

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ СЛАВЯНОВЕДЕНИЯ

БАЛТО-СЛАВЯНСКИЕ ИССЛЕДОВАНИЯ

XX

СБОРНИК НАУЧНЫХ ТРУДОВ



Москва

2019

УДК 811.16/811.17+39
ББК 81
Б 207

Редакционная коллегия:

А. В. Андронов, П. М. Аркадьев, Б. Вимер, П. У. Дини, В. А. Дыбо (отв. редактор серии), М. В. Завьялова (секретарь), К. А. Кожанов, Б. Лаумане, Е. Л. Назарова, М. В. Ослон, С. И. Рыжакова, Б. Стундзя, Т. М. Судник, А. Хольфут, Т. В. Цивьян, У. Шмальшитис

Оригинал-макет:

С. Г. Болотов

Б 207 Балто-славянские исследования—XX: Сб. науч. трудов. — М., 2019. — 480 с. DOI: 10.31168/2658-5766.2019.20

ISSN-13: 977-2658-576-00-8

ISSN-8: 2658-576-6

DOI: 10.31168/2658-5766

В сборнике представлены статьи ведущих специалистов в области балто-славянских языковых и этноязыковых контактов из России, Белоруссии, Литвы, Чехии, Германии. Статьи посвящены актуальным проблемам акцентологии и морфологии балтийских языков, балтийской и славянской диалектологии, этимологии, восточнославянской и балтийской мифологии. Темы, затронутые в сборнике, актуальны не только для славистов, балтистов, индоевропеистов, но и для лингвистов, фольклористов, этнологов, историков широкого профиля.

На фронтисписте — Фото Вяч. Вс. Иванова (Ю. Будрайтис, 16 июня 2009 г.).

ISSN 2658-5766

00020>

УДК 811.16/811.17+39

ББК 81



9 772658 576008

© Авторы статей, 2019
© Институт славяноведения РАН, 2019
© С. Г. Болотов, оригинал-макет, 2019

СОДЕРЖАНИЕ

<i>В. А. Дыбо, С. А. Крылов.</i> Вячеслав Всеволодович Иванов (1929–2017)	8
<i>П. М. Аркадьев, И. Б. Иткин.</i> Аблаут в литовских отглагольных именах: к описанию синхронного распределения. II	28
<i>С. Л. Николаев.</i> Реконструкция акцентных классов балто-славянских тематических глаголов с корнями на шумный. II. Акцентуация отглагольных производных с суффиксами *-tlo- (*-dhlo-), *-to-, *-sto- и *-tā- ...	41
<i>С. Г. Болотов, М. В. Ослон.</i> «Правило Лескина–Отрембского–Смочиньского» и мнимые исключения из закона де Соссюра	55
<i>М. Н. Саенко.</i> Периодизация истории праславянского субстантивного склонения	92
<i>Б. Вимер, К. А. Кожанов, А. Эркер.</i> Корпус славянских и балтийских говоров <i>TriMSo</i> : структура, цели и примеры применения	122
<i>Е. А. Сведенцова.</i> Сериализованные конструкции в старославянском языке ...	144
<i>В. Блажек.</i> Славянское * <i>tešьka</i> ‘медведица’	200
<i>Р. А. Агеева.</i> Синие камни в традиционной культуре восточных славян	207
<i>Д. Разаускас.</i> Символика отношения «рыба–птица» и его метафизические контексты	237
<i>Н. Лауринкене.</i> Мифический образ неба в балтийской традиции	292
<i>Т. В. Володина.</i> Люди и <i>хозяйева</i> : стратегии и регулирование межмировой коммуникации в представлениях белорусов Поднепровья	317
<i>В. А. Лобач.</i> Богатыри, осилки и девы-воительницы в топонимических преданиях восточных славян	350
<i>Р. Бальсис.</i> Язычество балтов: человеческие жертвоприношения	384
<i>В. Я. Петрухин.</i> Посох Криве-Кривайтиса в балто-славянской перспективе ...	395
<i>Р. Шмигельските-Стукене.</i> Из истории парламентаризма в Речи Посполитой: сеймики Новгородского воеводства, 1794 г.	409
<i>Е. Л. Назарова.</i> «В душе моей Вы живете рядом с моей матерью...» (о латышском поэте Карлисе Скалбе)	425
<i>Т. Бартеле.</i> Первое празднование Дня русской культуры в Риге в 1925 году (по материалам русской прессы)	437

In memoriam

Симас Каралюнас (20 января 1936 – 28 ноября 2015) (<i>Ю. Карацеюс</i>)	448
Альгирдас Сабаляускас (26 июля 1929 – 17 апреля 2016) (<i>С. Валянтас</i>)	459
«Наука всегда должна служить жизни, своему народу» (Зигмас Зинкявичюс) (<i>А. Юджентис</i>).....	467
Витаутас Амбразас (25 марта 1930 – 23 февраля 2018) (<i>П. М. Аркадьев</i>).....	475
Сведения об авторах	478

CONTENTS

<i>V. A. Dybo, S. A. Krylov.</i> Vyacheslav Vsevolodovich Ivanov (1929–2017)	8
<i>P. M. Arkadiev, I. B. Itkin.</i> Ablaut in Lithuanian deverbal nouns towards the description of the synchronic distribution. II	28
<i>S. L. Nikolayev.</i> Reconstruction of accent classes of Balto-Slavic thematic verbs with obstruent-final roots. II. Accentuation of deverbals with the suffixes *-tlo- (*-dhlo-), *-to-, *-sto- и *-tā-	41
<i>S. G. Bolotov, M. V. Oslon.</i> “Leskien–Otrębski–Smoczyński’s Rule” and alleged exceptions to de Saussure’s Law	55
<i>M. N. Saenko.</i> Periodization of the history of the Proto-Slavic noun declension	92
<i>B. Wiemer, K. A. Kozhanov, A. Erker.</i> The <i>TriMCo</i> corpus of Slavic and Baltic dialects: structure, goals and case studies	122
<i>E. A. Svedentsova.</i> Serialized constructions in Old Church Slavonic	144
<i>V. Blažek.</i> Slavic * <i>mečьka</i> ‘(she-)bear’	200
<i>R. A. Ageyeva.</i> Blue stones in the traditional culture of the Eastern Slavs.....	207
<i>D. Razauskas.</i> Symbolism of the “fish–bird” relationship and its metaphysical contexts.....	237
<i>N. Laurinkienė.</i> The mythical image of the sky in Baltic tradition.....	292
<i>T. V. Volodina.</i> People and <i>hosts</i> : strategies and regulation of inter-world communication in the culture of Belarusians of the Dnieper	317
<i>V. A. Lobač.</i> Bogatyrs, giants and virgin warriors in the toponymic traditions of the Eastern Slavs	350
<i>R. Balsys.</i> Paganism of the Balts: human sacrifice.....	384
<i>V. Ja. Petrukhin.</i> Krivè-Krivaitis’ staff in the Balto-Slavic perspective	395
<i>R. Šmigelskytė-Stukienė.</i> A case from the history of parliamentarism in the Polish-Lithuanian Commonwealth: “Sejmiks” (dietines) of the Nowogródek Voivodeship, 1794.....	409
<i>E. L. Nazarova.</i> “In my soul you live next to my mother...” (about the Latvian poet Kārlis Skalbe)	425
<i>T. Bartele.</i> The first celebration of the Russian Culture Day in Riga in 1925 (based on the materials of the Russian press)	437

In memoriam

Simas Karaliūnas (20 January 1936 – 28 November 2015) <i>(J. Karaciejus)</i>	448
Algirdas Sabaliauskas (26 July 1929 – 17 April 2016) <i>(S. Valentas)</i>	459
“Scholarship must always serve life and its own nation” (Zigmas Zinkevičius) <i>(A. Judžentis)</i>	467
Vytautas Ambrazas (25 March 1930 – 23 February 2018) <i>(P. M. Arkadievs)</i>	475
Information about the authors	478

В 2019 году исполняется 40 лет со «дня рождения» серии «Балто-славянские исследования»: в 1979 году Владимир Николаевич Топоров написал докладную записку от Сектора структурной типологии в дирекцию Института славяноведения и балканистики АН с предложением учредить ежегодник «Balto-balkanica» (текст этой записки был опубликован в «Балто-славянских исследованиях – XVIII», М., 2009). Официально решение об учреждении серии было принято годом позже, но фактически она уже существовала (в 1972 г. был издан «Балто-славянский сборник», а в 1974 г. — «Балто-славянские исследования»). Отсчет серии ведется от 1972 года, и в этом году мы выпускаем уже 20-й том.

