

Издавач:
Филолошки факултет „Блаже Конески“ – Скопје

Уредник на издавачката дејност:
Анета Дучевска, декан на факултетот

Уредник на изданието:
Биљана Мирчевска-Бошева

Меѓународен уредувачки одбор:
Branko Tošović, Karl-Franzens-Universität Graz
Ivana Vidović Bolt, Filozofski fakultet Sveučilišta u Zagrebu
Piotr Źmigrodzki, Instytut Języka Polskiego PAN Kraków
Лидија Танушевска, Филолошки факултет „Блаже Конески“ – Скопје

Лектура:
Милица Миркуловска

Компјутерска подготовка и печат:
БороГрафика

Дизајн на корица:
Наталија Лукомска

Тираж:
300 примероци

CIP - Каталогизација во публикација
Национална и универзитетска библиотека „Св. Климент Охридски“, Скопје

81'373.7(062)

ЗБОРНИК на трудови од Меѓународната конференција „Зборот збор
отвора“. - Скопје : Филолошки факултет „Блаже Конески“, 2017. - 336
стр. : табели ; 25 см

Трудови на повеќе јазици. - Фусноти кон текстот. - Библиографија кон
трудите. - Summaries кон трудите

ISBN 978-608-234-047-0

а) Фразеологија - Собири
COBISS.MK-ID 102333962

Универзитет „Св. Кирил и Методиј“ во Скопје
Филолошки факултет „Блаже Конески“ – Скопје

ЗБОРОТ ЗБОР ОТВОРА

Уредник на изданието Биљана Мирчевска-Бошева

Бранко ТОШОВИЧ (эм.)

Университет им. Карла и Франца Грац

Австрия

ГРАЛИС-КОРПУС

Резюме: Гралис-Корпус является параллельным корпусом, предназначенным для исследования славянских языков, в первую очередь южнославянских. Он состоит из двух подкорпусов: GralisText-Korpus-a и GralisSpeech-Korpus-a. GralisText-Korpus содержит письменные тексты. Speech-Korpus образуют три части: Wort-Korpus (записи отдельно произнесенных слов), Fix-Korpus (записи прочитанных текстов) и Frei-Korpus (записи и транскрипты спонтанной речи).

Ключевые слова: корпус, Гралис-Корпус, параллельные корпуса, аннотация, славянские языки

1. В 2007 году создан комплексный и многоязычный текстуальный и устный корпус для изучения славянских языков – Гралис-Корпус. В настоящее время им пользуются слависты из различных стран в рамках коллективных и индивидуальных проектов и в процессе обучения.¹ Он функционирует как многоязычный (параллельный) корпус славянских языков, а также немецкого языка. Гралис-Корпус относится к аннотированным корпусам (морфосинтаксическая разметка пока разработана для сербского, хорватского и бошняцкого языков). Пользователям стоит в распоряжении оболочка (интерфейс) на английском, немецком, русском и сербском/хорватском/бошняцком.

2. Создание Гралис-корпуса было осуществлено в рамках (а) научно-исследовательского проекта, финансируемого Австрийским фондом для поддержки научных исследований, (б) сотрудничества между Институтом славистики и Центром по моделированию информации Гуманитарного факультета Университета им. Карла и Франца в Граце (в) при поддержке

¹ Руководителем проекта является Бранко Тошович (Branko Tošović), координатором работ Арно Вонисх (Arno Wonisch), использованием и развитием программной поддержки занимается Хубер Штиглер (Hubert Stigler), а администрированием корпуса Ольга Ленер (Olga Lehner).

специалистов в области славянской корпусной лингвистики.² В качестве платформы используется Gralis – славистический лингвистический портал Университета им. Карла и Франца в Граце³, основанный в 2010 г.

3. В работе над Гралис-Корпусом упор делается на создание 1) параллельного устного и письменного корпусов для всех славянских языков, 2) текстуального корпуса, охватывающего все функциональные стили и доступного для как можно большего числа пользователей, при полном соблюдении авторских прав, 2) реляционных баз данных – „Slaw-Lexikarium“ с корпусными словарными единицами, „Slaw-Grammatikarium“ с корпусными грамматическими формами, „Slaw-Akzentarium“ с ударениями для всех слов в корпусе, „Slaw-Präskriptarium“ с правилами для написания корпусных слов и „Slaw-Bibliothekarium“, с важнейшей сопровождающей литературой.

