

**Герд Хентшель, Ян Патрик Целлер, Светлана Теш**

**Ольденбургский корпус  
белорусско-русской смешанной речи:  
ОК-БРСР**

**Документация**

*Состояние на 01.04.2014*



BIS-Verlag der Carl von Ossietzky Universität Oldenburg

ISBN: 978-3-8142-2337-7

## Содержание

1	Введение	1
2	Говорящие, их социальная характеристика и объем материала в корпусе	5
2.1	Объем материала отдельных говорящих	5
2.1.1	В семейном корпусе	5
2.1.2	В корпусе интервью	7
2.2	Города и респонденты	8
2.2.1	В семейном корпусе	8
2.2.2	В корпусе интервью	12
2.3	Возрастная структура респондентов и доля материала разных возрастных групп	15
2.3.1	В семейном корпусе	15
2.3.2	В корпусе интервью	17
2.4	Соотношение полов респондентов	19
2.4.1	В семейном корпусе	19
2.4.2	В корпусе интервью	21
2.5	Образование респондентов	23
2.5.1	В семейном корпусе	23
2.5.2	В корпусе интервью	25
2.6	О миграции из деревни в город	27
2.6.1	В семейном корпусе	27
2.6.2	В корпусе интервью	29
2.7	Вывод	31
3	Структура транскрипции в системе СНАТ	32
3.1	Названия файлов	32
3.2	Основные черты структуры транскрипции	33
3.3	Основная информация о разговорах	34
3.3.1	@Languages: Типы говорящих, принимающие участие в разговорах	34
3.3.2	@Participants: Роль или миграционный статус говорящих	36
3.3.3	@ID: Информация о говорящих	37
3.3.3.1	Тип говорящих	38
3.3.3.2	Корпус	38

3.3.3.3	Сокращенное обозначение говорящего / респондента	38
3.3.3.4	Представленность в корпусе	39
3.3.3.5	Миграционный статус / Роль в интервью	39
3.3.3.6	Образование	40
3.3.4	@L1: Первый язык	40
3.3.5	@FL: Язык семьи в детстве	40
3.3.6	@MT: Родной язык	41
3.3.7	@Comment: Комментарии к транскрипции	41
3.3.8	@Situation: Речевая ситуация	41
3.4	Воспроизведение разговоров	41
3.4.1	Деление речи на высказывания	42
3.4.2	Транскрипция словоформ	43
3.4.2.1	Гласные в ударной позиции	43
3.4.2.2	Гласные в первом предударном слоге	44
3.4.2.3	Гласные в других предударных и послеударных слогах	46
3.4.2.4	Особые случаи транскрипции гласных	47
3.4.2.5	Транскрипция согласных	48
3.4.2.6	Палатализация	48
3.4.2.7	Различия согласных в белорусском и русском языках	48
3.4.2.8	Передача результатов некоторых фонологических процессов и другие явления	50
3.4.2.9	Особые случаи	51
3.4.3	Языковая афинность	52
3.4.3.1	Фонетико-фонологический уровень	53
3.4.3.2	Морфологический уровень	55
3.4.4	Дополнительная информация по словоформам	59
3.4.5	Пунктуация	60
3.4.5.1	Простые знаки препинания	60
3.4.5.2	Прямая речь	61
3.4.5.3	Метаязыковые элементы	61
3.4.5.4	Повторения, исправления, переформулировки, переводы	62
3.4.5.5	Неполные предложения	63
3.4.5.6	Паузы и наложения речи одного говорящего на речь другого	64
3.5	Аннотации в строке %tag	64
3.5.1	Языковая афинность	65

3.5.2	Часть речи	66
3.5.3	Заголовочное слово	68
3.5.4	Склонение	73
3.5.4.1	Падеж	73
3.5.4.2	Род	74
3.5.4.3	Число	74
3.5.5	Спряжение	74
3.5.5.1	Время и повелительное наклонение	75
3.5.5.2	Лицо	75
3.5.5.3	Род	75
3.5.5.4	Число	76
4	Работа с ОК-БРСР в программе CLAN	77
4.1	Введение	77
4.2	Подготовка корпуса	77
4.3	Поиск словоформ	79
4.4	Поиск аннотаций в строке %tag	80
4.5	Поиск знаков препинания	81
4.6	Ограничение поиска по отдельным частям корпуса / городам / миграционному статусу	82
5	Список литературы	83

## 1 Введение

Белорусско-русская смешанная речь, т.е. частое чередование в речи «белорусских» и «русских» элементов и структур, представляет собой феномен, встречающийся у многих белорусов, по крайней мере, в определенных речевых ситуациях. *Ольденбургский корпус белорусско-русской смешанной речи* (ОК-БРСР) является первым объемным корпусом данного актуального явления, возникшего в результате языкового контакта между двумя близкородственными языками. Корпус был создан в рамках проекта *Трасянка в Беларуси: Смешанный код как продукт белорусско-русского языкового контакта. Языковое структурирование, социолингвистические механизмы идентификации и социальная «экономика» языка*, который проводился под руководством Герда Хентшеля (институт славистики, Ольденбургский университет) и Бернгарда Киттеля (раньше институт социальных наук, Ольденбургский университет / сегодня институт экономической социологии, Венский университет) в сотрудничестве с Сергеем Запрудским (кафедра истории белорусского языка, БГУ, Минск) и Давидом Ротманом (центр социологических и политических исследований, БГУ, Минск) в период с 2008 по 2013 гг. при финансовой поддержке фонда Фольксваген.