Все эти годы бессменным главным редактором серии был Вячеслав Всеволодович Иванов. Но этот том он не увидит... Прискорбно, что Вячеслав Всеволодович, до самых последних своих дней принимавший живейшее участие в составлении и редактировании тома, не дождался его выхода в свет. Мы не посвящаем этот выпуск памяти Вячеслава Всеволодовича, потому что ощущаем в нем его живую руку, слышим его голос. Мемориальным должен стать следующий том.

Мы уверенно надеемся, что новый главный редактор — Владимир Антонович Дыбо — сохранит традицию и приумножит славные дни «Балто-славянских исследований».

От редколлегии

Б. ВИМЕР, К. А. КОЖАНОВ, А. ЭРКЕР

Корпус славянских и балтийских говоров *TriMCo*: структура, цели и примеры применения

1. Введение

Цель проекта *TriMCo*¹ состоит в комплексном изучении проявлений лингвистической конвергенции и дивергенции в зоне контактов между восточнославянскими и балтийскими говорами на более широком лингвогеографическом фоне (см. § 2). В первую очередь такое изучение нацелено на охват и описание самих явлений конвергенции и дивергенции. В то же время в задачи проекта входит выяснение причин возникновения конвергентных черт и описание отдельных процессов, которые могут оказаться полезными для понимания того, как более глобальные изменения и взаимная адаптация между языками одного региона происходят в ограниченном масштабе, т. е. как бы под микроскопом.

Для этой последней цели в рамках проекта *TriMCo* был создан сопоставительный корпус устной речи белорусских и литовских говоров на пограничье Беларуси и Литвы (дополнительно белорусские разговоры записывались и в Латгалии), а также русских говоров в Псковской области и латгальских говоров в Латвии (см. § 3).

В настоящей статье мы обсудим исходные положения, которые легли в основу проекта (см. § 2), представим наш корпус (см. § 3) и дадим обоснование квантитативным корпусным исследованиям в области диалектологии (§ 4). На фоне такого методологического обсуждения мы дальше эскизно представим результаты первых исследований, проведенных с помощью этого корпуса (см. § 4.1.3). Эти примеры проиллюстрируют, каким образом работа с корпусом связанной устной речи может помочь нам приблизиться к более адекватному отражению процессов, которые происходят в обыденной речи. Вместе с тем работа с корпусом предоставляет неоценимую помощь в оценке вариативности, которая существенным образом влияет на возможности проявления конвергентных черт при языковом контакте. Такая работа незаменима другими, «докорпусными» методами, к которым, например, относятся опросники для атласов или диалектологические словари. Дело в том, что такие методы не позволяют (или позволяют не в такой же мере) охватить всю вариативность, распределение вариантов и их ко-

¹ Акроним выведен из названия проекта *Triangulation Approach for Modeling Convergence with a High Zoom-In Factor*. Проект получал финансовую поддержку с сентября 2013 до мая 2017 года (DFG, WI 1286/16-1). За эту поддержку мы выражаем свою глубокую благодарность немецкому Фонду фундаментальных исследований (Deutsche Forschungsgemeinschaft, DFG).

личественную оценку. Ниже мы покажем, что сопряжение поисковых возможностей электронных корпусов и статистических методов дает возможность увидеть тенденции в кажущемся хаосе. Выводы и первые обобщения формулируются в заключающем разделе (см. § 5).

2. Исходные положения

Проект *TriMCo* направлен на то, чтобы раскрыть механизмы, ведущие к возникновению и укреплению лингвистических ареалов, а также на то, чтобы охватить внутреннюю дифференциацию таких ареалов.

В рамках проекта принят многофакторный подход, основанный на «триангуляции» методов и результатов исследований из области типологии, ареальной и контактной лингвистики, сравнительно-исторического языкознания и диалектной географии (см. <https://www.trimco.uni-mainz.de/>; также Wiemer, Seržant, Erker 2014). Акцент ставится на ареалы небольшого объема (микроареалы) и на углубленный анализ вариативности фонологических и морфосинтаксических явлений. Тем не менее, вариативность на уровне микроареалов следует рассматривать на фоне более крупных ареалов (англ. *areal clusters*) и более объемных континуумов (англ. *areal clines*). Таким образом, можно проверить, не являются ли конвергентные черты какого-нибудь выбранного микроареала, которые причисляются к его особенностям (или выдающимся чертам), всего лишь частью большего ареального комплекса, а также — в какой мере эти особые черты совмещаются между собой (наподобие пучков изоглосс) за узкими пределами микроареала.

Заодно нужно задаться вопросом, отличаются ли постулируемые особенности данного микроареала от более широкого окружения качественно (т.е. как изолированные факты) или в отношении частотности и в плане распределения. При оценке распределения грамматических черт необходимо учитывать лексическое наполнение, ограничения другими грамматическими свойствами и т. п.

В дальнейшем мы употребляем выражение «говор» в широком смысле, т. е. применительно ко всем разновидностям языка, отличным от литературного и распространенным среди деревенского населения (о такой широкой трактовке см. § 3). Несколько упрощая, можно сказать, что контактная зона, в которой пересекаются (или граничат) друг с другом славянские и балтийские говоры, понимается как ядро, вокруг которого концентрическими кругами расположены ареалы всё более возрастающего объема. В конечном итоге, такое расположение напоминает устройство матрешек. Поэтому, пользуясь этой метафорой, можно сказать, что излагаемое здесь соображение опирается на «матрешечный» принцип (подробнее см. в Wiemer 2013; 2019).

Следуя этому принципу, а также вышеизложенным рассуждениям, касающимся распределения выбранных черт какого-нибудь небольшого ареала (входящего в более крупные ареалы), можно обратиться к давним постулатам системных подходов («каждый говор обладает своей системой»), поскольку структурные черты оцениваются не в изоляции от других (в том же говоре или в том же ареале). Однако вместе с тем не игнорируется и то обстоятельство, что ни одна разновидность (или ареал) не существует в изоляции от соседних или более объемных единиц как в географическом («горизонтальном», диатопическом) отношении, так и в отношении сетей общения (англ. *networks*) и взаимодействия между разновидностями одного и того же языка («вертикальное», диастратическое отношение, например, между говором и литературным языком).

Напоследок, сопряжение принципа матрешек с углубленным анализом распределения конкретных черт может выявить особенности выбранных микроареалов — либо в целом, либо отдельных разновидностей (говоров и т. п.) — которые остались незамеченными или недостаточно оцененными именно потому, что не было возможности «проецировать» конкретные черты на окружение (в горизонтальном или вертикальном плане) и не проводилось количественного и дистрибутивного анализа на фоне других разновидностей того же или прилегающих регионов. Тем самым анализ корпусных данных делает возможным эвристическое «добывание» (англ. *data mining*) отдельных черт, а комплексное исследование совокупности таких черт (англ. *aggregate analysis*) позволяет составить целые профили той или иной языковой разновидности (говора, микроареала) с большей точностью и адекватностью реальному употреблению.

Понятно, что корпуса связной и естественной речи нужно признать главным и незаменимым орудием для того, чтобы следовать названным только что принципам описания языковой действительности. Поэтому в проекте *TriMCo* мы постарались создать корпус, отражающий устную речь старшего поколения деревенского населения, которое проживало на балто-славянском пограничье в течение последних 30 лет. Этот корпус мы и представим сейчас в общих чертах.

3. Корпус *TriMCo*: структура, состав, состояние

Корпус состоит из образцов устной, в основном монологической речи представителей восточнославянских и балтийских говоров. Название «говор» здесь условное, поскольку записывались не обязательно люди, которых можно было бы считать «типичными» представителями говоров «традиционного» типа, но вообще люди, которые в выбранных

населенных пунктах и их непосредственном окружении проживали долго (если не всю жизнь) и речь которых отражала местный способ общения. Подбирались преимущественно люди старшего возраста, но безоговорочно считать их представителями традиционных говоров нельзя уже по той причине, что традиционных говоров, которые легли в основу описаний (типа Аванесов, Орлова 1965), в лучшем случае учитывавших состояние до начала 1970-х гг., в принципе почти не осталось.

«Выросла» речь, которую приходится считать намного более смешанной в смысле неоднородности «слияния» разных говоров в региолекты (об этом понятии см. Трубинский 1984: 50–69; Герд 2000: 47–49), а также в смысле нивелировки за счет влияния литературного русского (или соответственно литовского или латышского) языка или приближенных к нему вариантов. В речи наших информантов степень смешения (или выравнивания) с вариантами, близкими литературному языку, разная, поэтому в дальнейшем «диалектную речь» нужно понимать, имея в виду только что сделанные оговорки, хотя мы и старались отобразить как можно больше такой речи, в которой влияние литературных вариантов отражалось в меньшей степени.