4. Создание Гралис-Корпуса преследует цель (а) собрать материал для славянских языков, необходимый для анализа отношений между ними, их совпадений, сходств и различий, а также межязыкового расстояния. Это подразумевает: 1. количественное и качественное сопоставление категорий параллельных фрагментов и единиц текстов, 2. анализ статистики пар относительно эквивалентных единиц параллельных текстов каждого из языков по ряду параметров (например, по статистике типов соответствия: слово языка¹ – слово языка², слово языка¹ – словосочетание языка², словосочетание языка¹ – слово языка², слово языка¹ – $\emptyset$ языка² и т. п.), 3. интерпретация, качественный анализ статистических данных, 4. выработка и согласование параметров, по которым обследуются тексты каждого из языков по отдельности для последующей корректной сопоставимости получаемых результатов (составление частотных словарей словоформ, лексем, морфологических категорий и т. п. для каждого из текстов, для текстов определенного жанра, функционального стиля и всего корпуса текстов данного языка в целом), 5. исследование внутренних характеристик текстов каждого из языков. Гралис-Корпус преследует цель показать, как славянские языки функционируют и взаимодействуют на конкретном материале, в природном окружении и на всех языковых уровнях.

5. В Гралис-Корпусе существует интерфейс для всех славянских языков: белорусского, болгарского, бошняцкого, лужицко-сербского, македонского, польского, русского, сербского, словенского, словацкого, украинского, хорватского, черногорского, чешского, а также немецкого. По мере необходимости этот список можно расширить так наз. славянскими микроязыками: градишчанско-хорватским, кашубским

² В этих целях был проведен в летнем семестре 2006 года семинар на тему „Славянская корпусная лингвистика“.

³ <http://www-gewi.uni-graz.at/gralis>.

и русинским. В его подготовке делается упор на то, чтобы он был как можно больше репрезентативным и сбалансированным. Гралис-Корпус готовится в рамках концепции языковых корреляций, изложенной в книге *Korrelacionniy sintaksis* [Корреляционный синтаксис] (Tošović 2001) и в ряде статей, а также на основе взглядов о различии в языке и различии вообще. В качестве теоретической основы для типологической обработки и представления различных видов текстов служит книга *Funkcionalni stilovi* [Функциональные стили] (Tošović 2002). Согласно концепции, изложенной в ней, Гралис-Корпус разделен на пять функциональных стилей (литературно-художественный, публицистический, научный, официально-деловой и разговорный). Большое внимание уделяется фундаментальным критериям любого корпуса: (а) глубине и широте аннотирования, (б) возможностям поиска и получения информации, (в) представительному (репрезентативному) характеру, пропорциональности и сбалансированности, а также (г) свободе доступа. При заполнении корпуса предусмотрено уравновешенное включение текстов трех типов: оригинальных, модифицированных (адаптированных) и переведенных. Гралис-Корпус приспособлен в значительной степени к гипертекстуальным и корпусным стандартам (TEI, XML и др.). Его развитие направлено на количественное расширение новым содержанием, качественное улучшение оболочки, более широкое и глубокое аннотирование, функциональное ускорение, улучшение поисковой системы и программное обновление.

Корпус предназначен в первую очередь для специалистов в области сопоставительного языкознания и для лингвистов более широкого профиля (особенно в области общего и системного языкознания, а также социолингвистики), для интересующихся интралингвистическими, интерлингвистическими и экстралингвистическими отношениями между славянскими языками. Он является полезным для всех тех, кто сталкивается с проблемой взаимодействия славянских языков. Его можно целесообразно использовать в обучении, особенно в вузах.

6. Гралис-Корпус состоит из нескольких подкорпусов, разделенных на макроязыковые группы (восточнославянские, западнославянские и южнославянские языки) и микроязыковые группы (корпусы отдельных языков, а также индивидуальные корпусы – корпусы писателей, напр., Иво Андрича). Существует также возможность выбора двух макрогрупп (скажем, восточнославянской и южнославянской). Отдельный блок предоставляет возможность сопоставления славянских языков и немецкого языка.

7. В рамках Гралис-Корпуса выделяются два комплекса – Корпус устной речи (Speech-Korpus) и Корпус письменных текстов (Text-Korpus).