Несмотря на широкую распространенность в Беларуси, белорусско-русская смешанная речь долгое время не привлекала к себе внимания лингвистов. Только в 1990-е годы она стала объектом лингвистического исследования, причем заслуга в этом принадлежит Геннадию Цыхуну и Нине Мечковской, которые первыми ввели данный феномен в научную дискуссию (ср., напр., Цыхун 2013 [1998], Мечковская 2006, Мячкоўская 2007).<sup>1</sup> В Беларуси данная форма речи обычно называется «трасянкой», причем первоначально слово «трасянка» обозначало смешанное с соломой сено, т.е. «плохой» корм для скота (подробно о происхождении данного термина см. Цыхун 2013 [1998]). Т.к. термин «трасянка» имеет сильную негативную коннотацию – о чем свидетельствует уже сама метафоричность термина – и содействует

---

1 Более новое описание белорусской «трасянки» представлено в ХЕНТШЕЛЬ (2013, 2015), HENTSCHEL 2013, а также в сборнике HENTSCHEL, TARANENKO & ZAPRUDSKI 2014, в котором также рассматривается подобное явление в Украине – украинско-русский суржик.

стигматизации людей, использующих данную форму речи, в дальнейшем рассматриваемый феномен будет называться исключительно «белорусско-русская смешанная речь», или же сокращенно «БРСР». Такой подход необходим и по той причине, что некоторые характеризуют «трасянку» как речь людей, не владеющих ни русским, ни белорусским языком в «приемлемой» форме. А если к тому же рассматривать «трасянку» как характерный признак малообразованных людей, то получается замкнутый круг (ср. ХЕНТШЕЛЬ 2013). Наше исследование показывает, что носители БРСР образуют очень неоднородную группу и что смешанная речь используется также и людьми, хорошо владеющими русским, а частично также и белорусским языком. Недостаточная образованность говорящих на «трасянке» – это всего лишь негативный стереотип (ср. КИТТЕЛ et al. 2010; HENTSCHEL & KITTEL 2011a/b).

ОК-БРСР содержит языковой материал записей белорусско-русской смешанной речи, собранных в период с 2008 по 2014 гг. в восьми белорусских городах (ср. HENTSCHEL 2008с, 2013). Во-первых, были записаны спонтанные разговоры в кругу семьи и друзей (*семейный корпус*), а во-вторых, также открытые интервью с теми говорящими, которые «признали» себя носителями смешанной речи (*корпус интервью*).

В *семейном корпусе* содержатся непринужденные, спонтанные разговоры в семейном и дружеском кругу, причем каждый город представлен одной семьей<sup>2</sup>. Таким образом, участниками разговоров являются члены соответствующих семей, а также их родственники, друзья, соседи и иногда коллеги по работе, с которыми выбранные семьи тесно общаются. При проведении записей использовался метод включенного наблюдения. С целью получить как можно более аутентичную речь для проведения записей были выбраны люди, принадлежащие к рассматриваемым семьям либо к кругу их друзей. Включенное записывающее устройство находилось в помещении, причем некоторые участники разговора знали, что разговор записывается, другие же нет. В таком случае им было сообщено об этом позже. Т.к. диктофон, как правило, оставался включенным на протяжении не одного часа, то можно было рассчитывать на естественное, непринужденное ведение разговора даже теми

---

2 Только в одном случае, в городе Октябрьский, данные были собраны в двух семьях.

говорящими, которые знали о проведении записи. Проанализированы были записи приблизительно в объеме 36 часов, причем выбраны были такие отрывки, речь в которых определено была непринужденной. Так, например, не рассматривались отрывки разговоров сразу же после включения диктофона, отрывки с сильным шумовым фоном, отрывки, в которых несколько человек разговаривают одновременно, а также отрывки, в которых содержание разговора имеет очень личный или по разным причинам деликатный характер.<sup>3</sup>

*Корпус интервью* содержит языковой материал из открытых интервью, проведенных в рамках социологической части Ольденбургского проекта. Вначале в центре внимания социологического исследования находились закрытые интервью с 1400 респондентами, выбранными при помощи метода случайной выборки (ср. KITTEL et al. 2010; HENTSCHEL & KITTEL 2011a/b). Далее из этих 1400 опрошенных была выбрана группа респондентов, которые в закрытых интервью назвали БРСР своим основным средством коммуникации. С ними были проведены открытые интервью, составляющие вторую часть социологического исследования Ольденбургского проекта (ср. KITTEL & LINDNER 2011). Из открытых интервью были выбраны отрывки, содержащие смешанную речь, которые и составляют языковой материал корпуса интервью. Выбранные для интервью респонденты никоим образом не были связаны с говорящими из семейного корпуса и даже не были знакомы между собой.

В общей сложности семейный корпус содержит приблизительно 212000 словоформ, образующих 39000 высказываний из 444 разговоров различной продолжительности. Корпус же интервью состоит из приблизительно 170000 словоформ в 23000 высказываний из 54 разговоров. Общая информация о респондентах и их социодемографическая характеристика представлена во второй главе данной документации.

ОК-БРСР доступен для свободного пользования на странице <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/>, где представлены аудиофайлы для прослушивания и скачивания, а также корпус в трех различных письменных версиях:

---

3 Однако разговоры такого типа не были исключены из разного рода анализа данных, проводимого в рамках проекта.