Тексты, включенные в корпус *TriMCo*, представляют собой часть полевых записей, которые вошли в архив проекта. Эти записи довольно разного технического качества, но все они хранятся в цифровом виде (включая оцифрованные более старые записи), их общий объём составляет 1118 часов. Транскрибировалась речь из записей с 1988, 1994–1996, 1998, 2000–2011, 2013–2016 гг.², так что ее временной диапазон достаточно большой. В данный момент корпус *TriMCo* складывается из транскрипций, отражающих в общей сложности 103 с лишним часа записей на говорах белорусского, русского, литовского и латгальского языков. Транскрипции делались в системе *ELAN* (см. список сокращений), обеспечивающей возможность синхронизации оригинальной звукозаписи с ее транскрипцией, а также регистрацию метаданных по информантам. Каждый файл оснащен метаданными, которые дают сведения о следующих показателях:

Длина записи: В идеальном случае мы «резали» отрезки в 15–30 минут от каждого из выбранных информантов³; если времени

² Довольно большая часть записей была нам предоставлена коллегами из Москвы, Варшавы и Вильнюса (М. Толстой, М. Янковяком и В. Кардялисом соответственно), за что мы им приносим свою благодарность.

³ Для этой технической процедуры использовалась программа *MAGIX Audio Cleaning Lab MX*. Поток звучащей речи разделялся в программе *ELAN* на отдельные отрезки (фразы) на основании естественных пауз в речи говорящего, но без учета информационного или синтаксического членения речи.

пригодной записи было меньше, соответственно столько, сколько можно было использовать. Запись, вне зависимости от ее продолжительности, всегда бралась одним цельным куском.

Код информанта: инициалы или, если имя и фамилия были неизвестны, шифр, который позволяет идентифицировать совсем анонимного информанта.

Пол информанта.

Год рождения: преимущественно мы подбирали самых старших представителей данных населенных пунктов. Большинству информантов в момент записи было более 65 лет.

Место рождения.

Год записи.

Место записи.

Информация о том, является ли речь на записи преимущественно монологической или диалогической (в корпусе решительно перевешивает монологическая). В разговорах с информантами исследователи не руководствовались какими-либо заранее принятыми рамками.

Школьное образование: тип школы, сколько классов закончено.

Вероисповедание.

Происхождение родителей (в смысле географии).

(Бывшее) профессиональное занятие информанта.

(Бывшее) место работы информанта.

Информация о том, менялось ли место проживания информанта.

Названные выше языки в корпусе представлены неравно как в количественном отношении (см. Таблицу 1), так и в отношении дальнейшей обработки транскрипций. Наибольшая доля приходится на белорусскую речь. Она записывалась по обеим сторонам сегодняшней белорусско-литовской границы, причем бóльшая доля на белорусской стороне. Небольшая часть была записана в Латгалии. При выборе записей соблюдался принцип как можно большей географической противопоставленности при одновременном стремлении к достаточно большому и представительному количеству текстов.

Литовская речь, включенная в корпус, происходит из двух регионов, которые ни географически, ни исторически друг с другом прямым образом не связаны: из Дзукии и из северовосточной Аукштайтии. Согласно традиционному диалектному членению эти регионы относятся к южноаукштайтским и вильнюсским восточноаукштайтским говорам соответственно (лит. *pietų aukštaičių / rytų aukštaičių vilniškių patarmės*). Количество записей из последнего региона несколько перевешивает. К южноаукштайтскому диалектному ареалу относятся также записи, включен-

ные в корпус из деревень Пелеса и Ромашканцы, которые находятся на белорусской стороне в непосредственной близости к границе с Литвой.

Латгальские записи собирались в разных районах, за исключением западных.

Записи из Псковской области включают следующие районы: Гдовский район на севере, Невельский и Себежский районы на юге, Островский, Печорский, Пушкиногорский, Пыталовский — в центре Псковской области. С точки зрения традиционной русской диалектологии (см., например, Диалектологическую карту русского языка 1965 года в ДАРЯ) Гдовский район относится к гдовской группе западных среднерусских окающих говоров, в то время как Пыталовский, Пушкиногорский, Островский, Печорский районы представляют псковскую группу западных среднерусских акающих говоров. Расположенные на юге Псковской области Невельский и Себежский районы по своим диалектным чертам традиционно относят не к среднерусским, а к западной группе южнорусского наречия.

В качественном аспекте корпус неравномерно аннотирован. Литовский корпус полностью аннотирован морфологически⁴. Части, отражающие псковские (русские) и белорусские записи, аннотированы по другим принципам, разрабатываемым в рамках исследовательской сети *SpoSla* (см. ниже). Неаннотированными до сих пор остались записи латгальской речи. Эта часть корпуса также до сих пор наименее выверена.

Таблица 1.

Корпус *TriMCo* в формате *ELAN* (состояние: май 2017 г.)

разновидность; регион	часы:минуты	количество информантов	количество словоформ ⁵
<i>белорусский</i> ; Беларусь, пограничье с Литвой (от Лиды до Браслава)	27:00	83	179.041
<i>белорусский</i> ; южная и восточная Литва	17:12	71	122.273
<i>белорусский</i> ; Латгалия	7:00	9	40.476
<i>русский</i> ; псковский район	14:00	30	88.458
Σ восточнослав.	65:12	193	430.248

⁴ Аннотация осуществлялась полуавтоматически с помощью программы *Fieldworks Language Explorer* (FLEx; <http://fieldworks.sil.org/flex/>).

⁵ Эти цифры включают и сравнительно немногочисленные реплики интервьюеров (исследователей). Однако при анализе количественных данных токены, встречающиеся в таких репликах, исключались (см. §4).

разновидность; регион	часы:минуты	количество информантов	количество словоформ ⁵
литовский; игналинский район	10:19	41	65.593
литовский; Дзукия	06:06	24	42.319
литовский; Пелеса, Ромашканцы (Беларусь)	05:00	17	34.989
латгальский	13:00	22	94.107
Σ балт.	34:25	104	237.008
общая сумма	99:37	297	667.256

Дальнейшая разработка корпуса *TriMCo* еще с 2016 года ведется в сотрудничестве с исследовательской сетью *SpoSla* (см. список сокращений). Цель этой сети двояка: во-первых, она объединяет участников разных проектов в России, Германии и Польше, изучающих славянскую диалектную речь и ее взаимодействие с кодифицированными вариантами соответствующих языков, причем изучение ведется исключительно на основе звукозаписей и их транскрипций (в формате *ELAN*), а также метаданных об информантах и обстоятельствах записи. За этим направлением исследований стоит не только стремление к надежной долгосрочной документации быстро меняющихся и исчезающих нелитературных разновидностей этих языков, но и убеждение, что вариативность, наблюдаемую в этих разновидностях, необходимо изучать с точки зрения вариационной лингвистики (англ. *variationist linguistics*) и методами, основанными на употреблении (англ. *usage-based*), поскольку только такой подход обеспечивает приближение к реальной картине языковой действительности (см. выше). Во-вторых, в рамках *SpoSla* развиваются технические условия для создания инфраструктуры, предоставляющей не только возможность прочной архивации записей диалектной речи, но и программных средств для аннотирования и поисковых процедур, которые, с одной стороны, достаточно унифицированы, чтобы сделать возможным обмен данными из разных частных (как правило, низкобюджетных) проектов, но которые, с другой стороны, также достаточно гибки для того, чтобы приспособливаться к индивидуальным нуждам каждого отдельного проекта и его корпуса данных. В дальнейшей перспективе для предоставления таких средств и хранения корпусов должна быть создана платформа, доступная в интернете (англ. *repository*).

Сеть *SpoSla* устроена как неофициальное объединение ученых на основе взаимной поддержки, стремящееся к достижению выше-

описанных целей. Ее инициатором является Р. фон Вальденфельс (Йенский университет им. Фридриха Шиллера), а с российской стороны в нем участвует проект *Корпус бассейна реки Устья* (Ustyia River Basin Corpus). В данный момент в корпусе *TriMCo* разрабатывается аннотация для восточнославянских подкорпусов.

Прежде чем показывать на примерах, чем полезен и даже незаменим такого рода корпус, перейдем сначала к обсуждению коренного вопроса.

4. Зачем изучать диалектную речь с помощью электронных корпусов?

Говоры (и другие нестандартные разновидности того или иного языка) создают идеальную лабораторию, позволяющую раскрывать элементарные шаги, которые в своей совокупности могут привести к устойчивым изменениям (см. De Vogelaer, Seiler 2012). Тем не менее исследователи восточнославянских диалектов (а также таких нестандартных разновидностей, как городское просторечие) до сих пор почти не пытались воспользоваться вариативностью, обнаруживающейся внутри отдельных разновидностей, для того, чтобы смоделировать актуально наблюдаемые процессы, способные привести к изменениям (см. обзор в Wiemer, Seržant 2014).

