, глосс) содержит толкование терминов, иногда перевод на другой язык, комментарии и примеры. Компания «ProMT» предлагает более 120 готовых специализированных словарей для переводчика «ProMT Professional». К процессу перевода можно подключить несколько словарей как специализированных, так и собственных, и в зависимости от тематики документа установить приоритеты их

использования при переводе. Также появилась возможность создания собственных словарей для достижения более точного перевода документов, содержащих узкоспециализированную терминологию. Пользователи, имеющие собственные глоссарии терминов, легко могут создать на их основе словари для системы перевода «ProMT Professional». В узких предметных областях при большом количестве исходных текстов и устоявшейся терминологии переводчики могут использовать и машинный перевод, который может обеспечить хорошее качество перевода терминологии и устойчивых выражений в узкой области. Переводчик в этом случае осуществляет постредактирование полученного текста.

2. Работа с памятью переводов. Память переводов (англ. *translation memory*), иногда называемая «накопителем переводов», – база данных, содержащая набор ранее переведенных текстов. Одна запись в такой базе данных соответствует сегменту или «единице перевода», за которую обычно принимается одно предложение (реже – часть сложносочиненного предложения либо абзаца). Если единица перевода исходного текста в точности совпадает с единицей перевода, хранящейся в базе (точное соответствие), она может быть автоматически подставлена в перевод. Новый сегмент может также слегка отличаться от хранящегося в базе (неточное соответствие). Такой сегмент может быть также подставлен в перевод, но переводчику необходимо будет внести определенные изменения. В каждой конкретной системе памяти переводов данные хранятся в своем собственном формате (текстовый формат в Wordfast, база данных Access в Deja Vu), но существует международный стандарт TMX (англ. *Translation Memory eXchange format*), который основан на XML (англ. *eXtensible Markup Language* – расширяемый язык разметки) и который может генерироваться практически всеми системами памяти переводов. Большинство систем памяти как минимум поддерживают создание и использование словарей пользователя, создание новых баз данных на основе параллельных текстов, а также полуавтоматическое извлечение терминологии из оригинальных и параллельных текстов.

3. Использование лингвистических корпусов. Лингвистическим корпусом называют совокупность текстов, собранных в соответствии с определенными принципами, размеченных по определенному стандарту и обеспеченных специализированной поисковой системой. Целесообразность создания текстовых корпусов объясняется:

- 1) представлением лингвистических данных в реальном контексте;
- 2) достаточно большой представительностью данных (при большом объеме корпуса);
- 3) возможностью многократного использования единожды созданного корпуса для решения различных лингвистических задач.

Наличие большого количества текстов в электронной форме существенно облегчило задачу создания больших представительных корпусов размером в десятки и сотни миллионов слов, но не ликвидировало проблем: сбор тысяч текстов, снятие проблем с авторскими правами, приведение всех текстов в единую форму, балансировка корпуса по темам и жанрам отнимают много времени.

## СВЕДЕНИЯ ОБ АВТОРАХ

**Аматов Александр Михайлович**, доктор филологических наук, профессор, (Белгородский государственный университет, Белгород, Россия); amatov@bsu.edu.ru

**Андреев Сергей Николаевич**, доктор филологических наук, профессор (Смоленский государственный университет, Смоленск, Россия); smol.an@mail.ru

**Andrzejczuk Anna – Анджейчук Анна**, докторант (Институт польского языка Польской академии наук; Варшава, Польша); anna.andrzejczuk@gmail.com

**Асенчик Елена Федоровна**, ассистент (Гомельский государственный университет имени Ф. Скорины); elena-assenchik@rambler.ru

**Боркович Александр Аркадьевич**, кандидат филологических наук (Белорусский государственный университет); albark@tut.by

**Бойка Валянцін Юр'евіч**, аспирант (Мазырскі дзяржаўны педагагічны ўніверсітэт імя І. П. Шамякіна); v\_bojka@yahoo.com

**Богоедова Татьяна Николаевна**, кандидат филологических наук, доцент (Гомельский государственный университет имени Ф. Скорины); bogoedova@yandex.ru

**Брадзіхіна Ала Васільеўна**, кандыдат філалагічных навук, дацэнт (Гомельскі дзяржаўны ўніверсітэт імя Ф. Скарыны);

**Бусел Татьяна Викторовна**, преподаватель (Минский государственный лингвистический университет);

**Бутякова Ирина Игоревна**, старший преподаватель (Белорусский государственный университет транспорта); wolskaja@mail.ru

