

**Министерство образования и науки РФ
Научно-методический совет по иностранным языкам
Национальное общество прикладной лингвистики (НОПриЛ)
Факультет иностранных языков и регионоведения
МГУ имени М.В. Ломоносова
Центр Дистанционного Образования**

**Информационно-коммуникационные
технологии в лингвистике,
лингводидактике и межкультурной
коммуникации**

Сборник статей

Под редакцией А.Л. Назаренко

Выпуск 6

**Москва
«Университетская книга»
2014**

УДК 81'1(082.1)
ББК 81я43
И74

*Печатается по постановлению Ученого совета факультета
иностранных языков и регионоведения МГУ имени М. В. Ломоносова*

Рецензенты:

профессор, доктор педагогических наук *Е. М. Соловова*
профессор, доктор филологических наук *С. С. Хромов*

И74 Информационно-коммуникационные технологии в лингвистике, лингводидактике и межкультурной коммуникации. Вып. 6: сб. ст. / М-во образования и науки РФ [и др.]; под ред. А. Л. Назаренко. — М.: Центр Дистанц. Образования: Факультет иностр. яз. и регионоведения МГУ имени М. В. Ломоносова; Изд-во «Университетская книга», 2014. — 498 с.: табл., ил. ISBN 978-5-91304-571-3

Сборник включает материалы VI Международной научно-практической конференции по вопросам интеграции ИКТ в лингвистическое образование, состоявшейся в МГУ имени М.В. Ломоносова в 2014 г. Широкий тематический спектр и прагматическая направленность конференции позволяют использовать сборник в качестве учебного пособия по методически целесообразному использованию новейших технологий в лингвистических исследованиях, в преподавании иностранных языков и межкультурной коммуникации для создания обучающей среды, организации дистанционного, смешанного и мобильного обучения с использованием ресурсов и социальных сервисов Интернета, для проведения международных телекоммуникационных проектов, внедрения smart-технологий и др.

Издание адресовано студентам, аспирантам и преподавателям гуманитарных специальностей, а также широкому кругу читателей, интересующихся проблемами информатизации лингвистического образования.

УДК 81'1(082.1)
ББК 81я43

© Центр Дистанционного Образования,
ФИЯиР МГУ имени М.В. Ломоносова, 2014
© Коллектив авторов, 2014
© ИД «КДУ», 2014

ISBN 978-5-91304-571-3

Опыт использования метода проектов при проведении международных образовательных интернет-проектов по межкультурной коммуникации на факультете иностранных языков и регионоведения <i>Сизых Л.Г.</i>	448
--	-----

ГЛАВА X

ИКТ в лингвистических исследованиях

ИТ средства сопровождения научных лингводидактических исследований <i>Иванова Л.И.</i>	458
Проектирование сервиса для создания и обработки текстовых массивов, создаваемых в процессе усвоения английского языка <i>Иголкина Н.И., Аветисян С.Ю.</i>	463
Этика научного текста: опыт корпусного исследования <i>Нохрин А.В.</i>	471
Терминологическое гнездо Classroom в предметной области КЛД <i>Половец М.В.</i>	481
Многоцелевой лингвистический корпус региональных СМИ <i>Рычкова Л.В., Станкевич А.Ю.</i>	488

Многоцелевой лингвистический корпус региональных СМИ¹

Рычкова Людмила Васильевна

кандидат филологических наук, доцент,
филологический факультет ГрГУ им. Янки Купалы,
Гродно, rychkova@grsu.by

Станкевич Алеся Юрьевна

филологический факультет ГрГУ им. Янки Купалы,
Гродно, a.stan.lab@gmail.com

Peculiarities of the illustrative linguistic corpus created on the material of the authentic texts of mass media of the Grodno region (mainly Russian-speaking historically multiethnic region of Belarus bordering with both Poland and Lithuania), as the first corpus within the new module of the National corpus of Russian – that is the corpus of the regional and foreign press are presented. Multiple directions of its use both for the research and didactic purposes are shown.

Представлены особенности иллюстративного лингвистического корпуса, созданного на материале аутентичных текстов СМИ Гродненщины (преимущественно русскоязычного исторически полиэтничного региона Беларуси, граничащего с Польшей и Литвой), как первого корпуса в рамках нового модуля Национального корпуса русского языка – корпуса региональной и зарубежной прессы. Показаны возможности его применения для различных исследовательских и дидактических целей.