8. Устный Гралис-Корпус (GralisSpeech-Korpus) охватывает транскрибированные записи и предоставляет возможность звуковой материал

анализировать фонетически (акустически, артикуляционно), просодически и фонологически на уровне звука/фонемы, слога, лексемы, синтагмы, словосочетания и предложения. Существуют три подкорпуса: Wort-Korpus, Fix-Korpus и Frei-Korpus. Wort-Korpus содержит записи отдельно произнесенных слов на разных славянских языках (в настоящее время больше всего на сербском, хорватском и бошняцком). Fix-Korpus охватывает записи прочитанных текстов. Frei-Korpus состоит из записей и транскриптов спонтанной речи. Фонетическая и просодическая транскрипция вносится после верификации экспертов при помощи специальной программы – Valorisarium-a. Speech-Korpus связан с программой Akzentarium, предоставляющей возможность получить комплексную информацию об ударении (в данный момент для более 120 000 слов сербского, хорватского и бошняцкого языков) и позволяющей ставить ударение и искать/находить его в внесенных словах. В Transkriptarium-e дается информация о том, как произносятся слова, а в Suprasemgentarium-eo интонационной структуре предложения. Frei-Korpus является коллекцией свободно/спонтанно произнесенных текстов, относящихся ко всем функциональным стилям. BSpeech-Korpus-e предоставляется несколько возможностей поиска – по отмеченным предложениям, по частям с транскрипцией, акцентуацией и супrasegmentацией и т. д.

Для обработки данных используются особые процедуры, начинающиеся с заполнением бланка с основными персональными данными: Опрошенный, Запись и Корпус. В первой графе вносится основная информация о говорящем (пол, национальность, религия, год и место рождения, место жительства, профессия, место обучения, родной язык, важнейшие места проживания, иностранный язык и т. п.), дата и местозаписи, во второй следуют данные о теме, месте, обстоятельствах записи, оборудовании, длине и формате записи, операторе, а в третью вносятся данные о том, для какого типа исследования используется записанный материал (для устного, спектрального или какого-нибудь другого анализа).⁴ В конце бланка находится графа для письменного разрешения внести запись в Гралис. В этих целях выбирается пароль, состоящий из названия места записи, трехзначного номера и букв, обозначающий язык (напр., **b, k s** указывают на боснийский/бошняцкий, хорватский или сербский) типа split_001k. После внесения пароля открывается полный бланк с данными об опрошенном и о записи. Добавить дополнительную информацию можно в графе „Внести новое место или пароль“.

⁴ В анализе устного материала используются два метода – акустический (при помощи программы Praat измеряется расстояние между славянскими языками по основным параметрам: длине произнесенных звуков, изменению в интонации и т. п.) и слуховой.

9. Текстуальный Гралис-Корпус (GralisText-Korpus) дает возможность спаривать тексты на всех славянских языках, причем в настоящее время больше всего он пополняется материалом из южнославянских языков крупнейшего славянского языка – русского. Благодаря созданной инфраструктуре можно проводить параллелизацию(а) всех славянских языков и немецкого языка, (б) славянских языков по отдельным группам (восточнославянской, западнославянской, южнославянской), (в) языков одной группы (например, южнославянской). В рамках Текст-Корпуса существуют индивидуальные подкорпусы. Пока разработаны четыре из ряда запланированных – корпус Иво Андрича (1892–1975, лауреата Нобелевской премии за литературу), Бранко Чопича (1915–1984, величайшего славянского повествователя, юмориста и сатирика), Зорана Живковича (род. 1948, в настоящее время самого переводимого писателя бывшей Югославии) и Блаже Конеского (1921–1993, виднейшего македонского писателя, поэта и филолога). Особый типом является образовательный корпус, в состав которого входят тексты, необходимые дипломантам, аспирантам, докторантам, диссертантам для подготовки их научных работ.

С текстуальным корпусом связан онлайн словарь „Lexikarium“ и программа „Akzentarium“, позволяющие получать комплексную просодическую, лексикосемантическую и грамматическую информации. Корпус предоставляет два типа поиска: простой и сложный (при помощи наз. CQP-синтаксиса). Результаты (попадания) отмечаются желтым или синим цветом. Над полученными предложениями стоит стрелочка, указывающая на подлинник (если речь идет об интернет-адресе, приводится линк). Из корпуса можно извлекать частотные списки.