- Во-первых, простая версия доступна для чтения прямо в интернете; причем здесь же есть возможность прослушать некоторые разговоры. Данная версия была затранскрибирована в соответствии с правилами СНАТ<sup>4</sup>, разработанными в рамках проекта CHILDES (с небольшими изменениями, см. ниже). Она предназначена для простого чтения «на экране» и поэтому не содержит дополнительных грамматических характеристик, а также информации о так называемой языковой аффинности (т.е., вкратце, можно ли определенную словоформу определить как ‚белорусскую‘, ‚русскую‘, ‚общую‘ или ‚гибридную‘; см. главу 3.4.3). В данной версии можно пользоваться только обычными инструментами поиска, имеющимися в веб-обозревателе.
- Во-вторых, возможно скачивание ОК-БРСР в формате СНАТ, в качестве текстовых файлов в кодировке UTF8, содержащих дополнительную грамматическую информацию и информацию о языковой аффинности. Данную версию ОК-БРСР можно анализировать при помощи инструментов программы CLAN (ср. McWHINNEY 2013b). Подробнее об этом в главах 3 и 4.
- В-третьих, для интересующихся исключительно текстом обе части корпуса ОК-БРСР выставлены в формате PDF. Эта версия содержит только транслитерированный в соответствии с описанными здесь правилами текст, без названных ниже формальных обозначений формата СНАТ и без дополнительной (например, грамматической) информации. Социальная характеристика участвующих в разговоре респондентов представлена здесь в более доступной для читателя форме, чем в формате СНАТ. А также в отличие от других версий в данной версии в словах обозначено ударение.

ОК-БРСР представляет собой самостоятельную публикацию, которую можно цитировать следующим образом:

HENTSCHEL, G.; ZELLER, J. P.; TESCH, S. 2014: *Das Oldenburger Korpus zur weißrussisch-russischen gemischten Rede: OK-WRGR*. Oldenburg: BIS [<http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/>] [ISBN: 978-3-8142-2305-6]

---

4 По состоянию правил СНАТ на 2013 год; ср. McWHINNEY 2013a.



Список опубликованных в рамках Ольденбургского проекта работ, основанных на анализе ОК-БРСР (а также других публикаций в рамках Ольденбургского проекта), можно найти здесь: <https://www.uni-oldenburg.de/slavistik/forschung/sprachwissenschaft/schwerpunkt-mischvarietaten/oldenburger-publikationen-zur-weissrussisch-russischen-gemischten-rede/>.

## **2 Говорящие, их социальная характеристика и объем материала в корпусе**

### **2.1 Объем материала отдельных говорящих**

#### *2.1.1 В семейном корпусе*

Семейный корпус содержит языковой материал в общей сложности 129 говорящих. Конечно же, объем материала разных говорящих не одинаков, что соответствует характеру коммуникации в семьях. Как правило, из всех членов семьи обычно 2 или 3 человека участвуют в разговорах наиболее активно. В большинстве семейных социальных структур это чаще всего женщины из поколения, образующего ядро семьи. Остальные члены семьи принимают меньшее участие в разговорах. Кроме того, в разговоры были вовлечены люди, не являющиеся членами рассматриваемых семей, а только оказавшиеся у них в гостях во время проведения записи. К ним относятся более дальние родственники, друзья и знакомые, а также соседи и коллеги по работе. Вследствие этого объем языкового материала разных говорящих сильно различается. Минимальный объем материала в корпусе у респондента, произнесшего всего лишь одну единственную словоформу, а максимальный объем – у респондента с 12529 словоформами.<sup>5</sup> (Так как средняя длина высказываний, которые в большинстве случаев, однако не всегда соответствуют предложениям, составляет приблизительно от пяти до шести словоформ, получается,

---

<sup>5</sup> Количественные данные в данной главе основываются на используемой в проекте версии корпуса, содержащей также определенные отрывки личного либо «деликатного» характера, которые были удалены из опубликованной версии корпуса. Однако для общих количественных результатов данное отличие не играет никакой роли.

что речевой материал названного респондента с максимальным количеством словоформ представлен приблизительно 2000 высказываний.) Следующий график показывает количество респондентов и объем их материала:

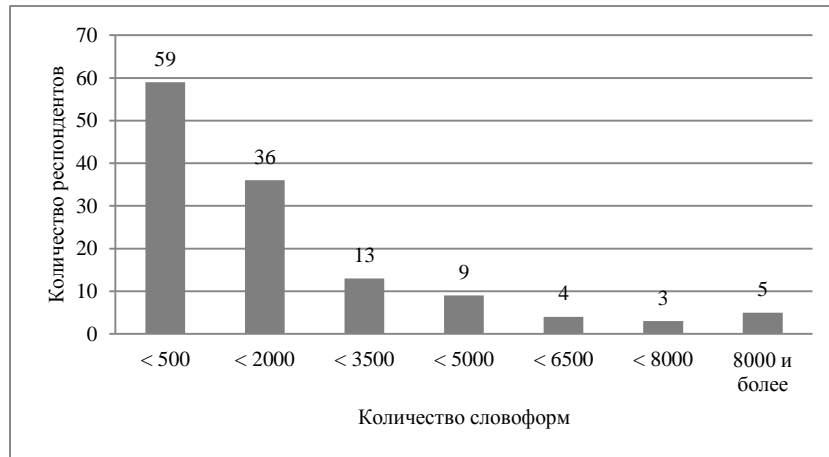


Рис. 1 Количество респондентов и объем их материала (т.е. количество словоформ) в семейном корпусе

Учитывая представленные различия, было правомерным разделить 129 респондентов, принявших участие в записях, на две группы. 59 человек было отнесено к группе «периферийных респондентов», объем материала которых не превышает 500 словоформ. Остальные же 70 человек относятся к группе «основных респондентов». Их материал составляет приблизительно 97 процентов от общего объема всего материала в корпусе (205938 из общей сложности 212279 словоформ).