Данные, собранные с помощью опросников и представленные на страницах атласов, способны отражать всего лишь более или менее грубые изоглоссы (или пучки изоглоссы), но они не оказываются особенно полезными, когда нужно установить параметры вариативности. В атласах вариативность обычно даже не видна, поскольку карты атласа, как правило, лишь статично и в двух измерениях отражают отсутствие или наличие того или иного признака. Такими данными трудно или вообще невозможно оперировать, когда нужно понять, с какой последовательностью соответствующий признак реализуется, насколько распространился он в словарном запасе исследуемого говора или каким дополнительным ограничениям он подчиняется. Кроме того, признаки часто необходимо разбить на более мелкие параметры, которые нагляднее показывают вариативность, а многие «особенности» отдельных говоров могут обернуться результатом взаимодействия ряда независимых (и раньше плохо осознанных) факторов, которые с трудом охватываются интуитивно и остаются незаметны даже при большом опыте работы с данным говором.

В связи с этим следует отметить, что без статистического анализа, учитывающего разные типы «мелких» факторов, часто может создаться впечатление, будто соответствующий говор сильно отличается

по данному признаку от литературного языка (или говоров в других регионах). При более пристальном рассмотрении количественных данных может оказаться, что такое впечатление возникло лишь потому, что исследователи приблизительно прикинули, что им казалось «необычным», не вникая в частотность и распределение изучаемых явлений и не оценивая отдельные наблюдения на фоне градуируемых свойств этих явлений. Из примеров, которые обсуждаются ниже, особенно четко это показывает распространение генитива отрицания в западноаукштайтских литовских говорах (см. § 4.2).

Для того чтобы приблизиться к ответам на затронутые выше вопросы, необходимо, во-первых, оценивать соотношение между частотностью в системе (англ. *type frequency*) — что практически соответствует степени распространения данного признака в словарном запасе — и частотностью в тексте (англ. *token frequency*). Это соотношение можно считать достаточно надежным показателем продуктивности. Во-вторых, весьма полезно (а иногда необходимо) применение статистических методов анализа многих переменных, например, регрессии.

Легко понять, что без электронных корпусов практически нельзя установить сочетаемостные тенденции (англ. *collocational patterns*) и совокупности признаков (англ. *feature aggregates*). Ни те, ни другие нельзя увидеть невооруженным глазом, а интуиция часто подводит (или не «наводит»). Как раз применение мультивариантных методов в отношении частотности в тексте позволяет внести важные коррективы в сложившиеся впечатления о распределении признаков (грамматических показателей и т. д.), см., например, Szmrecsányi 2013. Применительно к типологии такие коррективы были внесены с помощью анализа параллельных корпусов; см., например, Levshina 2016 о показателях каузативных отношений, а также Wälchli 2012, который исследовал совокупности признаков на основе текстовой частотности с точки зрения ареальной типологии. При этом также подчеркивается, что внимание не надо сосредотачивать на одних только особенно ярких, необычных свойствах. На самом деле, если своего рода «профиль» для выбранного говора (или языка) определять эмпирическим путем, то эвристическая процедура должна включать и менее очевидные свойства, причем разнородные. В славянских странах единственным исследователем, который подвергал принятые со времен Р. И. Аванесова черты восточнославянских говоров эвристической процедуре на основе совокупностей признаков, была Н. Н. Пшеничнинова (см. Пшеничнинова 1996 и многие статьи начиная с 1970-х гг.). Однако и она пользовалась данными не из связного дискурса (записей текстов), а собранными с помощью опросников для диалектных атласов; тем самым,

ее исходной точкой остались типы, а не текстовая частотность. На уровне типов также осталось исследование А. Эркера (Эркер 2015: 186–193), которое в этом смысле сравнимо с работами Пшеничной.

За пределами славянских стран статистический анализ, опирающийся на электронные корпуса и нацеленный на текстовую частотность признаков, был проведен применительно к английским говорам Великобритании (см. Szmracsányi 2013) и к эстонским говорам (см. Uiboed 2013; Uiboed et al. 2013). В обоих случаях такой анализ показал, что следует пересмотреть традиционное деление исследуемых говоров. Поэтому можно утверждать, что на результаты диалектологических исследований — или исследований, рассматривающих внутреннее членение какого-либо ареала — и, тем самым, на наше понимание языковой действительности выбранных говоров (ареалов) могут существенно повлиять применяемые методы: либо основанные на довольно статических признаках, типах, неградуируемых признаках и без учета динамики, либо основанные на токенах (словоформах) и градуируемых свойствах, принадлежащих разным языковым уровням и на возможности разложения принятых свойств на менее сложные признаки. В общей сложности, анализ совокупностей признаков, анализ соотношения частотностей на уровне типов и на уровне текста, который является предпосылкой для установления продуктивности тех или иных признаков (показателей, конструкций и т. д.), может производиться только с помощью электронных корпусов с разными видами аннотаций. Только корпуса могут предоставить материал для статистических процедур, без которых часто нельзя раскрыть распределение выбранных признаков, нельзя увидеть раннее незамеченные признаки и установить одновременное влияние разных факторов и веса отдельных факторов среди их множества. Эти методы часто не может заменить интуиция опытных исследователей, даже если они очень хорошо знакомы с соответствующим говором. Превосходно показывали это в своих исследованиях смешанной белорусско-русской речи (часто именуемой «трасянкой») Г. Хентшель и его сотрудники (см. Hentschel 2013 и др.).

Особо большой заслугой Г. Хентшеля и его сотрудников следует признать то, что им удалось показать, что распределение белорусских, русских и гибридных форм в речи носителей «трасянки» следует определенным закономерностям (см. Tesch 2014). При этом немаловажно осознать, что особенно большая доля приходится на формы, в которых нельзя определить «русскость» или «белорусскость», поскольку они являются общими для этих близкородственных языков. Результаты, полученные в этом проекте, очень важны и для диалектологии, в частности для диалектной географии, если она готова

не замыкаться на факты, в принципе очевидные. К ним относятся, с одной стороны, процессы нивелирования диалектных различий прежде всего за счет сильного воздействия литературного русского языка⁶ и отсутствия стабильной, укорененной в речевой практике широких слоев общества нормы белорусского языка. С другой стороны, сюда относятся механизмы и тенденции к установлению вариантов узуса в выборе альтернативных («собственно русских», «собственно белорусских», гибридных) форм на разных участках континуума между грамматикой (напр., словоизменительные морфемы) и лексиком (напр., слов, обозначающих «родителей»), а также функциональных слов типа *опять* и *зноў*, *очень* и *натта*). Обнаружить (или хотя бы проверить) такие механизмы и тенденции возможно только с помощью статистического анализа и на основе корпусных данных.

Вместе с тем, электронные корпуса представляют собой незаметное орудие для комплексного исследования языковой вариативности на диатопической и диастратической осях, если в их метаданных учитываются разные сведения по языковой биографии информантов.

Следует отметить еще следующее. Кроме преимуществ для статистического анализа, такие корпуса, если они включают оригинальную запись, намного уменьшают риск ложной интерпретации конкретных языковых фактов, отраженных в естественной речи. При этом следует учесть, что, помимо грамматической аннотации, транскрипция поглощает самую большую часть рабочей силы и времени, особенно если акцент ставится на фонетически верное отражение звучащей речи. Транскрипция звучащей речи всегда в большей или меньшей степени чревата ложной интерпретацией, а невозможность вернуться к оригинальной записи преграждает путь к пересмотру уже вынесенной интерпретации.

В оставшейся части этого раздела мы обсуждаем ряд избранных выбранных явлений, которые стали предметом первых специальных исследований, проведенных с помощью корпуса *TriMCo*. Эти явления широко известны в диалектологической и/или в типологической литературе, но еще не изучались с точки зрения реальной вариативности на корпусной основе.

4.1. Распределение типов предударного вокализма в восточнославянских говорах

Материал восточнославянской части корпуса *TriMCo* послужил основой для масштабного количественного и статистического исследования предударного вокализма белорусских и русских псковских говоров.

⁶ На воздействии этих процессов сосредоточивается упомянутый в §3 проект *Корпус бассейна реки Устья*.