Материал сначала вносится в „Roh-Korpus“ („Сырой корпус“). Тексты, находящиеся в интернете и не нуждающиеся в решении вопроса авторских прав, включаются в „Warte-Korpus“ („Корпус в ожидании“), в его часть „FS-Korpus“, расчлененную на пять функциональных стилей („FS-Korpus“). В рубрике „Мета-Корпус“ („Meta-Korpus“) рассматриваются теоретические и практические вопросы создания корпуса и проводится дискуссия между сотрудниками проекта.

Обработка материала проходит в двух этапах. На первом отдельно готовятся тексты для каждого славянского языка. Важнейшей частью такой работы является аннотация (паспортизация) – метаязыковая, экстралингвистическая и грамматическая. Метаязыковая аннотация состоит из указания об источнике (авторе, заглавии текста, месте и годе издания, количестве страниц, издательстве, переводе и др.). Грамматическая аннотация указывает на морфологическую структуру, словоизменение и тип сочетае-

мости с другими словами в рамках словосочетания и предложения. Она проводится при помощи программы „MorphoGenerator“. Благодаря ей можно получать парадигму любого слова, что особенно важно в процессе обучения. Межъязыковые различия в частоте использования словоформ между языками не определяются произвольно, в общих формулировках, а объективно при помощи корпусных данных. „Gralis-MorphoGenerator“ предоставляет статистическую информацию о каждой части речи в Гралис-Корпусе.

Морфологическая разметка проведена в настоящее время для сербского, хорватского, бошняцкого и черногорского языков и охватывает 100 461 слов. Их парадигмы порождаются на основе 822 правил. Под правилом подразумевается (а) процедура выявления морфологических изменений и варьирования форм, определения способа чередований и их описания в форме коротких элементарных инструкций, (б) матрица соотношений между кодовыми позициями и парадигматическим наполнением, (в) прием раскрытия системы формоизменений в целях порождения всех форм и целостных парадигм. Существуют синтагматические и парадигматические, отдельные и общие, структурные, категориальные и интеркатегориальные правила. Особую группу составляют эксклюзивные и инклюзивные правила, а также флексийные и кумулятивные. Процедура разметки выглядит следующим образом. а) Создается список всех слов в рамках одного литературного языка. б) Слова разделяются по частям речи. в) В рамках каждой части речи слова объединяются в отдельный список по грамматическому признаку, важному для порождения парадигмы (для существительных это категория рода, для прилагательных наличие/отсутствие степеней сравнения, для глаголов вид). г) Полученный список расчленяется на подписки, учитывая дополнительный грамматический признак (для существительных это одушевленность, для глаголов переходность). д) Проводится анализ списка под г и выявляются парадигматические маркеры – падежи, лица, являющиеся релевантными для порождения целостной парадигмы (напр., для существительных это им., род., зв. и тв. п. ед. ч. и им. и род.п. мн. ч.). е) Создается список слов с общими грамматическими маркерами (скажем, для существительных мужского рода на согласный, имеющих в им. п. ед. ч. беглую гласную а). ж) Выбирается типичное слово в этой группе и для него создается таблица со всеми 20-ю позициями и всеми словоформами. По этому слову и номеру типа называется каждое правило (напр. 174^{мудрац}). з) Материал, полученный в процедурах а – ж, превращается в реляционную базу данных MySQL, распределяя слова не по окончаниям, а по конечным буквам. и) Создаются две маски – одна для поиска информации по грамматической разметке и вторая для отображения

результатов поиска. Обе объединены в MorphoGenerator. к) Образуются информационные каналы в двух направлениях – от MorphoGenerator-а к Корпусу и от Корпуса к MorphoGenerator-у, для того, чтобы любую форму, приведенную в MorphoGenerator-е, можно было найти в Корпусе и получить информацию о ее частоте, и чтобы для любого словоупотребления в Корпусе можно было получить парадигму в MorphoGenerator-е.⁵