Естественно, при анализе языкового материала значительную роль играет социальная характеристика респондентов. С целью проверить важность отдельных социальных параметров респондентам было предложено заполнить анкету, содержащую соответствующие вопросы. Как правило, анкета была предложена для заполнения только 70 «основным» респондентам, так как социальные параметры «периферийных» респондентов количественно все равно не повлияли бы на полученные результаты. В процессе работы, к сожалению, были утеряны данные о социальной характеристике трех основных

респондентов из поселка Октябрьский, одного из семи населенных пунктов, в которых проводились записи. Если не принимать во внимание этих трех говорящих, то объем материала остальных 67 основных респондентов составляет 203110 словоформ, т.е. почти 96 процентов от общего объема корпуса.

### 2.1.2 В корпусе интервью

Данные, находящиеся в основе корпуса интервью, были собраны в тех же населенных пунктах (за исключением двух городов, см. ниже), что и данные семейного корпуса, причем их основной целью было проведение качественного социологического исследования (см. пример в KITTEL & LINDNER 2011). Выбраны были такие респонденты, которые в закрытых интервью, находящихся в основе исследования HENTSCHEL & KITTEL (2011), ответили, что используют БРСР в качестве основного средства коммуникации. Фрагменты из этих интервью, содержащие частые высказывания на БРСР,<sup>6</sup> были таким же образом затранскрибированы и аннотированы (т.е. указаны грамматические показатели, а также информация о языковой афинности), как и языковой материал семейного корпуса. Как было упомянуто раньше, выбранные для интервью респонденты не имели никакого отношения к семьям из семейного корпуса, а также не были знакомы между собой.

Конечно же, в корпусе интервью следует различать высказывания респондентов и высказывания интервьюеров. Кроме того, в некоторых разговорах есть высказывания людей из окружения респондентов, присутствующих при интервью. Из общего объема корпуса интервью в 167523 словоформы 141570 словоформ, т.е. приблизительно 85 процентов, принадлежат респондентам.

Как и следовало ожидать, разница в объеме языкового материала отдельных респондентов в корпусе интервью значительно меньше, чем в семейном корпусе. О общей сложности для анализа были выбраны речевые фрагменты из получасовых либо часовых интервью, проведенных с 54 респондентами. Как было отмечено ранее, имеются в виду отрывки, содержащие интенсивное смешение обоих языков в речи, тогда как отрывки на русском языке (как правило, с более или менее

---

6 В самом начале интервью, длившихся один час, респонденты употребляли БРСР значительно меньше, чем дальше на протяжении разговора. В начале же интервью они обращались к русскому языку, не к белорусскому.

явными фонетическими и фонологическими интерференциями со стороны белорусского языка, т.е. на такой его разновидности, которая является немаркированным кодом общения с чужими в Беларуси) не принимались во внимание. На следующем графике представлено соотношение между количеством респондентов и объемом их материала:

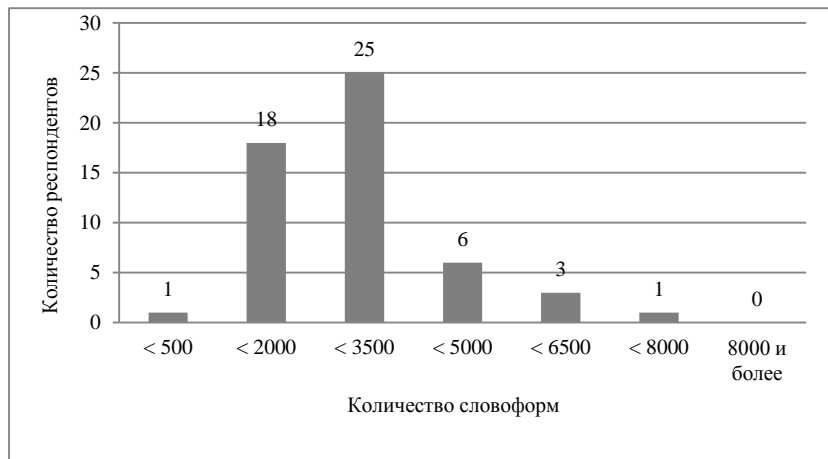


Рис. 2 Количество респондентов и объем их материала (т.е. количество словоформ) в корпусе интервью

Также была собрана информация о социальной характеристике респондентов из корпуса интервью, о чем еще будет сказано дальше.

## 2.2 Города и респонденты

### 2.2.1 В семейном корпусе

Материал семейного корпуса был собран в шести небольших<sup>7</sup> городах, которые расположены на территории трех основных белорусских диалектов (северо-восточного, юго-западного и центрального переходного), а также в столичном городе Минске (расположенном географически на территории центрального диалектного региона, но,

<sup>7</sup> С населением от приблизительно 7000 человек в Шарковщине до 168000 в Барановичах.

как известно, представляющем собой особый случай с диалектологической точки зрения).<sup>8</sup> Из каждого диалектного региона был рассмотрен один скорее западный и один скорее восточный город:

Табл. 1 Города, представленные в семейном корпусе, и их население (по состоянию на 2009 г.)