В анализе были задействованы корпусные данные белорусских говоров на территории Беларуси (общее количество релевантных⁷ для исследования словоформ — 39.720), данные белорусских говоров на территории Литвы (количество используемых в анализе словоформ — 19.537), данные белорусских говоров на территории Латгалии (число релевантных вхождений — 3.399), а также данные русских псковских говоров (число релевантных для анализа словоформ — 16.242). Подготовительной стадией исследования стала типизация корпусного материала, т. е. для каждой отдельной релевантной для анализа словоформы был определен тип⁸, характеризующийся различной реализацией в нем следующих факторов: (i) вариант реализации предударного гласного неверхнего подъема (напр., *o* > [a], *o* > [ɤ], *e* > [i], *e* > [a]), (ii) какой гласный в ударном слоге, (iii) палатализован или не палатализован согласный, предшествующий предударному гласному. В итоге были выделены 18 типов, представляющих собой более мелкое дробление принятых в традиционной восточнославянской диалектологии типов предударного вокализма, основанных на понятиях о диссимильности⁹ и недиссимильности¹⁰; см. к примеру, ДАБМ 1963; Аванесов, Орлова 1965; ДАРЯ 1986; Мяцельская, Блінава 1991 среди многих других.

⁷ На основе транскрибированных в программе ELAN корпусных данных с помощью специально для этой цели разработанной И. Хаитом и И. Сержантом программы были созданы экспериментальные таблицы в формате MS Excel. Данные таблицы дают перечень всех используемых носителями диалектной речи словоформ, кроме нерелевантных для исследования, и служат базой для квантитативного и статистического анализа. Нерелевантными для исследования типов предударного вокализма являются словоформы с одним слогом (например, *tam*, *gd'e*), словоформы с ударным начальным слогом (например, *pěsn'a*, *ráma*), а также словоформы с этимологическими гласными верхнего подъема в предударном слоге (напр., *přsát'*, *myčát'*, *stučát'*).

⁸ При этом под «типом» здесь понимается определенная организация фонологической структуры каждой отдельно взятой словоформы.

⁹ В традиционной восточнославянской диалектологии диссимильное аканье характеризуется тем, что гласные [a, o] в первом предударном слоге переходят в [a], если под ударением находятся [i, y, u, e, o]; если же под ударением находится [a], то в предударном слоге на месте [a, o] возникает редуцированный гласный [ɤ], напр., *vědá*, *vady*. Основа диссимильного яканья лежит также в расподоблении гласных предударного и ударного слогов. Существует несколько типов диссимильного яканья: обоянский (архаический), жиздринский, дмитриевский, донской, суджанский, мосальский, шигровский и другие. Подробнее о типах диссимильного яканья см.: Аванесов, Орлова 1965: 38–39, 47–53; Касаткин 1998: 102–108.

Такой более широкий взгляд на систему предупредарного вокализма, т. е. не сведение ее рассмотрения только к позиции диссимилятивности и недиссимилятивности, позволил выявить детальную картину продуктивности отдельных типов реализаций предупредарных /e/, /o/, /a/ по подкорпусам. В таблице 2 представлены доли (в процентах) всех типов реализаций на основе данных, полученных при помощи квантитативного анализа корпуса для каждого из исследуемых здесь говоров.

Таблица 2. Продуктивность типов реализаций /e/, /o/, /a/ в предупредарном слого в белорусских и русских псковских говорах

Типы реализаций	Подкорпус белорусских говоров в Беларуси	Подкорпус русских псковских говоров	Подкорпус белорусских говоров в Литве	Подкорпус белорусских говоров в Латгалии
C'e > [i] (тип <i>všná, všmý</i>)	29,3%	11%	24,7%	45%
C'e > [a] (тип <i>všná, všmý</i>)	45,2%	51%	28,5%	30%
C'e > [e] (тип <i>všná</i>)	14%	31%	40,5%	20%
C'e > другой, кроме [i, e, a] (тип <i>všná</i>)	11,5%	7%	6,3%	5%
Co > [a] (тип <i>kazá, kazý</i>)	99%	83,8%	97%	97%
Co > [ь] (тип <i>kъzá, kъzý</i>)	0,05%	6%	0%	1,4%
Co > [o] (тип <i>molokó</i>)	0,85%	10,1%	2,5%	1,4%
Co > другой, кроме [a, o, ь] (тип <i>pugódy</i>)	0,1%	0,1%	0,5%	0,2%
Ca > [ь] (тип <i>mъvěrna</i>)	0,1%	4,8%	0%	0%
Ca > [a] (тип <i>nazvát'</i>)	99%	94,1%	100%	100%
Ca > другой, кроме [ь, a] (тип <i>molína</i>)	0%	1,1%	0%	0%
C'a > [i] (тип <i>abízátelna</i>)	4%	11%	0,7%	23%
C'a > [a] (тип <i>abázátelna</i>)	95,2%	87%	98,3%	77%
C'a > другой, кроме [i, a] (тип <i>abézátelna</i>)	0,8%	2%	1%	0%

Для реализаций предупредарных /e/, /o/, а также /a/ после непалатализованных согласных в исследование дополнительно было включено рассмотрение вариативности с учетом влияния наличия и отсутствия гласного [a] в ударном слоге, постулируемого в традиционной восточнославянской диалектологии в качестве основополагающего критерия для определения всей системы предупредарного вокализма как диссимилятив-

¹⁰ Под аканьем недиссимилятивного типа понимается нейтрализация предупредарных [a, o] в [a], независимо от гласного под ударением; напр., *vadá, vadý*. Недиссимилятивное (сильное) яканье представляет собой реализацию гласных неверхнего подъема в первом предупредарном слоге после палатализованных согласных как [ʼa], независимо от ударного гласного; напр., *všná, všmý* (Аванесов, Орлова 1965: 38–39, 47–53).

ной или недиссимилятивной (см. выше). Вариативность гласного /e/, как реализация /e/ > [i] или /e/ > [a] при [á] / не [á], представлена в таблицах 3–6.

Таблица 3. Вариативность реализаций /e/ → [i] и /e/ → [a] при [á] / не [á] в белорусских говорах на территории Беларуси

	не [á]		[á]	
/e/ → [i]	1.771	33%	1.074	59%
/e/ → [a]	3.647	67%	744	41%
Всего	5.418	100%	1.818	100%

Результат теста хи-квадрат 397.299 при $p < 0.0001$ с учетом степеней свободы (для $p = 0,05$ при степени свободы (*df*) $1 \rightarrow 3,84$) свидетельствует о высокой сигнификативности данных и отчетливо указывает на наличие статистически значимого различия в вариативности $e \rightarrow [i]$ и $e \rightarrow [a]$ при [á] / не [á] в белорусских говорах на территории Беларуси.

Таблица 4. Вариативность реализаций /e/ → [i] и /e/ → [a] при [á] / не [á] в белорусских говорах на территории Латгалии

	не [á]		[á]	
/e/ → [i]	163	44%	172	97%
/e/ → [a]	211	56%	6	3%
Всего	374	100%	178	100%

Результат теста хи-квадрат 142.249 при $p < 0.0001$ с учетом (*df*) 1 свидетельствует о высокой сигнификативности данных и указывает на наличие статистически значимого различия между группами.

Таблица 5. Вариативность реализаций /e/ → [i] и /e/ → [a] при [á] / не [á] в белорусских говорах на территории Литвы

	не [á]		[á]	
/e/ → [i]	951	42%	369	61%
/e/ → [a]	1.288	58%	232	39%
Всего	2.239	100%	601	100%

Результат теста хи-квадрат 68.2065 при $p < 0.0001$ с учетом (*df*) 1 свидетельствует о высокой сигнификативности данных и указывает на наличие статистически значимого различия между группами.

Таблица 6. Вариативность реализаций /e/ → [i] и /e/ → [a] при [á] / не [á] в русских псковских говорах

	не [á]		[á]	
/e/ → [i]	240	16%	214	23%
/e/ → [a]	1281	84%	735	77%
Всего	1.521	100%	949	100%

Результат теста хи-квадрат 17.856 при $p < 0.0001$ с учетом (*df*) 1 свидетельствует о высокой сигнификативности данных и указывает на наличие статистически значимого различия между группами.

Представленные выше результаты тестов подтверждают наличие статистически значимого распределения /e/ > [i] и /e/ > [a] в зависимости от наличия или отсутствия [a] в ударном слоге во всех исследуемых говорах, но при этом показывают существенные различия в частотности реализаций /e/ > [i] или > [a] по говорам. При наличии [a] под ударением в белорусских говорах Беларуси, Латгалии и Литвы более частотной реализацией является /e/ > [i]. В белорусских говорах Беларуси /e/ > [i] при [á] реализуется в 59% случаев, в белорусских говорах Латгалии — в 97%, в Литве — в 61% случаев. Однако при отсутствии [a] в ударном слоге во всех названных выше говорах более частотной является реализация e > [a]: в Беларуси ее реализация составляет 67% случаев, в Латгалии — 56%, в Литве — 58% случаев. Русские псковские говоры в этом отношении существенно отличаются от белорусских, здесь реализация /e/ > [a] является более частотным вариантом, как при наличии, так и при отсутствии [a] в ударном слоге. При [á] реализация /e/ > [a] составляет 77% от общего количества релевантных случаев, при отсутствии [á] — 84%.