После проведения разметки текст расчленяется на предложения, в результате чего получится система, в которой каждому предложению языка А соответствует предложение языка В, С... Параллелизация состоит в том, что тексты двух или более славянских языков объединяются, потом проводится их выравнивание. Если, например, в одном языке абзац состоит из трех предложений, а в другом из пяти, приходится устранить такое неравновесие. Для автоматизации данного процесса используются имеющиеся (не)модифицированные разработки. Когда существующие программы не в состоянии выполнить поставленные задачи, создаются новые инструментальные средства, позволяющие в частности в одном комплексе объединять тексты различных славянских языков, автоматически находить несовпадения в числе предложений и автоматически делать исправления. Для серверной работы используются IMS CoprusWorkbench (CQP) и Asset-Management: из корпусного материала создаются списки языковых единиц, которые затем превращаются в реляционные базы данных при помощи программы MySQL, на основании которых можно готовить словари различного типа. IMS OpenCorpusWorkbench служит для администрирования, подготовки и осуществления поиска в больших текстовых корпусах с лингвистической аннотацией. Его главным компонентом является поисковая программа CQP (CorpusQueryProcessor). CWB содержит инструменты для кодировки, индексации, сжатия, декодирования и частотных распределений, общий реестр, в котором хранится информация о корпусе (название, атрибуты, место нахождения), поисковую программу (CQP), осуществляющую быстрый поиск с использованием синтаксиса регулярных выражений по значениям атрибутов отдельных позиций (например, по морфологическим тэгам). Простой и расширенный поиск базируются на CQP-синтаксисе (Suchemit CQP-Syntax) и предлагают очень широкие и разнообразные комбинации. Результаты (попадания) вертикально выводятся на монитор. Название текста/источника обозначается

⁵ Существуют различные методы и кодировки, используемые для грамматического аннотирования. Так как Гралис-Корпус требует полностью унифицированной аннотации (чтобы искать информацию для всех СлЯз), выбрана та, которая больше всего годилась для изучаемых языков. Самой целесообразной оказалась Multext-East кодировка (Multilingual Texts and Corpora for Eastern and Central European Languages – multilingual dataset for language engineering research and development: MultiText East-www), разработанная в 2004

желтым цветом. Нажатием на стрелочку с левой стороны заглавия получается основная метаинформация (об авторе, месте и времени издания, издательстве, числе страниц и т. п.). Пользователь может определить размер отображения. Допускаются различные виды сортировки строк, подсчитываются частоты (например, для комбинации слов), составляется многоязычный индекс для параллельных корпусов.

Корпусная работа состоит из шести процедур: 1. выработки инструментальных средств для создания корпуса, его ведения (пополнения, предобработки текстов, паспортизации, контроля их параметров и т.п.), структурирования (разметки структурными пометами), категоризации (качественной квалификации его фрагментов и единиц, выделенных в ходе структурирования текстов), выравнивания и фиксации связи между относительно эквивалентными элементами параллельных текстов корпуса, 2. создания, тестирования, наполнения, пробной эксплуатации и развития совершенствование оболочки (программного средства) для выполнения работ по созданию, ведению, структурированию, категоризации, выравниванию элементов параллельных текстов, 3. выработки и согласования схемы метаязыка, а также структурной и категориальной разметки фрагментов и единиц текстов, 4. переработки текстов каждого из языков, контроль их параметров и характеристик паспортизации, структурную разметку текстов, категориальную разметку фрагментов и единиц текстов (как минимум предусматривается осуществить лемматизацию и морфологическую квалификацию словоформ), выравнивание (Alignment) соответствующих эквивалентных фрагментов параллельных текстов, 5. выработки и согласования функционально-стилистической (жанровой) схемы сбора параллельных текстов для корпуса, 6. сбора и исходной паспортизации (метаразметки) параллельных текстов по каждому из языков.

10. В процессе создания корпуса кроме подготовки базовых текстов и аннотации метаданных в XML-формате перерабатываются цифровые ресурсы (MP3-, WAV-данные и т. п.) в рамках Speech-Korpusa. Asset Management System (AMS) дает такую возможность в форме Workflows (системы хранения, управления и подготовки цифровых ресурсов, накапливаемых в большом количестве). В отличие от Content Management System, в Asset Management выступает на передний план, прежде всего, архивирование постоянных метаданных, доступных для цитирования и гибкого использования цифровых ресурсов. Asset Management System предоставляет очень гибкий способ хранения документов, основанных на XML, и управление корпусом.