Северо-вост. диалект		Центр. диалект			Юго-зап. диалект	
зап.	вост.	зап.	центр.	вост.	зап.	вост.
Шарковщина (Sh)	Хотимск (Ch)	Сморгонь (Sm)	Минск (Mi)	Рогачев (Ra)	Барановичи (Ba)	Октябрьский (Ak)
6.900	7.100	36.300	1.836.800	33.700	168.200	7.400
Витебская обл.	Могилевская обл.	Гродненская обл.	Минская обл. (город)	Гомельская обл.	Брестская обл.	Гомельская обл.

Даже не принимая во внимание сильно выделяющуюся, как метрополия и миллионный город, столицу Минск, заметны различия также и в структуре остальных рассматриваемых городов (ср. HENTSCHEL & KITTEL 2011). Если Сморгонь и Рогачев являются городами средней величины, а Барановичи представляют собой в белорусском понимании даже относительно большой город, то остальные три города, Шарковщина, Хотимск и Октябрьский – маленькие города, напоминающие поселения сельского типа.

Изначально было запланировано собрать из каждого из семи городов, т.е. каждого семейного контекста, языковой материал в объеме 25000 словоформ. По возможности должны были быть представлены респонденты двух поколений, причем женщины и мужчины в таком количественном соотношении, которое позволило бы провести определенные сравнения. Однако при выбранном способе записи (преимущество которого – как было отмечено в начале – состояло в возможности записать в значительной мере непринужденную, свободную речь и, таким образом, преодолеть парадокс наблюдателя) проконтролировать подобного рода количественные соотношения не было возможности. В четырех городах (Хотимск, Минск, Рогачев и

<sup>8</sup> Многие респонденты из Минска являются уроженцами населенных пунктов, относящихся к северо-восточному диалектному региону Беларуси.

Сморгонь) пришлось проводить дополнительный сбор материала, чтобы хотя бы уменьшить значительные отклонения от названной цели. В итоге в этих городах был собран дополнительный языковой материал в объеме от приблизительно 8000 до 10000 словоформ:

Табл. 2 Количество словоформ из отдельных городов в семейном корпусе

<b>Город</b>	<b>Количество словоформ</b>	<b>В процентах</b>	<b>Сумма процентов</b>
<b>Ak</b>	24946	11,8	11,8
<b>Ba</b>	25386	12,0	23,7
<b>Ch</b>	34212	16,1	39,8
<b>Mi</b>	33987	16,0	55,8
<b>Ra</b>	33023	15,6	71,4
<b>Sh</b>	25263	11,9	83,3
<b>Sm</b>	35462	16,7	100,0
<b>Всего</b>	<b>212279</b>	<b>100,0</b>	

Объем материала (т.е. количество словоформ) из разных городов колеблется в пределах 11,8 и 16,7 процента и является, таким образом, сравнительно сбалансированным.

Значительно сильнее различается количество говорящих в разных городах. Оно представлено в следующей таблице, в которой, к тому же, говорящие поделены на три группы: периферийные респонденты с объемом материала менее 500 словоформ и две группы основных говорящих: с объемом материала от 500 до 4999 словоформ и более 5000 словоформ.

Табл. 3 Количество респондентов в разных городах и объем их материала (семейный корпус)

Город	менее 500 словоформ	от 500 до 4999	5000 и более	
<b>Ak</b>	14	14	0	28
<b>Ba</b>	6	4	2	12
<b>Ch</b>	2	8	2	12
<b>Mi</b>	7	9	2	18
<b>Ra</b>	0	9	1	10
<b>Sh</b>	4	5	2	11
<b>Sm</b>	26	9	3	38
<b>Всего</b>	<b>59</b>	<b>58</b>	<b>12</b>	<b>129</b>

Так как количество словоформ может сильно различаться даже у говорящих, относящихся к одной группе, имеет смысл сравнить доли языкового материала трех рассматриваемых здесь групп говорящих. Такое сравнение предлагает следующий график:

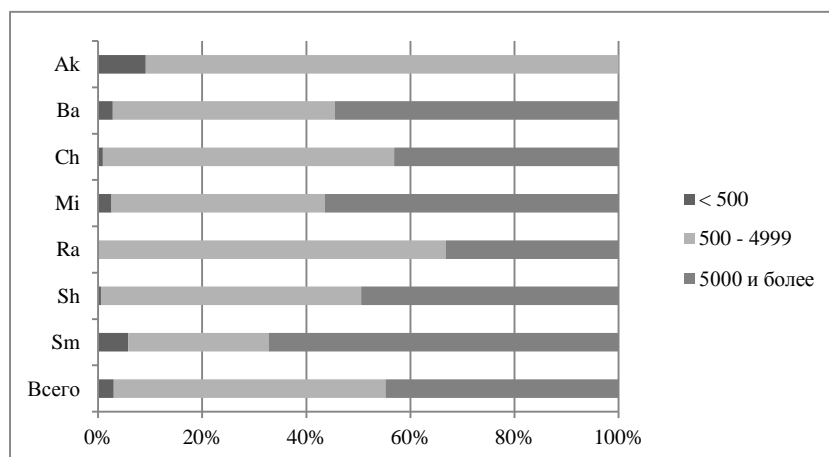


Рис. 3 Доля материала трех групп говорящих, различающихся количеством словоформ, в разных городах (семейный корпус)

Характерно, что как в общем корпусе, так и в отдельных его семи частях (т.е. из семи городов) периферийным говорящим, составляющим большинство почти во всех городах, в каждом из рассматриваемых городов принадлежит всего лишь небольшая доля материала – ни в одном из населенных пунктов она не превышает 10 процентов. Что же касается «количественно доминантных» респондентов, то здесь разные части корпуса показывают значительные различия. К количественно доминантным относятся такие говорящие, объем материала которых составляет 5000 и более словоформ. Так, если в Октябрьском нет ни одного количественно доминантного говорящего, то в других городах такие респонденты есть, причем в некоторых населенных пунктах их даже двое или трое. Доли материала этих говорящих варьируют между двумя третями в Сморгони (три доминантных говорящих в этом городе) и приблизительно одной третью в Рогачеве (только один доминантный говорящий). Объем материала доминантных респондентов в остальных городах, в каждом из которых их по два, составляет соответственно от одной трети до двух третей. Чем больше доля материала доминантных говорящих, тем индивидуальнее является этот материал.