Исследования региональных и глобальных вариантов языков получили объективную фактологическую базу в виде аутентичных электронных языковых ресурсов, среди которых особое место занимают лингвистические, то есть метатекстово и собственно лингвистически размеченные, корпуса. Широко известным, к примеру, является Международный корпус английского языка – ICE (<http://ice-corpora.net/ice/index.htm>), первичной целью создания которого было собрание материала для дальнейшего сравнения различных вариантов английского языка (collecting material for comparative studies of English worldwide). Создание этого корпуса,

¹ Подготовлено в рамках проекта, реализуемого при поддержке Белорусского республиканского фонда фундаментальных исследований (договор № Г13Р-050 от 16.04.2013 г.).

основанного на единой концепции и использующего общие принципы разметки, независимо от того, где именно и на каком англоязычном материале формируется подкорпус, позволило не только обеспечить «внутриязыковые» сопоставительные исследования объективной фактологической базой, но и активизировать исследования самих региональных и национальных вариантов английского языка.

Образование на постсоветском пространстве новых независимых государств результировало не только в формировании в них новой культуры государственности, но и в изменении отношения к русскому языку и/или его статусу. В любом случае, новая геополитическая языковая ситуация актуализировала проблему новых состояний русского языка, отличных от его былого идеологически поддерживаемого статуса «великого языка великого советского народа». Последний, обслуживая сознательно формируемую советской политической элитой (читай – властью) надэтническую общность, должен был быть единым для всех. Отсюда – такое внимание к кодифицированной форме языка, которая возводилась в абсолют, и, будучи, по сути, формой искусственной, часто подменяла (в том числе и в лингвистических исследованиях) понятие языка, естественного, «живого», языкового материала, отражающего фактический узус, включая обобщение реальных практик использования литературного стандарта. Процессы дивергенции региональных и национальных вариантов русского языка и кодифицированного языка метрополии отмечаются сегодня многими лингвистами как в России, так и за ее пределами (Г.И.Кустова², С.О. Савчук, А. Мустайоки³ и др.). Вполне ожидаемым в свете сказанного было развитие Национального корпуса русского языка – НКРЯ (www.ruscorgora.ru) путем включения в него нового модуля, отражающего региональные варианты русскоязычного узуса.

Первым в новом модуле НКРЯ, получившим название корпуса региональной и зарубежной прессы, стал иллюстративный лингвистический корпус на материале аутентичных текстов СМИ Гродненщины – преимущественно русскоязычного исторически

² Кустова Г.И., Савчук С.О. Изучение лексико-семантической и социокультурной специфики русской речи на территории Республики Беларусь (на материале текстов СМИ) // Труды международной конференции «Корпусная лингвистика – 2013». – СПб.: С.-Петербургский гос. университет, Филологический факультет, 2013. – С. 344 – 352.

³ Мустайоки А. Разновидности русского языка: анализ и классификация // Вопросы языкознания. – 2013. – № 5. – С. 3–27.

полиэтнического региона Республики Беларусь, граничащего с Польшей и Литвой. Корпус СМИ Гродненщины является частью белорусско-русского научного проекта, направленного на изучение лексико-семантической и социокультурной специфики русской речи данного региона. Проект осуществляется совместно учеными Института русского языка им. В.В. Виноградова РАН (при поддержке РГНФ) и кафедры общего и славянского языкознания Гродненского государственного университета им. Янки Купалы (при поддержке БРФФИ). Выбор именно текстов СМИ не случаен – ведь именно они наиболее быстро реагируют на социокультурные изменения, происходящие в обществе. К тому же, наличие газетных текстов как в основном корпусе НКРЯ, так и в собственно газетном корпусе, создает реальную базу для внутриязыкового сопоставления с целью выявления, прежде всего, отличий на номинативно-лексическом уровне, который, как известно, и является основным индикатором социокультурных и семантических сдвигов.