В качестве платформы для трансфера материала, сохранения цифровых web-ресурсов и управления ими используется Open Source Projekt

„Fedora“ (Flexible Extensible Digital Object Repository Architecture). Она содержит полный текстовый регистр (Volltext Index), базируемый на Apache Lucene, и версионное управление (Versionsverwaltung) Asset-содержанием (Assetinhalte). Для нее характерно определение тонко раздробленных прав на доступ к Assets и к его частичным энтитетам на основании XACML (Extensible Access Control Markup Language), а также стандартные импорт- и экспортформаты: METS (Metadata Encoding and Transmission Standard). Существует поддержка стандартных протоколов обмена метаданными типа OAI-PMH. На основе носящей технологии (Trägertechnologie) Apache Tomcat можно осуществить при помощи Repository Clustering и загрузочную балансировку (Load Balancing) системного окружения с соответствующим числом пользователей (Concurrent Usern).

Для успешного и эффективного проведения корпусных работ существует особый on-line Projekt-Management, в рамках которого созданы базы данных „Персоналиум“, „Библиотекариум“, а также „Слав-Форум“. „Персоналиум“ содержит все основные данные о сотрудниках в проекте. В эту базу данных вносятся тексты для публикаций, резюме, хендауты и презентации для совместных конференций. В „Библиотекариуме“ стоит в распоряжении литература, относящаяся к исследуемой теме. Для постоянного контакта и быстрого обсуждения определенных вопросов используется Slawistik-Forum (<http://www.gewi.uni-graz.at/gralis/slav/forum>).

11. Гралис-Корпус является важным для македонистики, так как в его рамках создан первый электронный корпус македонского языка – Гралис-Мак. Договор об австрийском-македонском научно-исследовательском проекте „Одноязычный и многоязычный электронный корпус македонского языка“ подписали 18 декабря 2008 г. Институт славистики Университета им. Карла и Франца в Граце и Филологический факультет им. Блаже Конеского Университета им. Кирилла и Мефодия в Скопье (Договор 2013). Основной целью проекта было исследование и популяризация македонского языка, а также обучение ему. Корпус македонского языка состоит из двух частей: 1) текстов для исследования македонского языка (Одноязычный корпус македонского языка), 2) тексты для сопоставительного исследования македонского языка по отношению к (а) другим славянским языкам, (б) к немецкому языку (Многоязычный корпус македонского языка).

На собранном корпусном материале был проведен в 2012 г. первый австрийско-македонский симпозиум на тему „Поэтика, стилистика и лингвистика текстов Блаже Конеского в Гралис-Корпусе“ (Symposium Koneski 2012) и опубликован сборник докладов (Тошович 2013^a). Второй том в рамках данного проекта охватывает переводы текстов Б. Конеского на разные языки; эти переводы предназначены для македонского Гралис-Корпуса (Тошович 2015). На основе Гралис Мак-Корпуса написана одна

докторска дисертација, една дипломна работа, ряд курсових работ и неколку одделних истражувањия. Мак-корпус подробно опишан в ряде статей (см. список литературы).

В настоящее время македонский корпус содержит 406 текстов на различных языках, но его основную часть составляют тексты Конеского (рассказы, стихотворения и очерки). Авторами остальных текстов являются Арсовски, Бошковски, Георгиевский, Малеский, Павловский, Пандев, Пендовский, Старова, Чашуле, Яневский и др.

Источники:

- Gralis-Korpus-www: *Gralis-Korpus*, http://www-gewi.uni-graz.at/gralis/korpusarium/gralis_korpus.html. 15. 10. 2016.
 Mak-Korpus: <http://www-gewi.uni-graz.at/gralis/projektarium/Mak-Korpus/index.html>. 15. 10. 2016.
 SymposiumKoneski. 2012: *Симпозиум: Поетикајџа, стилстикјџа и лингвистикајџа на текстовите од Блаже Конески во Корпусот „Галис“*, http://www-gewi.uni-graz.at/gralis/projektarium/Mak-Korpus/Symposium_mk1.html. 15. 10. 2016.