**Важнік Сяргей Аляксандравіч**; кандыдат філалагічных навук, дацэнт (Беларускі дзяржаўны ўніверсітэт); waznik@yandex.ru

**Валодзіна Аляксандра Вадзімаўна**, студэнтка 4 курса гістарычнага факультэта (Беларускі дзяржаўны ўніверсітэт); sushka77@rambler.ru

**Галло Ян**, кандидат филологических наук, старший преподаватель (Университет Константина Философа, Нитра, Словакия); jgallo@ukf.sk

**Гильмуллин Ринат Абрекович**, кандидат физико-математических наук, заведующий отделом НИИ «Прикладная семиотика» Академии наук Татарстана (Казань, Россия);

**Головня Анастасия Ивановна**, кандидат филологических наук, доцент (Белорусский государственный университет); GolovnjaAI@bsu.by

**Гомонова Инна Геннадьевна**, ассистент (Гомельский государственный университет имени Ф. Скорины); inna-gomonova@mail.ru

**Говорушкина Наталья Андреевна**, ассистент кафедры автоматизированных систем обработки информации (Гомельский государственный университет имени Ф. Скорины);

**Давыденко Ирина Тимофеевна**, магистрант (Белорусский государственный университет информатики и радиоэлектроники);

**Егоров Николай Николаевич**, кандидат физико-математических наук, доцент (Мозырский государственный педагогический университет имени И. П. Шамякина);

**Елисеева Ольга Евгеньевна**, кандидат технических наук, доцент (Белорусский государственный университет); olae@open.by

**Ерашко Наталья Владимировна**, студентка 4 курса филологического факультета (Гомельский государственный университет имени Ф. Скорины);

**Житко Владимир Александрович**, магистрант (Белорусский государственный университет информатики и радиоэлектроники); minsk.nes@gmail.com

**Жулего Александр Владимирович**, докторант (Варшавский университет, Варшава, Польша); alexjoul@gmail.com

**Зубов Александр Васильевич**, доктор филологических наук, профессор (Минский государственный лингвистический университет); proscien@mslu.by

**Зубова Ирина Ивановна**, доцент (Минский государственный лингвистический университет); proscien@mslu.

**Киеня Светлана Николаевна**, лаборант (Гродненский государственный университет имени Янки Купалы);

**Кобыш Инна Ивановна**, кандидат филологических наук, доцент (Гомельский государственный университет имени Ф. Скорины); ANK-78@mail.ru

**Коваль Владимир Иванович**, доктор филологических наук, профессор (Гомельский государственный университет имени Ф. Скорины); vlad-kov@mail.ru

**Капылоў Ігар Лявонавіч**, кандыдат філалагічных навук, загадчык аддзела лексікалогіі і лексіграфіі, вучоны сакратар Інстытута мовы і літаратуры імя Я. Коласа і Я. Купалы НАН Беларусі; iharkap@gmail.com

**Кошчанка Уладзімір Аляксандравіч**, малодшы навуковы супрацоўнік Інстытута мовы і літаратуры імя Я. Коласа і Я. Купалы НАН Беларусі;

**Кривонос Екатерина Александровна**, доцент (Белорусский государственный университет); kerris@inbox.ru, rki@bsu.by

**Крохмальник Антон Юрьевич**, аспирант (Гомельский государственный университет имени Ф. Скорины);

**Крупа Людмила Федоровна**, аспирант (Мозырский государственный педагогический университет имени И. П. Шамякина); text2005@mail.ru

**Кураш Сергей Борисович**, кандидат филологических наук, доцент (Мозырский государственный педагогический университет имени И. П. Шамякина); kurash-persmail@tut.by

**Лапицкая Надежда Ивановна**, кандидат филологических наук, доцент (Гомельский государственный университет имени Ф. Скорины);

**Малинина Юлия Александровна**, аспирант (Гродненский государственный университет имени Янки Купалы); luczik@mail.ru

**Махина Александр Иосифович**, студент факультета информационных технологий и управления (БГУИР);

**Миклашевич Анна Генриховна**, магистрант (Гродненский государственный университет имени Янки Купалы); sag030288@yandex.ru

**Ничипорчик Елена Владимировна**, кандидат филологических наук, доцент (Гомельский государственный университет имени Ф. Скорины); evnich@gmail.com

**Ничипорчик Алексей Иванович**, студент 4 курса факультета прикладной математики и информатики (Белорусский государственный университет);

**Падажецкая Юлия Михайловна**, студентка 5 курса филологического факультета (Гомельский государственный университет имени Ф. Скорины);

**Паршикова Елена Владимировна**, магистрант (Мозырский государственный педагогический университет имени И. П. Шамякина); bagira1290@list.ru