Таким образом, в основу иллюстративного лингвистического корпуса СМИ Гродненщины легли электронные версии районных газет («Берестовицкая газета», «Ивьевский край», «Островецкая правда», «Праца»⁴ («Труд»), «Свислочская газета») и ряда изданий, выходящих в областном центре, то есть в городе Гродно («Вечерний Гродно», «Перспектива»). Временной охват, в силу иллюстративного характера корпуса (обусловлен большой трудоемкостью его создания), ограничен 2012 годом. Текущая (не окончательная) версия корпуса по объему уже превышает 2 миллиона словоупотреблений. Поскольку корпус СМИ Гродненщины создается как часть модуля НКРЯ, то его метаразметка базируется на метаразметке НКРЯ и содержит следующие группы параметров: «Информация об авторе», «Информация о тексте», «Информация об аудитории», «Библиографическое описание», «Служебная информация». Специфика исходного языкового материала обусловила необходимость незначительного обновления базовой системы метаразметки, затронувшего, в основном, параметры группы «Информация о тексте». Так, к пометам (тегам) значений параметра «Тематика» был добавлен тег «происшествие»; для параметра «Тип текста» применение тега «поздравление» было распространено на сферу электронной коммуникации (до этого данный тег

⁴ Мы не переводим данное название, а лишь даем в скобках его семантизацию, поскольку в Республике Беларусь имеет хождение иное издание СМИ под названием «Труд».

использовался лишь применительно к сфере публицистики и к официально-деловой сфере); для этого же параметра была введена привязка тега «совет» к церковно-богословской сфере (до этого данный тег использовался исключительно по отношению к сфере публицистики). Были введены также новые теги, отражающие релевантные для языка СМИ Гродненщины значения: для сферы электронной коммуникации был введен тег «комментарии читателей»; для художественной сферы введены теги «стихотворение» и «басня» (в НКРЯ стихотворные и прозаические тексты представлены в различных модулях); для сферы публицистики введен тег «надпись» (применялся при разметке самостоятельных текстов, представляющих собой подпись под фотографией).

Традиционно важную роль в корпусной метаразметке играет тематическое аннотирование. Тематическое ядро иллюстративного лингвистического корпуса СМИ Гродненщины представлено следующими темами: политика и общественная жизнь; здоровье и медицина; сельское хозяйство; образование; искусство и культура; досуг, зрелища и развлечения; спорт; бизнес, коммерция, экономика, финансы; частная жизнь; администрация и управление; происшествия; криминал; религия; право; природа. Жанровое ядро корпуса представлено следующими типами текстов: заметка; информационное сообщение; статья; репортаж; интервью; хроника; объявление; поздравление; комментарий; комментарии читателей; отчет; анонс; совет; очерк. Язык СМИ Гродненщины подтверждает отмечаемое исследователями текстов масс-медиа принципиальное функционально-стилевое разнообразие, изучение которого все чаще становится предметом исследований в сфере «медиа лингвистики» (термин Т.Г. Добросклонской). Так, хотя преобладающее число текстов различных СМИ Гродненщины можно отнести к публицистическому стилю, среди текстов, представленных в них и, соответственно, в корпусе, встречаются репрезентанты таких сфер функционирования, как художественная (в т.ч. поэтическая), официально-деловая, научно-популярная, церковно-богословская, бытовая. Особо следует отметить сферы рекламы и электронной коммуникации.

Параметры метаразметки, описанные выше, не исчерпывают всего многообразия языкового материала, представленного в корпусе СМИ Гродненщины (как, впрочем, и всего разнообразия самой метаразметки, применяемой в данном электронном языковом ресурсе). Тем не менее, и их достаточно, чтобы можно было получить представление о тематическом, стилистическом и жанровом разнообразии данного лингвистического корпуса, обуславливающим

потенциальные направления его использования для самых различных исследовательских и дидактических целей. Основной спецификой корпуса СМИ Гродненщины, позволяющей говорить о нем как о принципиально новом типе лингвистического корпуса, является его «смешанный» характер, отражающий коммуникативно несбалансированный (в пользу русского языка) близкородственный билингвизм белорусского общества⁵. Так, корпус отражает особый характер сосуществования русскоязычного и белорусскоязычного дискурсов, которое проявляется на разных уровнях.