#### *2.2.2 В корпусе интервью*

В корпусе интервью представлены те же города, что и в семейном корпусе, за исключением двух городов: во-первых, город Барановичи пришлось заменить на соседний с ним Слоним, город с населением немного менее 50000 жителей. Во-вторых, не были рассмотрены данные интервью из Минска, так как в этом городе не удалось получить достаточный для анализа объем материала на БРСР. Хотя респонденты из Минска и ответили при предшествующем интервью опросе, что регулярно используют БРСР, на вопросы интервью они отвечали преимущественно по-русски.



Табл. 4 Города, представленные в корпусе интервью, и их население (по состоянию на 2009)

Северо-вост.		Центр.		Юго-западн.	
зап.	вост.	зап.	вост.	зап.	вост.
Шарковщина (Sh)	Хотимск (Ch)	Сморгонь (Sm)	Рогачев (Ra)	Слоним (Sl)	Октябрьский (Ak)
6900	7100	36300	33700	49000	7400
Витебская обл.	Могилев- ская обл.	Гроднен- ская обл.	Гомельская обл.	Гроднен- ская обл.	Гомельская обл.

В каждом из названных городов был собран объем языкового материала, сравнимый с объемом в семейном корпусе, т.е. от 25000 до 30000 словоформ.

Табл. 5 Количество словоформ из отдельных городов в корпусе интервью

Город	Количество словоформ	В процентах	Сумма процентов
<b>Ak</b>	29073	17,4	17,4
<b>Sl</b>	29159	17,4	34,8
<b>Ch</b>	29708	17,7	52,5
<b>Ra</b>	25035	14,9	67,4
<b>Sh</b>	24778	14,8	82,2
<b>Sm</b>	29770	17,8	100,0
<b>Всего</b>	167523	100,0	

Доля языкового материала из разных городов составляет приблизительно от 15 до 18 процентов и является, таким образом, относительно сбалансированной.

Также в корпусе интервью, в отличие от семейного корпуса, не очень сильно различается количество говорящих в разных городах, ср. следующую таблицу:

Табл. 6 Количество респондентов в разных городах (корпус интервью)

Город	Количество респондентов	В процентах	Сумма процентов
<b>Ак</b>	7	13,0	13,0
<b>Сл</b>	13	24,1	37,1
<b>Сн</b>	11	20,4	57,4
<b>Ра</b>	9	16,7	74,1
<b>Ш</b>	7	13,0	87,1
<b>См</b>	7	13,0	100,0
<b>Всего</b>	54	100,0	

Естественно, следует обратить внимание на то, какую долю от общего объема материала корпуса интервью составляет материал респондентов и какую материал интервьюеров (а также других говорящих, принявших спорадическое участие в некоторых разговорах). Такое разделение материала представлено в следующей таблице:

Табл. 7 Доля материала респондентов, интервьюеров и других говорящих в корпусе интервью (в количестве словоформ)

Город		Респонденты	Интервьюеры	Другие	Всего
<b>Ак</b>	<i>N</i>	21776	7297	0	29073
	%	74,9%	25,1%	0,0%	100,0%
<b>Сн</b>	<i>N</i>	26801	2907	0	29708
	%	90,2%	9,8%	0,0%	100,0%
<b>Ра</b>	<i>N</i>	21526	3509	0	25035
	%	86,0%	14,0%	0,0%	100,0%
<b>Ш</b>	<i>N</i>	20403	3289	1086	24778
	%	82,3%	13,3%	4,4%	100,0%
<b>Сл</b>	<i>N</i>	25122	3739	298	29159
	%	86,2%	12,8%	1,0%	100,0%
<b>См</b>	<i>N</i>	25942	3828	0	29770
	%	87,1%	12,9%	0,0%	100,0%
<b>Всего</b>	<i>N</i>	141570	24569	1384	167523
	%	84,5%	14,7%	0,8%	100,0%

В общей сложности объем материала респондентов составляет 141570 словоформ, т.е. приблизительно 85 процентов от общего объема корпуса интервью. Только в Октябрьском этот показатель не достигает 80 процентов. Все представленные далее данные будут относиться только к языковому материалу респондентов.

## **2.3    Возрастная структура респондентов и доля материала разных возрастных групп**

### *2.3.1    В семейном корпусе*

70 «основных» респондентов (из общего количества 129 респондентов) из семи городов могут быть разделены на следующие возрастные группы (отсутствуют сведения о возрасте трех респондентов из Октябрьского<sup>9</sup>):

---

9    Другие сведения о социальном фоне этих респондентов также отсутствуют.