**Половец Маргарита Владимировна**, магистрант (Гродненский государственный университет имени Янки Купалы); mail@grsu.by, magy@tut.by

**Палуян Алена Мікалаеўна**, кандыдат філалагічных навук, дацэнт (Гомельскі дзяржаўны ўніверсітэт імя Ф. Скарыны),

**Палуян Сяргей Сяргеевіч**, студэнт 4 курса матэматычнага факультэта (Гомельскі дзяржаўны ўніверсітэт імя Ф. Скарыны);

**Ревуцкий Олег Викторович**; кандидат филологических наук, доцент (Мозырский государственный педагогический университет имени И. П. Шамякина);

**Рогозина Елена Андреевна**, старший преподаватель (Санкт-петербургский государственный университет, Санкт-Петербург, Россия); Renehorn.r@gmail.com

**Рычкова Людмила Васильевна**, кандидат филологических наук, доцент (Гродненский государственный университет имени Янки Купалы); l.v.rychkova@mail.ru

**Савченко Любовь Васильевна**, Дослідження етнокультурної етимологізації фразеологізмів через інтернет-ресурси; заведуюча кафедрою культури українського мови, к.ф.н., доцент, докторант (Тавричеський національний університет імені В. І. Вернадського, Сімферополь, Україна); lubov29@bk.ru

**Савчук Светлана Олеговна**, старший науковий співробітник, кандидат філологічних наук, доцент (Інститут російського мови імені В. В. Виноградова Російської академії наук, відділ корпусної лінгвістики і лінгвістическої поезії, Москва, Росія); savsvetlana@mail.ru

**Самодумкин Сергей Александрович**, асистент (БГУИР); samodumkin@bsuir.by

**Серикова Ирина Валерьевна**, асистент (Гомельський державний університет імені Ф. Скорины);

**Силков Сергей Васильевич**, член Російської гільдії управляючих документацією; SILKOV\_S\_V@mail.ru

**Сичинава Дмитрий Владимирович**, кандидат філологічних наук, науковий співробітник (відділ лінгвістических досліджень Всеросійського інституту наукової і технічної інформації РАН, Москва, Росія); mitrius@gmail.com

**Сорока Сергей Иванович**, студент факультету інформаційних технологій і управління (Білоруський державний університет інформатики і радіоелектроніки);

**Станкевіч Аляксандра Аляксандраўна**, доктар філалагічных навук, прафесар (Гомельскі дзяржаўны ўніверсітэт імя Ф. Скарыны);

**Станкевич Алеся Юрьевна**, старший преподаватель (Гродненский державный университет имени Янки Купалы); a.stan.lab@gmail.com

**Стрижак Артем Леонидович**, аспирант (Гомельський державний університет імені Ф. Скорины); strizhak2011@yandex.ru

**Супрунчук Никита Викторович**, кандидат філологічних наук, доцент (Білоруський державний університет); suprunchuk@mail.ru

**Телепень Сергей Валерьевич**, кандидат історических наук, доцент (Мозьрський державний педагогічеський університет імені І. П. Шамякіна); telepen\_serg@tut.by

**Убоженко Елена Владимировна**, студентка 5 курсу філологічеського факультету (Гомельський державний університет імені Ф. Скорины);

**Фоминых Татьяна Евгеньевна**, магістрант (Гомельський державний університет імені Ф. Скорины); t.fominykh@hotmail.com

**Хакимов Булат Эрнстович**, ведучий науковий співробітник, кандидат філологічних наук (Державне бюджетне закладення «Прикладна семиотика» Академії наук Республіки Татарстан, Казань, Росія); khakeem@yandex.ru

**Шабулдаева Наталья Ивановна**, кандидат філологічних наук, доцент (Гомельський державний університет імені Ф. Скорины); shabuldaeva@mail.ru

**Шведова Мария Алексеевна**, кандидат філологічних наук (Інститут українського мови НАНУ, Київ, Україна);

**Шиманская Ольга Юрьевна**, кандидат філологічних наук (Закладення освіти Федерації профсоюзів Білорусі «Міжнародний інститут трудових і соціальних відносин»); shimans@tut.by

**Шукайлова Алина Геннадьевна**, студентка 4 курсу філологічеського факультету (Гомельський державний університет імені Ф. Скорины);

**Юдкин-Рипун Игорь Николаевич**, доктор мистецтвознавства, член-кореспондент Національної Академії Художеств України, ведучий науковий співробітник; заведуючий відділом театрознавства і культурології Інституту мистецтвознавства, фольклористики і етнології НАН України (Інститут мистецтвознавства, фольклористики і етнології НАН України, Київ, Україна); iyudkin@yandex.ru

**Юрина Елена Андреевна**, доктор філологічних наук, професор (Томський державний університет, Томск, Росія); yourina2007@yandex.ru.