1. В изданиях Гродненщины, вошедших в корпус, сосуществуют как русскоязычные, так и белорусскоязычные тексты. Процентное соотношение словоупотреблений русскоязычной vs белорусскоязычной частей в разных изданиях таково: «Вечерний Гродно»: 99,40% и 0,60%; «Перспектива»: 98,03% и 1,97%; «Свислочская газета»: 87,22% и 12,78%; «Ивьевский край»: 86,86% и 13,14%; «Берестовицкая газета»: 66,87% и 33,13%; «Праца»: 48,27% и 51,73%; «Островецкая правда»: 15,25% и 84,75%. По преобладанию текстов на том или ином языке можно определить следующие группы газет: 1) газеты с незначительной долей белорусскоязычных текстов («Вечерний Гродно» и «Перспектива»), 2) газеты с малой долей белорусскоязычных текстов («Свислочская газета», «Ивьевский край»), 3) газета со средней долей белорусскоязычных текстов («Берестовицкая газета»), 4) газета с условно равными долями белорусскоязычных и русскоязычных текстов («Праца»), 5) газета с существенно преобладающей долей белорусскоязычных текстов («Островецкая правда»).

2. Языки заголовка и тела статьи могут быть различны. Показательны в этом случае газеты «Островецкая правда» и «Праца»: в on-line версиях изданий, за редкими исключениями, заголовки белорусскоязычных материалов поданы на русском. Как следует из записей в форумах на сайте газеты, редакция в этих случаях руководствовалась сугубо утилитарной целью привлечь на страницу с материалом и тех пользователей, которые ввели в поисковике запрос на русском языке. Такое решение любопытно наличием предположения, что выдача страниц на белорусском языке при запросе на русском будет оценена пользователем как приемлемая, понимаемая.

⁵ В Республике Беларусь 2 официальных («государственных») языка – белорусский и русский.

3. В текстах статей могут присутствовать **вкрапления** на языке, отличном от языка основного текста, и здесь типичны белорусскоязычные вкрапления в русскоязычных статьях и русскоязычные вкрапления в белорусскоязычных статьях (редкие случаи вкраплений на польском, английском и др. языках здесь не рассматриваются). В корпусе вкрапления типа «белорусское в русском» количественно преобладают над вкраплениями типа «русское в белорусском». В качестве причин сознательного использования вкраплений в текстах СМИ Гродненщины можно назвать следующие факторы:

а) официальность периодического издания, определяющая такой источник вкраплений, как названия официальных республиканских, областных, районных, городских акций, фестивалей, праздников и т.п. (такие названия могут быть даны на одном из двух государственных языков РБ (белорусском, русском) и будут встречаться в анонсах, информационных сообщениях, объявлениях (но не только в них));

б) цитирование действующих официальных документов, литературных источников, архивных источников (преимущественно – русскоязычных источников времен Российской империи, времен СССР); языковой материал свидетельствует о том, что, как правило, авторы в подобных случаях не придерживаются практики сопровождения цитаты переводом / семантизацией;

в) обращение к обычной для периодических изданий практике самоцитирования (реализуемой через ссылки на заголовки и фрагменты ранее опубликованных статей, названия рубрик) при «смешанном» характере издания, то есть при наличии в этом издании статей и на русском, и на белорусском языках;

г) выбор корреспондентом такого типа текста, который фиксирует диалог (ведущийся в условиях полиэтнического региона). В качестве примеров таких текстов можно назвать следующие: коллективное интервью (круглый стол, «горячая линия» и др.); анкета (то есть «заочное интервью», когда на некоторый набор вопросов отвечает несколько респондентов, язык ответов при этом может не совпадать с языком вопросов); текст структуры «текст корреспондента, вводящий и комментирующий микротексты читателей, созданные на одну тему (к примеру, на тему, заданную фотографией)».

В корпусе основная часть вкраплений представлена именами собственными (названиями); здесь также вкрапления типа «белорусское в русском» количественно существенно преобладают над вкраплениями типа «русское в белорусском» и распределяются по

следующим группам: **названия событий** (акций; конкурсов (в том числе инсценировок обряда); праздников; фестивалей; ярмарок; выставок; концертов; спектаклей); **названия организаций** (издательств; музеев и галерей; спорт-, тур- и агрокомплексов; лагерей; предприятий, фирм и магазинов); **названия объединений** (клубов; творческих объединений: ансамблей, хоров и вокальных групп; театров; спортивных объединений); **названия газет и журналов** (и их частей: названия статей, циклов, рубрик газеты), **названия наград** (названия знаков, орденов и медалей; номинации фестивалей). Более редко встречающиеся вкрапления относятся к следующим группам: **названия предметов культа** (название иконы); **названия книг, буклетов** (в том числе: названия серий книг); **названия песен, музыкальных и музыкально-хореографических композиций**; **названия продуктов**; **название серии товаров**; **псевдонимы**; **имена персонажей**; **топонимы**; **гидронимы**.