Табл. 8 Основные респонденты семейного корпуса из разных городов и их год рождения

Город		до 1961	1961-1980	после 1980	отсутствует	Всего
<b>Ак</b>	<i>N</i>	1	6	4	3	14
	%	7,1%	42,9%	28,6%	21,4%	100,0%
<b>Ba</b>	<i>N</i>	4	2	0	0	6
	%	66,7%	33,3%	0,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	2	6	2	0	10
	%	20,0%	60,0%	20,0%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	5	3	3	0	11
	%	45,5%	27,3%	27,3%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	5	2	3	0	10
	%	50,0%	20,0%	30,0%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	2	1	4	0	7
	%	28,6%	14,3%	57,1%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	2	7	3	0	12
	%	16,7%	58,3%	25,0%	0,0%	100,0%
<b>Всего</b>	<i>N</i>	21	27	19	3	70
	%	30,0%	38,6%	27,1%	4,3%	100,0%

Очевидно, что возрастная структура в разных городах, т.е. в разных семейных контекстах, очень различается. Тем не менее в целом респонденты распределены по возрастным группам относительно равномерно.

Однако более важной является информация о доле материала разных возрастных групп, нежели о количестве респондентов в этих группах. Она будет представлена в следующей таблице, причем доля материала снова будет выражена в количестве словоформ из разных городов. Из общего количества 212297 словоформ 205938 словоформы (или 97 процентов) приходится на 70 «основных говорящих» семейного корпуса, которые в разных городах могут быть разделены на следующие возрастные группы:

Табл. 9 Доля анализируемого языкового материала респондентов разных возрастных групп (в количестве словоформ) из разных городов в семейном корпусе (только основные говорящие)

Город		до 1961	1961-1980	после 1980	отсутствует	Всего
<b>Ак</b>	<i>N</i>	1432	14162	4245	2828	22667
	%	6,3%	62,5%	18,7%	12,5%	100,0%
<b>Ba</b>	<i>N</i>	17455	7212	0	0	24667
	%	70,8%	29,2%	0,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	4293	24763	4854	0	33910
	%	12,7%	73,0%	14,3%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	18427	10310	4383	0	33120
	%	55,6%	31,1%	13,2%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	22553	2461	8009	0	33023
	%	68,3%	7,5%	24,3%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	6332	5257	13534	0	25123
	%	25,2%	20,9%	53,9%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	1415	29810	2203	0	33428
	%	4,2%	89,2%	6,6%	0,0%	100,0%
<b>Всего</b>	<i>N</i>	71907	93975	37228	2828	205938
	%	34,9%	45,6%	18,1%	1,4%	100,0%

Как видно из таблицы, в разных частях корпуса (т.е. частях корпуса из разных городов) не одинаково представлены разные возрастные группы говорящих. В то время как в Барановичах, Минске и Рогачеве большая часть материала приходится на представителей старшей возрастной группы, в Октябрьском, Хотимске и Сморгони основную роль играет средняя группа (особенно сильно в Сморгони). И, наконец, в Шарковщине преобладает материал младшей возрастной группы.

### 2.3.2 В корпусе интервью

54 респондента из шести городов могут быть разделены на следующие возрастные группы:

Табл. 10 Респонденты корпуса интервью из разных городов и их год рождения

Город		до 1961	1961-1980	после 1980	Всего
<b>Ak</b>	<i>N</i>	6	1	0	7
	%	85,7%	14,3%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	5	8	0	13
	%	38,5%	61,5%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	7	4	0	11
	%	63,6%	36,4%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	3	5	1	9
	%	33,3%	55,6%	11,1%	100,0%
<b>Sl</b>	<i>N</i>	5	2	0	7
	%	71,4%	28,6%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	6	1	0	7
	%	85,7%	14,3%	0,0%	100,0%
<b>Всего</b>	<i>N</i>	32	21	1	54
	%	59,3%	38,9%	1,9%	100,0%

Самую большую группу составляют старшие респонденты, к младшей же группе может быть отнесен только один единственный респондент. Тем не менее наблюдаются различия между городами: в Сморгони и Октябрьском значительно преобладают старшие респонденты; в Рогачеве и Слониме их преобладание выражено менее ярко; в Хотимске и Шарковщине большинство составляют респонденты средней возрастной группы.

Что же касается объема языкового материала (выраженного в количестве словоформ, произнесенных респондентами), то среди представителей разных возрастных групп он распределен следующим образом:

Табл. 11 Доля языкового материала респондентов разных возрастных групп (в количестве словоформ) из разных городов в корпусе интервью

Город		До 1961	1961-1980	После 1980	Всего
<b>Аk</b>	<i>N</i>	19320	2456	0	21776
	%	88,7%	11,3%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	11701	15100	0	26801
	%	43,7%	56,3%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	12460	9066	0	21526
	%	57,9%	42,1%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	6496	11871	2036	20403
	%	31,8%	58,2%	10,0%	100,0%
<b>Sl</b>	<i>N</i>	15382	9740	0	25122
	%	61,2%	38,8%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	23538	2404	0	25942
	%	90,7%	9,3%	0,0%	100,0%
<b>Всего</b>	<i>N</i>	88897	50637	2036	141570
	%	62,8%	35,8%	1,4%	100,0%

Респонденты средней возрастной группы уступают старшей группе также и по количеству произнесенных ими словоформ. Младшее же поколение и вовсе представлено только маргинально.