## СОДЕРЖАНИЕ

### Компьютерная лингвистика: теория и практика. Содержание учебной дисциплины, ее становление и развитие

<i>Аматов А. М.</i> Машинная обработка текста и перевод: проблемы и поиск решений .....	3
<i>Зубов А. В.</i> Компьютерная лингвистика в Беларуси: эволюция задач .....	7
<i>Капылю́ І. Л., Кошчанка У. А.</i> Корпус тэкстаў як аснова аптымізацыі падрыхтоўкі фундаментальнага “Глумачальнага слоўніка беларускай мовы” .....	12
<i>Коваль В. И.</i> Компьютерная лингвистика в Гомельском университете: становление и перспективы .....	17
<i>Рычкова Л. В.</i> «Специалист по компьютерной филологии» – проблемы идентификации .....	21
<i>Савчук С. О.</i> Использование Национального корпуса русского языка в преподавании филологических дисциплин .....	26
<i>Сичинава Д. В., Шведова М. А.</i> Параллельные корпуса в составе национального корпуса русского языка: технологии и решаемые задачи ....	30
<i>Хакимов Б. Э., Гильмуллин Р. А.</i> Система морфологической разметки для корпуса татарского языка .....	34
<i>Юдкин-Рипун И. Н.</i> Дескрипторы идиоматики корпуса драматических текстов (методика тэггированных списков) .....	38

### Компьютерное обеспечение лингвистических исследований

<i>Андреев С. Н.</i> Многомерный анализ языковых и речевых данных .....	43
<i>Бойка В. Ю.</i> Стылістычная функцыя аказіяналізмаў у беларускім інтэрнэт-маўленні .....	46
<i>Бусел Т. В.</i> Структура системы автоматического порождения англоязычных деловых электронных писем .....	50
<i>Бутякова И. И.</i> Использование компьютерной программы «Wordstat» при полевом исследовании поэтического текста .....	53
<i>Важнік С. А.</i> Ветлівыя звароты ў беларускім сеціве .....	56
<i>Валодзіна А. В.</i> Выкарыстанне камп’ютарнага рэсурсу “Patrologia Latina” для даследавання канцэпту “Pauperes catholici” .....	60
<i>Говорушкіна Н. А.</i> Использование графической программы Corel Draw при верстке текстовых блоков .....	63
<i>Ерашко Н. В.</i> Лексико-семантическое поле «Современная женщина» по данным Интернет-источников .....	65
<i>Зубова И. И.</i> Формальное определение средств выражения неискренности в англоязычных высказываниях .....	68
<i>Крупа Л. Ф.</i> Устойчивая номинативная единица <i>мобильный телефон</i> и ее синонимический ряд в современной русской речи (по данным Интернет-источников) .....	71
<i>Ничипорчик Е. В., Ничипорчик А. И.</i> Использование метода семантической разметки в организации поиска по текстовым документам .....	74
<i>Паршикова Е. В.</i> Лексические особенности феминно-ориентированных рекламных текстов (на материале интернет-источников) .....	78

<i>Савченко Л. В.</i> Дослідження етнокультурної етимологізації фразеологізмів через інтернет-ресурси .....	81
<i>Станкевіч А. А.</i> З вопыту выкарыстання праграмнага камп'ютарнага забеспячэння пры даследаванні лексікі іншамоўнага паходжання .....	84
<i>Стрижак А. Л.</i> Использование ресурсов НКРЯ при изучении детерминологизации в современных публицистических текстах .....	88
<i>Фоминых Т. Е.</i> Феминные номинации в Национальном корпусе русского языка (диахронический аспект) .....	93
<i>Шукайлова А. Г.</i> Символика слова <i>заяц</i> в русской и китайской культурах (по данным интернет-источников) .....	97