Вкрапления-названия, относящиеся к типу «русское в белорусском», малочисленны и сравнительно равномерно распределяются по следующим группам: ***названия школьных исследовательских работ**; **названия газет и журналов** (и их частей: названия статей, циклов, рубрик газеты); **названия книг, буклетов** (в том числе: названия серий книг); **названия наград** (названия знаков, орденов и медалей); **названия объединений** (клубов; *команд интеллектуальных игр); **названия организаций** (предприятий, фирм и магазинов); **названия песен, музыкальных и музыкально-хореографических композиций**; **названия продуктов**; ***названия разделов сайта**; **названия событий** (акций; конкурсов; праздников; спектаклей). В приведенном выше списке знаком звездочки отмечены группы (подгруппы), включающие в себя только примеры типа «русское в белорусском».

Коммуникативно несбалансированный характер билингвизма находит отражение в практике использования перевода на русский язык либо пояснений на русском языке в белорусскоязычных текстах: *Вядома, што новы бераставіцкі касцёл названы ў гонар Перамянення Пана (па-руску «Преображения Господнего»); Сярод падарункаў асабліваю цікаваць выклікае мангольскі нацыянальны музычны інструмент морын хуур (інструмент с лошадиной головой); Нездарма і назву калектыў мае польскую: «стакроткі» – «маргарыткі»* (все примеры – «Берестовицкая газета»). В последнем примере автор материала отклоняет оба возможных варианта толкования на белорусском: и *стакротка*, почти идентичное польскому *stokrotka* (а значит, непригодное для пояснения), и слишком «русское» *маргарытка*.

«Смешанный» характер корпуса проявляется и в достаточно распространенном способе включения вкраплений в текст, когда вкрапление реализует связь управления, заданную опорным словом-«невкраплением» либо согласуется (в косвенном падеже) с таким опорным словом⁶. Показательно, что редки примеры использования неверной морфологической формы (*на выставках-продажах в «Кулинарии», «Родном куте»* (вместо *куце*; «Берестовицкая газета»).

Еще одним подтверждением смешанного характера корпуса является существование параллелей **«белорусское – адаптированное»**. Будем считать, что параллель **«белорусское – адаптированное»** определена, если у выступающего как вкрапление и частотного в языковой практике белорусского слова существует «орфографически адаптированный» вариант (язык которого, строго говоря, не является ни русским общенациональным, ни белорусским языком). Приведем примеры: параллель *«Дажынкі – Дожинки»* (название республиканской ярмарки-фестиваля, данное по названию традиционного обряда; эквивалентный перевод на русский отсутствует); параллель *«выцінанка – вытинанка»* (‘техника вырезания орнаментального или сюжетного узора из бумаги; результат применения такой техники’). С определенной долей

⁶ Примеры для типа «белорусское в русском»: *В «Бераставіцкай газеце» раз в полугодие публикуюцца спіскі <...>; Мне нравится ваш язык за «мілагучнасць, за прыгажосць»* («Берестовицкая газета»); *На эти и другие вопросы корреспондента «Ўзеўскага краю» ответил военный комиссар <...>; И она <...> выгатавлявала разные «прысмакі» для усердствовавших на здешних полях и фермах, новостройках сельчан* («Ивьевский край»); *Побывали участники круглого стола и в «Гарадзенскім маёнтку «Каробчыцы»*; *В том обсуждении пришли к выводу, что у Беларусі нет брэнда, но есть репутация страны «памяркоўных, трудолюбивых, незлобивых, радушных, гостеприимных людей* («Перспектива»); *Столица нынешних областных «Дажынак» – Новогрудок – готавіцца к традыцыйнаму фестывалю-чэстыванню тружэннікоў сёла; Кроіла, шіла наряды мясным модніцам, вязала світэра, ткала «дываны», «пасякі», паловікі* («Свислочская газета»).