Литература

на кириллице:

- Гајдова, Убавка, Јачева-Улчар, Елка 2015. Суфиксот ***-ина** како зборообразувачки во текстовите на Блаже Конески во корпусот „Галис“, *Македонски јазик*. Скопје: Филолошки факултет, год. 66, с. 193–204.
 Договор за австриско-македонскиот научноистражувачки проект „Моно- и мултилингвален електронски корпус на македонскиот јазик“. 2013. Тошовиќ, Бранко (Hg./ur.). *Поетикајџа, стилстикјџа и лингвистикајџа на текстовите од Блаже Конески во корпусот Галис / Poetik, Stilistik und Linguistik der Texte von Blaže Koneski im Gralis-Korpus*. Грац – Скопје / Graz – Skopje: Institut für Slawistik der Karl-Franzens-Universität Graz – Филолошки факултет „Блаже Конески“ во Скопје – Институт за македонски јазик „Крсте Мисирков“ во Скопје, 231–233.
 Лаброска, Веселинка. 2013. За лексичките маркери за евиденцијалност во текстовите на Блаже Конески (според материјалите достапни во корпусот), Тошовиќ, Бранко (Hg./ur.). *Поетикајџа, стилстикјџа и лингвистикајџа на текстовите од Блаже Конески во корпусот Галис / Poetik, Stilistik und Linguistik der Texte von Blaže Koneski im Gralis-Korpus*. Грац – Скопје / Graz – Skopje: Institut für Slawistik der Karl-Franzens-Universität Graz – Филолошки факултет „Блаже Конески“ во Скопје – Институт за македонски јазик „Крсте Мисирков“ во Скопје, 95–103.

Попоски, Димитар. 2013. Корпусот Галис-мак и творештвото на Блаже Конески како модели за отворени образовни ресурси, Тошовиќ, Бранко (Hg./ur.). *Поетикајџа, стилстикјџа и лингвистикајџа на текстовите од Блаже Конески во корпусот Галис / Poetik, Stilistik und Linguistik der Texte von Blaže Koneski im Gralis-Korpus*. Грац – Скопје / Graz – Skopje: Institut für Slawistik der Karl-Franzens-Universität Graz – Филолошки факултет „Блаже Конески“ во Скопје – Институт за македонски јазик „Крсте Мисирков“ во Скопје, 155–163.

Попоски Димитар, Бојковска Емилија, Тошовиќ, Бранко, Вонич, Арно. 2012. Создавање на текстови во сите функционални стилови и нивните авторски права за македонскиот Галис-Корпус („Галис-Мак“), Трајкова, Катица (и др.ред.), *XXXVIII Научна конференција на Меѓународниот семинар за македонски јазик, литература и култура (Охрид, 14–15 јули 2011)*. *Лингвистика*, 391–396.

Тошовиќ, Бранко. 2008. Сопоставителноизучение славјанских јазиков при помошмногоязычног „Галис-Корпуса“, *Izučavanje slovenskih jezika, književnosti i kultura kao i noslovenskih i stranih*, Beograd: Slavističko društvo Srbije, 336–340.

Тошовиќ, Бранко. 2011. Корпусот на македонски јазик во рамнике на Корпусот Галис, *XXXVII научна конференција на меѓународен семинар за македонски јазик, литература и култура: Лингвистика*, Скопје: Универзитет „Св. Кирил и Методиј“, Меѓународен семинар на македонски јазик, литература и култура, 281–290.

Тошовиќ, Бранко (Hg./ur.). 2013^a. *Поетикајџа, стилстикјџа и лингвистикајџа на текстовите од Блаже Конески во корпусот Галис / Poetik, Stilistik und Linguistik der Texte von Blaže Koneski im Gralis-Korpus*. Грац – Скопје / Graz – Skopje: Institut für Slawistik der Karl-Franzens-Universität Graz – Филолошки факултет „Блаже Конески“ во Скопје – Институт за македонски јазик „Крсте Мисирков“ во Скопје.

Тошовиќ, Бранко. 2013^b. Лексичките доминанти во текстовите на Блаже Конески во корпусот „Галис“, Тошовиќ, Бранко (ур.) / Tošović, Branko (Hg.). *Поетикајџа, стилстикјџа и лингвистикајџа на текстовите од Блаже Конески во корпусот Галис / Poetik, Stilistik und Linguistik der Texte von Blaže Koneski im Gralis-Korpus*, Грац – Скопје, Graz – Skopje: Institut für Slawistik der Karl-Franzens-Universität Graz – Филолошки факултет „Блаже Конески“ во Скопје – Институт за македонски јазик „Крсте Мисирков“ во Скопје, 189–198.

Тошовиќ, Бранко. 2013^c. Галис-Корпус, *Wiener slawischer Almanach*, München: Verlag Otto Sagner, c/o Kubon&Sagner. [Sonderband], 83, 89–111.