В свою очередь, процентное соотношение реализаций гласных /o/ и /a/ (/Co/ > [Ca], /Co/ > [Cъ], /Ca/ > [Ca], /Ca/ > [Cъ]) с учетом влияния фактора [á] / не [á] в ударном слоге представляет существенно иную картину распределений; см. таблицу 7.

Таблица 7. Вариативность реализаций предударных /o/ и /a/ в зависимости от [á] / не [á] в ударном слоге

Фактор	Фактор не [á]		Фактор [á]		Фактор не [á]		Фактор [á]	
Реализации	/Co/ > [Ca]	/Co/ > [Cъ]	/Co/ > [Ca]	/Co/ > [Cъ]	/Ca/ > [Ca]	/Ca/ > [Cъ]	/Ca/ > [Ca]	/Ca/ > [Cъ]
Подкорпус бел. гов. в Беларуси	99,99%	0,01%	99,8%	0,2%	99,98%	0,02%	99,8%	0,2%
Подкорпус бел. гов. в Литве	100%	0%	100%	0%	100%	0%	100%	0%
Подкорпус бел. гов. в Латгалии	99,6%	0,4%	96%	4%	100%	0%	100%	0%
Подкорпус рус. псков. говоров	94%	6%	92%	8%	95%	5%	93%	7%

Частотность употреблений указывает на значительное преобладание реализации /o/ > [a] над /o/ > [ъ] при наличии и отсутствии [á] в ударном слоге во всех группах говоров.

Если рассматривать полученные здесь результаты с точки зрения понятий о диссимилиативности и недиссимилиативности, доминирующих в концепциях традиционной восточнославянской диалектологии, то в отношении реализаций /e/ тенденция к недиссимилиативной редукции /C'e/ > [C'a] — т. е. к реализации [a] после палатализованного согласного даже перед ударным [á] — наблюдается только в подкорпусе псковских говоров. Эта реализация наиболее соответствует традиционному пониманию недиссимилиативного (или «сильного») яканья; см. таблицу 6. Напротив, белорусские говоры на территории Беларуси, Литвы и Латгалии проявляют явную тенденцию к редукции предударных гласных неверхнего подъема после палатализованных согласных, которую, скорее всего, следовало бы охарактеризовать как диссимилиативное яканье: здесь четко преобладает /C'e/ > [C'a], но только если в ударном слоге не следует [á]; в противном случае реализуется тип /C'e/ > [C'i]. Этим белорусские говоры существенно отличаются от соседних русских псковских говоров. Что касается редукции /o/ (т. е. гласного неверхнего подъема после непалатализованных согласных), то в говорах всех четырех регионов (подкорпусов) довольно четко преобладает тип, совпадающий с недиссимилиативным аканьем (/Co/ > [Ca] или /Ca/ > [Ca], независимо от ударного гласного).

В целом можно констатировать, что дифференциация ареалов между собой существенно различается в зависимости от того, редукция какого гласного неверхнего подъема оценивается: в то время как поведение заднего гласного (т. е. /o/) не обнаруживает сколько-нибудь заметных различий, передний гласный (т. е. /e/) ведет себя весьма по-разному. Как интерпретировать этот факт, в данный момент сказать трудно.

Однако, как бы то ни было, полученные нами результаты также показывают, что подход к описанию предударного вокализма только со стороны диссимилиативности/недиссимилиативности не дает возможности получить представление о системе предударного вокализма конкретного говора с учетом всех представленных в ней возможностей реализаций в их количественном соотношении; о чем свидетельствуют данные, обобщенные в таблице 2.

4.2. Генитив отрицания в литовских говорах (переходные глаголы)

Данные языков балто-славянского ареала дают прекрасный материал для изучения вариативности в падежном маркировании, которая часто выступает объектом квантитативных исследований. На основе данных литовской части корпуса был проведен анализ того, как маркируется прямой объект переходных глаголов при отрицании; подробнее см. Kozhanov 2017.

Как правило, прямой объект в литовском языке оформляется винительным падежом, но при наличии у глагола отрицания падеж прямого объекта меняется на родительный.

- | | | | | | | |
|-----|----|------------------|---------------------|----|---------------------|-----------------------|
| (1) | a. | <i>Mat-au</i> | <i>med-j</i> | b. | <i>Ne-mat-au</i> | <i>medži-o</i> |
| | | see-PRS.ISG | tree-ACC.SG | | NEG-see-PRS.ISG | tree-GEN.SG |
| | | ‘Я вижу дерево.’ | | | ‘Я не вижу дерева.’ | |

Само это явление под общим названием «генитив отрицания» или «партитив отрицания» известно и в других языках региона (литература, посвященная этому вопросу, чрезвычайно обширна, здесь сошлемся на недавний обзор в: Аркадьев 2017). Степень обязательности, с которой аккузатив меняется на родительный, зависит как от конкретного языка, так и от контекста отрицания. В литературном литовском генитив отрицания обязателен в локальных контекстах (прямой объект непосредственно отрицаемого предиката), в то время как вариативность обнаруживается в дистантных контекстах (прямой объект при нефинитной глагольной форме, зависящей от предиката под отрицанием), подробнее см. Arkadiev 2016, где прямой объект может быть как маркирован генитивом, так и сохранять аккузатив.

За исключением небольшого количества импрессионистических наблюдений (в частности относительно западноаукштайтских говоров, см. ниже), «витальность» генитива отрицания никогда не проверялась на материале диалектов. Так как литовская часть корпуса обладает морфологической аннотацией, поиск релевантных примеров не составил большого труда. Контексты были разделены на два типа — локальные и дистантные. Анализ данных показал, что генитив преобладает в контекстах обоих типов (98% и 87% соответственно). Более низкий процент генитивного маркирования в дистантных контекстах объясняется не столько преобладанием аккузатива (имеется лишь один пример), сколько активностью так называемых номинативных объектов в восточноаукштайтской части корпуса (т. е. корпусные примеры показывают, что номинативные объекты не реагируют на отрицание при глаголе). В этих диалектах номинативные объекты, т. е. прямые объекты, которые маркируются номинативом, чаще всего встречаются при инфинитивах, которые зависят от глагола *reikėti* ‘быть необходимым’, ср.:

- | | | | |
|-----|---|-----------------------|--|
| (2) | <i>raik-s</i> | <i>ko:ji-a</i> | <i>atⁱ-pji/Au-tie</i> ¹¹ |
| | надо-FUT3 | нога-NOM.SG | PVB-резать-INF |
| | ‘надо будет ногу (докл. нога) отрезать’ | | |

(Šiulėnai, Игналинский р-н)

¹¹ Заглавными буквами обозначены ударные гласные, двоеточие (:) указывает на долготу гласных, а индексный *j* (*i*) — на палатализованность согласных.

Преимущество квантитативного подхода к описанию генитива отрицания в литовских говорах можно хорошо проиллюстрировать следующим примером. Часто в описаниях западноаукштайтских говоров диалектологи отмечают редкое использование генитива отрицания, ср. утверждение о том, что в говорах рядом с населенным пунктом Гришкабудис (лит. *Griškabūdis*) генитив отрицания «почти не употребляется» (Vacevičiūtė, Sakalauskienė 2008: 32). Однако подсчет всех примеров прямого объекта при отрицании, взятых из опубликованных материалов по этим говорам, дает совершенно другие результаты. Хотя в этих говорах аккузатив действительно встречается чаще, чем, например, в юго-восточных литовских диалектах или литературном языке (отсюда и впечатление диалектологов о «редкости» генитива отрицания), в реальном употреблении генитивное маркирование или преобладает, или выступает так же часто, как и аккузативное.

4.3. Диапазон функций творительного падежа в литовских и восточнославянских говорах

Хотя вопрос о вариативности в маркировании семантической роли ‘инструмент’ — беспредложным творительным или творительным с предлогом ‘с’¹² — уже рассматривался на материале языков балто-славянского ареала, см. Непокупный 1968: 54–88, наличие корпуса (а следовательно, и данных, собранных принципиально другим способом — для литовской части исследования А. П. Непокупный использовал ответы на два вопроса из анкеты к диалектологическому атласу) позволяют взглянуть на проблему по-другому (подробнее см. Kozhanov, рукопись).

Для проведения корпусного исследования сначала были собраны все примеры, маркированные творительным падежом. Затем из них были отобраны примеры, относящиеся к семантической зоне ‘инструмент’. Эта семантическая роль была поделена нами на четыре подтипа: i) непосредственно орудие, ср. *резать ножом*, ii) средство, *писать чернилами*, iii) часть тела, ср. *рвать траву руками*, и iv) средство передвижения, ср. *приехать конём*. Только в русской части корпуса встретилось несколько примеров, которые можно обозначить как ‘стихийная сила’, ср. *ударило грозой*. Кроме семантической роли, все примеры также получили пометы, содержащие информацию о языке (литовский, русский, белорусский) и регионе (для каждого языка отдельно).