Примеры для типа «русское в белорусском»: *Беларускіх кніг шмат і іх купляюць, – расказала мне прадавец «Детского мира» Зоя Хілоты; Былі падведзены і вынікі асабістага перышынства ігракоў – інтэлектуалаў у 4-м раённым чэмпіянаце па «Своей ігре»; Да таго ж кожны твор прагучаў у своеасаблівым «прочтении» кіраўніка ансамбля і аркестра <...>* («Берестовицкая газета»).

условности к этому ряду примеров можно добавить параллели «*рушнік – рушник*», «*рушнік – ручник*» (‘домотканое полотенце с узором’), «*капліца – капліца*» (‘часовня’): второе слово каждой из этих параллелей хотя и фиксируется в словарях современного русского языка, но имеет помету *областное* (или сходную).

В корпусе адаптированный вариант может конкурировать с русским переводом: в качестве примера приведем параллель «*кірмаш – кирмаш*». В текстах корпуса зафиксирована неполная парадигма слова *кірмаш* (ср. белорусское *кірмаш*), где представлен не только по понятным причинам преобладающий именительный / винительный единственного числа, но и свидетельствующие об освоенности этого слова косвенные формы единственного и множественного чисел *кірмаша, кірмаше, кірмашей, кірмашах*. Контексты показывают, что все случаи орфографически адаптированной записи относятся к традиционной ярмарке «*Ганненскі кірмаш*». Русскоязычная запись, «*Анненская ярмарка*», в корпусе наименее частотна (12 случаев при 19 случаях для адаптированной записи и 43 случаях для белорусскоязычной записи). Еще одним доказательством вытеснения русскоязычного варианта адаптированной записью является факт наличия двух намеренно переведенных на русский заголовков белорусскоязычных статей (см. ранее о практике такого перевода у газет «*Островецкая правда*» и «*Праца*»), где вместо ожидаемого ярмарка видим *кірмаш*: ***Кірмаш без коней – не кірмаш***; ***В Зельве провели кірмаш*** (оба примера из газеты «*Праца*»).

Отметим, что случаи проявления пресловутой «трасянки», понимаемой нами как близкородственная межъязыковая интерференция, возникающая как результат индивидуального неосознанного двуязычия смешанного типа, достаточно редки: *Вучні разважалі над паняццямі чалавечнасці, міласэрнасці, дабрыні і гуманнасці* («*Берестовицкая газета*»); *К участію допускаюцца как коллективныя, так і індывідуальныя работы: четверостишся, стихи, аднострачныя (скетчы), слоганы і т.д.* («*Островецкая правда*»); *С пачатком снегападоў на прадарожных участках, дзе нет зашчытных лесопалос, рабочыя пачалі ўстаноўку спецыяльных **снегазатрымліваючых** цытоў* («*Праца*»). Данное обстоятельство, как и оооооо функционирования вкраплений, подтверждает нашу мысль, уже высказанную выше, о том, что авторы СМИ Гродненщины используют оба кода сознательно, то есть мы имеем реальную фактологическую базу для исследования сознательного переключения кодов в медийных письменных текстах.

Подводя итог вышесказанному, можно с уверенностью утверждать, что лингвистический корпус на материале СМИ 496

Гродненщины представляет собой многоцелевой электронный ресурс, позволяющий проводить исследования в самых разных лингвистических направлениях, включая взаимодействие языков. Впервые корпус рассматривается как эффективный инструмент для объективного (основанного на данных) выявления социокультурной специфики белорусского диалекта русского языка и отражения эволюционной ситуации в русскоязычном узусе полиэтничного нероссийского региона с продолжительной ситуацией несбалансированного многоязычия, а также для выявления отличий в аксиологических картинах мира представителей близких, но не тождественных культур, являющихся носителями одного мирового языка. Особую роль данный корпус может и должен сыграть в дидактике, представляя уникальный материал как с точки зрения положений теории языка и социолингвистики, так и отражая реальные узуальные практики, учет которых необходим при обучении русскому языку, особенно русскому языку как иностранному. Электронный характер создаваемого языкового ресурса и его доступность посредством Интернета делают его незаменимым инструментом обеспечения различных видов самостоятельной работы учащихся, в том числе в режиме дистанционного обучения.