Тошовиќ, Бранко / Tošović, Branko (Hg./ur.). 2015. *Blaže Koneski / Блаже Конески: Паралелен / Паралели*. Graz: Institut für Slawistik der Karl-Franzens-Universität.

Тошовиќ, Бранко. 2014. Грамматическая аннотация для Галис-Корпуса, *Казанская школа-семинар по компьютерной и когнитивной лингвистике: TEL'2014 „Языковая семантика: модели и технологии“*, Казань: Научно-исследовательский институт „Прикладная семиотика“ Академии наук Республики Татарстан, 142–147.

- Тошовић Бранко (Грац), Бојковска Емилија. 2013. Корпусот на Блаже Конески во Гралис-мак, Тошовић, Бранко (Hg./ur.). *Поетикајта, стилскајта и лингвистикајта на текстовите од Блаже Конески во корпусот Гралис / Poetik, Stilistik und Linguistik der Texte von Blaže Koneski im Gralis-Korpus*. Грац – Скопје / Graz – Skopje: Institut für Slawistik der Karl-Franzens-Universität Graz – Филолошки факултет „Блаже Конески“ во Скопје – Институт за македонски јазик „Крсте Мисирков“ во Скопје, 199–211.
- Тошовић, Бранко, Бојковска, Емилија. Попоски Димитар, Вониш Арно. 2012. Македонски корпус „Галис“ (Галис/Мак“), Трајкова, Катица (и др.ред.), XXXVIII Научна конференција на Меѓународниот семинар за македонски јазик, литература и култура (Охрид, 14–15 јули 2011). *Лингвистика*, 371–382.
- Тошовић, Бранко, Бојковска, Емилија. Попоски Димитар, Вониш Арно. 2012. Значењето и предизвиците на македонско јазичниот поткорпус „Галис-мак“ во рамките на корпусот „Галис“, Трајкова, Катица (и др.ред.), XXXVIII Научна конференција на Меѓународниот семинар за македонски јазик, литература и култура (Охрид, 14–15 јули 2011). *Лингвистика*, 383–390.
- на латинице:**
- Tošović, Branko. 2001. *Korelaciona sintaksa. Projektional*. Graz: Institut für Slawistik der Karl-Franzens-Universität.
- Tošović, Branko (Hg). 2008^a. Das Gralis-Korpus, *Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen*, Graz: Lit, 724–49.
- Tošović, Branko. 2008^b. Das Gralis-Akzentarium, *Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen*, Wien, Berlin: LIT, 770–776.
- Tošović, Branko. 2009. Leksička distanca između bosanskog/bošnjačkog, hrvatskog i srpskog jezika u Gralis-Korpusu, Branko Tošović (Hg.). *Die Unterschiede zwischen dem Bosnischen/Bosniakischen/Kroatischen und Serbischen: Lexik – Wortbildung – Phraseologie* (2), Wien, Berlin: LIT, 17–63.
- Tošović, Branko. 2012. Uputežakorišćenje Čopićevog Gralis-Korpusa, Tošović, Branko (Hg./ur.). *Poetika, stilistika i lingvistika pripovijedanja Branka Čopića / Poetik, Stilistik und Linguistik der Erzählens von Branko Čopić*. Graz – Banjaluka: Institut für Slawistik der Karl-Franzens-Universität Graz – Narodna i univerzitetska biblioteka Republike Srpske, 369–374.
- Tošović, Branko. 2016. *Korelaciona morfologija srpskog, hrvatskog i bošnjačkog jezika: Imenice. Konstrukcijsko izdanje*. Graz: Institut für Slawistik der Karl-Franzens-Universität.
- Tošović, Branko; Wonisch, Arno. 2009. Gralis-Korpus, Granić, Jagoda (ur.). *Jezična politika i jezična stvarnost*, Zagreb: Hrvatsko društvo za primijenjenu lingvistiku, 117–125.

Branko TOŠOVIĆ

GRALIS-KORPUS

Summary

Gralis-Korpus is a parallel corpus intended to explore Slavic languages, especially South Slavic languages. It consists of two subcorpora: Gralis Text-Korpus and Gralis Speech-Korpus. The first one contains written texts, the second one recordings of speech. Speech-Korpus consists of three parts: Wort-Korpus, Fix-Korpus and Frei-Korpus. Wort-Korpus is composed of spoken words, Fix-Korpus of images of a certain text, and Frei-Korpus of recordings of spontaneous speech.