Результаты квантитативного анализа примеров с общей семантикой ‘инструмент’, которые кодируются творительным падежом с или без предлога ‘с’, показали, что, во-первых, языки региона допускают разную степень вариативности в маркировании данной обобщённой

¹² Лит. *su*, рус. *с*, бел. *с/з*.

семантической роли: наибольшая вариативность наблюдается в литовском, тогда как в белорусском и русском языках вариативность сравнительно низкая, ‘инструмент’ обычно кодируется беспредложным творительным падежом. Иными словами, существует статистически значимая зависимость между наличием или отсутствием предлога ‘с’ и принадлежностью примера к определенному языку. Во-вторых, существует статистически значимая зависимость между семантическим типом ‘инструмента’ и наличием предлога ‘с’. Чаще всего творительным с предлогом ‘с’ маркируются транспортные средства, несколько реже аргументы, обозначающие орудие, совсем редкие аргументы с семантикой ‘средство’ и ‘части тела’.

Корпусные данные позволяют также с большей точностью взглянуть на географическое распределение примеров. Например, в обобщающем тексте А. П. Непокупного, посвященном вариативности в литовских диалектах, предлагается «противопоставление „предложного“ северо-запада „беспредложному“ юго-востоку» (Непокупный 1964: 56). Представленные в корпусе *TriMCo* данные из «беспредложного» юго-востока Литвы показывают, однако, более тонкую дистрибуцию. Для исследования были выделены три зоны — i) восточноаукштайтские говоры, ii) южноаукштайтские говоры в Литве и iii) южноаукштайтские говоры в Беларуси. Хотя во всех этих зонах примеры беспредложного инструмента преобладают над предложным, их частотность различается: в целом предложное маркирование инструмента частотнее в южноаукштайтских говорах на территории Литвы, чем в восточноаукштайтских говорах или даже близкородственных южноаукштайтских диалектах, но на территории Беларуси.

Наличие корпуса (хотя здесь необходимо сделать оговорки о размере корпуса и количестве примеров) позволяет также обнаружить более тонкие факторы, определяющие вариативность. Так, анализ примеров со значением ‘транспортное средство’ из литовского корпуса выявил статистически значимую зависимость между типом маркирования и транзитивностью глагола, управляющего аргументами: непереходные глаголы чаще допускают маркирование с предлогом ‘с’, чем переходные глаголы.

5. Итоги

В этой статье мы показали результаты первых попыток проведения количественных анализов с помощью подкорпусов *TriMCo*. Подготовка таких корпусов — очень трудоемкое занятие, особенно если требуется надежная морфологическая аннотация. Понятно также, что корпуса никогда не могут охватить всю вариативность, которая «чувствуется»

опытными исследователями и знатоками соответствующих говоров. Однако вряд ли какой-либо, пусть и опытный, исследователь в состоянии оценить распределение вариантов того или иного явления, так же как без помощи адекватных статистических методов нельзя взвесить «удельный вес» каждого отдельного фактора по отношению к другим. В связи с этим выбранные нами случаи из балто-славянской контактной зоны подтверждают то, что известно уже другим представителям «вариационной лингвистики» (см. введение в § 4): классификация говоров (или других разновидностей того или иного языка) оказывается другой, если за основу сравнения берется большее количество факторов, и то на основе токенов речи. Картина часто меняется, если за основу сравнения берутся примеры из живой речи, а не разрозненные наблюдения (или впечатления), которые не могут поспеть за потоком речи, или данные, основанные на типах, как, например, в опросниках или атласах.

Эти последние также не в состоянии охватить всю вариативность, но не столько из-за ограниченности материала (как в случае корпусов), сколько потому, что при опросах нельзя оценить одновременное влияние разных факторов. К тому же корпуса имеют больше шансов, чем опросники, отражать особенности спонтанной речи. Это обстоятельство дает корпусам преимущество, если учитывать, что между тем, что в том или ином языке или говоре возможно сказать, и тем, что действительно и обычно говорят, могут быть немалые различия.

Таким образом, квантитативное корпусное исследование балто-славянского пограничья не только предлагает большие возможности для исследования отдельных явлений, в той или иной степени описанных в литературе, но и задает новую планку для диалектологических исследований в балтийских и восточнославянских лингвистических традициях.

ЛИТЕРАТУРА

- Аванесов, Орлова 1965 — *Р. И. Аванесов, В. Г. Орлова*. Русская диалектология. М., 1965.
- Аркадьев 2017 — *П. М. Аркадьев*. Объектный партитив отрицания: ареально-типологическая перспектива // *Acta linguistica Petropolitana*. СПб., 2017. Т. XIII. Ч. 3.
- Герд 2000 — *А. С. Герд*. Несколько замечаний касательно понятия «диалект» // *Русский язык сегодня* 1. М., 2000.
- ДАБМ 1963 — *Р.І. Аванесаў, К. К. Крапіва, Ю. Ф. Мацкевіч* (ред.). Дыялекталагічны атлас беларускай мовы. Мінск, 1963.
- ДАРЯ 1986 — *Р. И. Аванесов, С. В. Бромлей* (ред.). Диалектологический атлас русского языка. Центр Европейской части СССР. Выпуск I: Фонетика. М., 1986.

- Касаткин 1998 — *Л. Л. Касаткин*. Некоторые фонетические особенности современных русских говоров, описанных в начале XX века О. Брокком // J. I. Bjørnflaten, G. Kjetsaa, T. Mathiassen (eds.). *A Centenary of Slavic Studies in Norway. Olaf Broch Symposium*. Papers. Oslo, 12–14 September 1996. Oslo, 1998.
- Мяцельская, Блінава 1991 — *Е. С. Мяцельская, Э. Д. Блінава*. Беларуская дыялекталогія. Практыкум. Мінск, 1991.
- Непокупный 1964 — *А. П. Непокупный*. Ареальные аспекты балто-славянских языковых отношений. Киев, 1964.
- Пшеничнова 1996 — *Н. Н. Пшеничнова*. Типология русских говоров. М., 1996.
- Трубинский 1984 — *В. И. Трубинский*. Очерки русского диалектного синтаксиса. Ленинград, 1984.
- Эркер 2015 — *А. Эркер*. Структурные черты смешанных белорусских говоров на балто-славянском пограничье. Leipzig, 2015.
- Arkadiev 2016 — *P. Arkadiev*. Long-distance genitive of negation in Lithuanian // A. Holvoet, N. Nau (eds.). *Argument Realization in Baltic*. Amsterdam, Philadelphia, 2016.
- Vacevičiūtė, Sakalauskienė 2008 — *R. Vacevičiūtė, V. Sakalauskienė*. Griškabūdžio apylinkių tekstai. Vilnius, 2008.
- De Vogelaer, Seiler 2012 — *G. De Vogelaer, G. Seiler*. The dialect laboratory: Introductory remarks // G. De Vogelaer, G. Seiler (eds.). *The Dialect Laboratory (Dialects as a testing ground for theories of language change)*. Amsterdam–Philadelphia, 2012.
- ELAN — <https://tla.mpi.nl/tools/tla-tools/elan/>.
- Hentschel 2013 — *G. Hentschel*. Zwischen Variabilität und Regularität, „Chaos“ und Usus: Zu Lautung und Lexik der weißrussisch-russischen gemischten Rede // G. Hentschel (ed.). *Variation und Stabilität in Kontaktvarietäten: Beobachtungen zu gemischten Formen der Rede in Weißrussland, der Ukraine und Schlesien*. Oldenburg, 2013.
- Kozhanov 2017 — *K. Kozhanov*. Studying variation in case marking: The genitive of negation in Aukštaitian dialects of Lithuanian // *Baltic Linguistics*. 2017. Vol. 8.
- Kozhanov, рукопись — *K. Kozhanov*. Variation in instrumental marking: a case-study of South-Eastern Lithuanian dialects.
- Levshina 2016 — *N. Levshina*. Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages // *Folia Linguistica*. 2016. Vol. 50. Issue 2.
- SpoSla* — Slavic Spoken Corpora: <http://parasolcorpus.org/Spoken-Slavic/>.
- Szmrecsányi 2013 — *B. Szmrecsányi*. Grammatical variation in British English Dialects (A Study in Corpus-Based Dialectometry). Cambridge etc., 2013.