### **Корпусная лингвистика: проблемы разработки, ведения и использования корпусов**

<i>Andrzejczuk A.</i> Korpusy językowe w pracy językoznawcy: nowe możliwości czy droga na manowce? .....	100
<i>Барковіч А. А.</i> Экстралінгвістычны кантэкст рэпрэзентатыўнасці беларускага корпуса тэкстаў .....	104
<i>Гомонова И. Г.</i> Исследование паремиологических единиц с использованием материалов Национального корпуса русского языка .....	108
<i>Жулего А. В.</i> Корпусы европейских языков: современное состояние и перспективы использования .....	112
<i>Крохмальник А. Ю.</i> Поисковый запрос как способ выявления языковых конструкций, репрезентирующих невербальную коммуникацию в художественных текстах .....	116
<i>Кураш С. Б., Ревуцкий О. И.</i> От метафоры корпуса к корпусам метафор ....	120
<i>Миклашевич А. Г.</i> Исследование семантики национально-маркированных единиц при помощи лингвистических корпусов письменных и устных текстов.....	126
<i>Рогозина Е. А.</i> Корпус агиографических текстов СКАТ: разметка элементов содержания .....	130
<i>Рычкова Л. В., Киеня С. Н.</i> Лингвистические корпусы в социокультурном аспекте .....	132
<i>Серикова И. В.</i> О мультимедийной представленности публицистического подкорпуса в Национальном корпусе русского языка .....	135
<i>Шабулдаева Н. И.</i> Об употреблении конструкции <i>хлеб-соль</i> (на материале Национального корпуса русского языка) .....	138

### **Прикладная лексикография.**

#### **Автоматизированные лексикографические системы**

<i>Житко В. А., Давыденко И. Т.</i> Естественно-языковой интерфейс вопросно-ответных систем .....	144
<i>Кобыш И. И.</i> Частотные словари и лингвистический анализ текста .....	148
<i>Малинина Ю. А.</i> Фрейм текста описания зámка для русского и английского языков (на материале дескрипторного словаря лексем) .....	152
<i>Падажецкая Ю. М.</i> Структурно-семантическая характеристика компьютерных эмоций .....	155

<i>Самодумкин С. А., Сорока С. И., Махина А. И.</i> Интеграция геоинформационных данных в интеллектуальных справочных системах .....	158
<i>Станкевич А. Ю.</i> Связность сочетаний квазилексем .....	162
<i>Шиманская О. Ю.</i> Использование метода базовых определений и семантических оппозиций при составлении электронного двуязычного словаря метафор .....	166
<i>Юрина Е. А.</i> Актуальные задачи компьютеризации диалектного архива говоров Среднего Приобья и пути их решения .....	170

### **Использование компьютерных технологий в учебно-методической деятельности**

<i>Асенчик Е. Ф.</i> Поиск информации в Интернете: некоторые аспекты обучения .....	176
<i>Богоедова Т. Н.</i> Использование ресурсов Национального корпуса русского языка в курсе русской диалектологии .....	180
<i>Брадзіхіна А. В.</i> Адметнасці выкарыстання мультымедыя-тэхналогій пры выкладанні курса «Гісторыя беларускай літаратуры 19 ст.» замежным студэнтам .....	183
<i>Галло Я.</i> К вопросу об электронном образовании (e-learning) .....	187
<i>Головня А. И.</i> Компьютерные экспертные системы в обучении РКИ .....	191
<i>Елисеева О. Е.</i> Сеть Интернет как инструмент и средство изучения прикладной и компьютерной лингвистики .....	196
<i>Кураш С. Б., Егоров Н. Н.</i> Структура и содержание учебной дисциплины «Компьютерная лингвистика» в системе подготовки будущих учителей русского языка и литературы .....	200
<i>Кривонос Е. А.</i> Возможности применения исследовательского корпуса русских текстов по экономике в лингвометодических целях .....	204
<i>Лапицкая Н. И.</i> Компьютерная презентация как вид наглядности на уроках русского языка .....	207
<i>Палуян А. М., Палуян С. С.</i> Выкарыстанне інфармацыйных тэхналогій пры правярцы ведаў студэнтаў .....	210
<i>Половец М. В.</i> Функциональный аспект лингвистических средств диалога в интернет-дидактике при обучении русскому языку как иностранному .....	214
<i>Силков С. В.</i> Построение терминологических информационных ресурсов по документальным процессам, терминоведению и поиску информации .....	217
<i>Супрунчук Н. В.</i> Использование системы "e-University" в преподавании языкознания .....	221
<i>Телепень С. В.</i> Латинский язык в Интернете .....	225
<i>Убоженко Е. В.</i> Автоматизированный перевод и способы его использования .....	228
<i>Сведения об авторах</i> .....	230

*Научное издание*

**Компьютерная лингвистика:  
научное направление и учебная дисциплина**

Сборник научных статей

В авторской редакции

Подписано в печать . Формат 60x84<sup>1/16</sup>.  
Бумага офсетная. Ризография. Усл. печ. л. .  
Уч.-изд. л. . Тираж 100 экз. Заказ № .

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.