- Tesch 2014 — *S. Tesch*. Morphological hybrids: Belarusian-Russian word forms in Belarusian Trasjanka // G. Hentschel, O. Taranenko, S. Zaprudski (eds.). *Trasjanka und Suržyk — gemischte weißrussisch-russische und ukrainisch-russische Rede (Sprachlicher Inzest in Weißrussland und der Ukraine?)*. Frankfurt/M. etc., 2014.
- TriMCo* — Triangulation Approach for Modelling Convergence with a High Zoom-In Factor: www.trimco.uni-mainz.de.
- Uiboaed 2013 — *K. Uiboaed*. Verbiühendid eesti murretes [Verb Constructions in Estonian Dialects]. Tartu, 2013.
- Uiboaed et al. 2013 — *K. Uiboaed, C. Hasselblatt, L. Lindström, K. Muischnek, J. Nerbonne*. Variation of Verbal Constructions in Estonian Dialects // LLC: Journal of Digital Scholarship in the Humanities, 2013.
- Ustyia River Basin Corpus — <http://www.parasolcorpus.org/Pushkino/login.php>.
- Wälchli 2012 — *B. Wälchli*. Grammaticalization clines in space: Zooming in on synchronic traces of diffusion processes // B. Wiemer, B. Wälchli, B. Hansen (eds.). *Grammatical Replication and Borrowability in Language Contact*. Berlin–Boston, 2012.
- Wiemer 2013 — *B. Wiemer*. Zur arealen Stufung im baltisch-slavisches Kontaktgebiet (und dabei auftretenden methodischen Desideraten) // S. Kempgen, M. Wingender, N. Franz, M. Jakiša (eds.). *Deutsche Beiträge zum 15. Internationalen Slavistenkongress, Minsk 2013*. München, 2013.
- Wiemer, Seržant 2014 — *B. Wiemer, I. A. Seržant*. East Slavic Dialectology: Achievements and Perspectives of Areal Linguistics // I. A. Seržant, B. Wiemer (eds.). *Contemporary Approaches to Dialectology. The Area of North, North-West Russian and Belarusian Dialects*. Bergen, 2014.
- Wiemer, Seržant, Erker 2014 — *B. Wiemer, I. Seržant, A. Erker*. Convergence in the Baltic-Slavic contact zone (Triangulation approach) // J. Besters-Dilger, C. Dermarkar, S. Pfänder, A. Rabus (eds.). *Congruence in Contact-induced Language Change (Language Families, Typological Resemblance, and Perceived Similarity)*. Berlin, New York, 2014.
- Wiemer 2019 — *B. Wiemer*. *Matrěška* and areal clusters involving varieties of Slavic: On methodology and data treatment // A. Danylenko, M. Nomachi (eds.). *Slavic in the Language Map of Europe*. Berlin, Boston (*TiLSM*).

СВЕДЕНИЯ ОБ АВТОРАХ

Руфь Александровна АГЕЕВА

: ruf09382008@yandex.ru

кандидат филологических наук, заместитель председателя Топонимической комиссии Московского отделения Российского географического общества

Петр Михайлович АРКАДЬЕВ

: alpgurev@gmail.com

кандидат филологических наук, старший научный сотрудник Института славяноведения РАН, доцент Института лингвистики РГГУ (Москва)

Римантас Бальсис (Rimantas Balsys)

: Rimantas.Balsys@ku.lt

доктор гуманитарных наук, профессор, декан Факультета социальных и гуманитарных наук Клайпедского университета (Литва)

Татьяна Бартеле (Tatjana Bartele)

: tmb@inbox.lv

доктор истории, ведущий исследователь

Рижской высшей школы экономики, искусств и технологий (Латвия)

Вацлав Блажек (Václav Blažek)

: blazek@phil.muni.cz

хабилированный доктор филологии, профессор
Университета им. Масарика (Брно, Чехия)

Сергей Григорьевич Болотов

: baltaitis@yandex.ru

научный сотрудник Института мировой культуры
МГУ им. М. В. Ломоносова (Москва)

Скирмантас Валянтас (Skirmantas Valentas)

доктор гуманитарных наук, профессор : skirmantas.valentas@gmail.com
Шяуляйского университета (Литва)

Бьёрн Вимер (Björn Wiemer)

: wiemerb@uni-mainz.de

хабилированный доктор филологии, профессор
Майнцкого университета имени Иоганна Гутенберга (Германия)

Татьяна Васильевна Володина (Тацяна Васільеўна Валодзіна)

доктор филологических наук, заведующая отделом : tanja_volodina@tut.by
Центра исследований белорусской культуры, языка и литературы
Национальной академии наук Беларуси

Владимир Антонович Дыбо

: vdybo@mail.ru

доктор филологических наук, академик РАН,
главный научный сотрудник Института славяноведения РАН (Москва)

Илья Борисович Иткин

: ilya.borisovich.itkin@gmail.com

кандидат филологических наук, старший научный сотрудник
Института востоковедения РАН, доцент Школы филологии,
Факультета гуманитарных наук НИУ ВШЭ (Москва)

Юозас Карацейус (Juozas Karasiėjus)

доктор филологии (Вильнюс, Литва)
Baltistikos katedra, Vilniaus universitetas, Universiteto g. 5, LT-01513, Vilnius

Кирилл Александрович Кожанов

: kozhanov.kirill@gmail.com

кандидат филологических наук, научный сотрудник
Института славяноведения РАН (Москва)

- Сергей Александрович Крылов** ✉: krylov-58@mail.ru
доктор филологических наук, ведущий научный сотрудник
Института востоковедения РАН (Москва)
- Нийоле Лауринкене (Nijolė Laurinkienė)** ✉: nlaurinkiene@yahoo.com
хабилированный доктор гуманитарных наук, старший научный сотрудник
Института литовской литературы и фольклора (Вильнюс, Литва)
- Владимир Александрович Лобач (Уладзімір Аляксандравіч Лобач)**
доктор исторических наук, профессор ✉: nordic972@gmail.com
Полоцкого государственного университета (Беларусь)
- Евгения Львовна Назарова** ✉: ezis08@gmail.com
кандидат исторических наук, старший научный сотрудник
Института всеобщей истории РАН (Москва)
- Сергей Львович Николаев** ✉: sergenicko@mail.ru
доктор филологических наук, ведущий научный сотрудник
Института славяноведения РАН (Москва)
- Михаил Владимирович Ослон** ✉: neoakut@gmail.com
кандидат филологических наук (Львов, Украина)
- Владимир Яковлевич Петрухин** ✉: vladimir.petrukhin@gmail.com
доктор исторических наук, главный научный сотрудник
Института славяноведения РАН (Москва)
- Дайнюс Разаускас (Dainius RAZAUSKAS)** ✉: dainius.razauskas@lnkc.lt
доктор гуманитарных наук, старший научный сотрудник
Института литовской литературы и фольклора (Вильнюс, Литва)
- Михаил Николаевич Саенко** ✉: veraetatis@yandex.ru
кандидат филологических наук, научный сотрудник
Института славяноведения РАН
- Елена Анатольевна Сведенцова** ✉: svedentsova@gmail.com
лаборант-исследователь Лаборатории моделирования речевой деятельности
филологического факультета
Санкт-Петербургского государственного университета
- Рамуне Шмигельските-Стукене (Ramunė Šmigelskytė-Stukienė)**
доктор гуманитарных наук, ✉: smigelskyte.stukiene@gmail.com
старший научный сотрудник Института истории Литвы (Вильнюс)
- Аксана Эркер (Aksana ERKER)** ✉: aksana.erker@gmx.de
доктор филологии, доцент языковой Академии
и Академии для детей (Кирхгейм, Германия)
- Артурас Юджентис (Artūras JUDŽENTIS)** ✉: arturas.judzentis@gmail.com
доктор гуманитарных наук, доцент
Вильнюсской художественной академии (Литва)

Научное издание

Балто-славянские исследования

XX

Сборник научных трудов

Подписано в печать 14.12.2019.

Формат 60 × 90¹/16.

Гарнитуры:

Times New Roman

© 2017 The Monotype Corporation (Monotype Imaging, Inc.)

(дизайн: Monotype Type Drawing Office — Stanley Morison, Victor Lardent, 1932).

Їзнітса © 2009 ParaType, Inc.

(дизайн: НПО «Полиграфмаш» — Светлана Ермолаева, 1988)

Печ. л. 30. Усл. печ. л. 30,13. Уч.-изд. л. 27,07.

Бумага офсетная. Цифровая печать.

Тираж 300 экз.

Заказ № 7854.

Отпечатано с электронной версии в типографии ООО «Буки Веди».
115 093, г. Москва, М-93, Партийный переулок, д. 1, корп. 58, стр. 2.

☎: +7-495-926 63 96, 8-800-333 42 06, 10:00 – 19:00.

Сайт: <http://www.bukivedi.com>; ✉: info@bukivedi.com.