## 2.4 Соотношение полов респондентов

### 2.4.1 В семейном корпусе

Среди 70 «основных» респондентов семейного корпуса количество женщин приблизительно в два раза превышает количество мужчин:

Табл. 12 Пол респондентов из разных городов в семейном корпусе

Город		Женск.	Мужск.	Отсутствует	Всего
<b>Ак</b>	<i>N</i>	8	3	3	14
	%	57,1%	21,4%	21,4%	100,0%
<b>Ва</b>	<i>N</i>	4	2	0	6
	%	66,7%	33,3%	0,0%	100,0%
<b>Сh</b>	<i>N</i>	7	3	0	10
	%	70,0%	30,0%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	7	4	0	11
	%	63,6%	36,4%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	7	3	0	10
	%	70,0%	30,0%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	5	2	0	7
	%	71,4%	28,6%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	7	5	0	12
	%	58,3%	41,7%	0,0%	100,0%
<b>Всего</b>	<i>N</i>	45	22	3	70
	%	64,3%	31,4%	4,3%	100,0%

Преобладание респондентов женского пола становится еще более явным, если рассматривать количество произнесенных словоформ:

Табл. 13 Доля анализируемого языкового материала респондентов женского и мужского пола (в количестве словоформ) из разных городов в семейном корпусе (только основные говорящие)

Город		Женск.	Мужск.	Отсутствует	Всего
<b>Ак</b>	<i>N</i>	13847	5992	2828	22667
	%	61,1%	26,4%	12,5%	100,0%
<b>Ва</b>	<i>N</i>	18576	6091	0	24667
	%	75,3%	24,7%	0,0%	100,0%



<b>Ch</b>	<i>N</i>	21559	12351	0	33910
	%	63,6%	36,4%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	29451	3669	0	33120
	%	88,9%	11,1%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	26781	6242	0	33023
	%	81,1%	18,9%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	17056	8067	0	25123
	%	67,9%	32,1%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	20366	13062	0	33428
	%	60,9%	39,1%	0,0%	100,0%
<b>Всего</b>	<i>N</i>	147636	55474	2828	205938
	%	71,7%	26,9%	1,4%	100,0%

Преобладание женского пола наблюдается во всех городах, при этом в одних городах оно более сильное, в других немного слабее; наиболее ярко оно выражено в Рогачеве.

#### 2.4.2 В корпусе интервью

Также и среди 54 респондентов корпуса интервью количество женщин примерно в два раза превышает количество мужчин:

Табл. 14 Пол респондентов из разных городов в корпусе интервью

Город		Женск.	Мужск.	Всего
<b>Ak</b>	<i>N</i>	7	0	7
	%	100,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	10	3	13
	%	76,9%	23,1%	100,0%
<b>Ra</b>	<i>N</i>	5	6	11
	%	45,5%	54,5%	100,0%
<b>Sh</b>	<i>N</i>	5	4	9
	%	55,6%	44,4%	100,0%

<b>Sl</b>	<i>N</i>	6	1	7
	%	85,7%	14,3%	100,0%
<b>Sm</b>	<i>N</i>	4	3	7
	%	57,1%	42,9%	100,0%
<b>Всего</b>	<i>N</i>	37	17	54
	%	68,5%	31,5%	100,0%

Что же касается количества произнесенных респондентами словоформ, то здесь преобладание женского пола становится еще более явным:

Табл. 15 Доля анализируемого языкового материала респондентов женского и мужского пола (в количестве словоформ) из разных городов в корпусе интервью

<b>Город</b>		<b>Женск.</b>	<b>Мужск.</b>	<b>Всего</b>
<b>Ак</b>	<i>N</i>	21776	0	21776
	%	100,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	20257	6544	26801
	%	75,6%	24,4%	100,0%
<b>Ra</b>	<i>N</i>	10729	10797	21526
	%	49,8%	50,2%	100,0%
<b>Sh</b>	<i>N</i>	13842	6561	20403
	%	67,8%	32,2%	100,0%
<b>Sl</b>	<i>N</i>	23350	1772	25122
	%	92,9%	7,1%	100,0%
<b>Sm</b>	<i>N</i>	18111	7831	25942
	%	69,8%	30,2%	100,0%
<b>Всего</b>	<i>N</i>	108065	33505	141570
	%	76,3%	23,7%	100,0%

## 2.5 Образование респондентов<sup>10</sup>

### 2.5.1 В семейном корпусе

БРСР, или «трасянку», принято считать показателем плохого образования, т.е. характеристикой людей, которые по причине недостаточной образованности не в состоянии говорить ни на «чистом белорусском», ни на «чистом русском» языке. Лисковец (ЛИСКОВЕЦ 2005) даже считает «необразованность» дефиниционным критерием носителей трасянки. Хентшель (ХЕНТШЕЛЬ 2013) отвергает такой подход, обосновывая это в первую очередь тем, что в случае эмпирической ориентации только на подобную группу говорящих<sup>11</sup> реальный масштаб распространения БРСР *a priori* останется неизвестным. В семейном корпусе ОК-БРСР представлены говорящие с абсолютно разными уровнями образования, о чем свидетельствуют данные в следующей таблице.

Табл. 16 Уровень образования респондентов из разных городов в семейном корпусе

Город	Еще незаконч. школьн.	Средн. / школьн.	Средн.- спец.	Выш.	Незак. школьн. ное	Инф. отсутс твует	Всего
<b>Ак</b>	N 0	2	4	4	0	4	14
	% 0,0%	14,3%	28,6%	28,6%	0,0%	28,6%	100%
<b>Ва</b>	N 0	2	3	1	0	0	6
	% 0,0%	33,3%	50,0%	16,7%	0,0%	0,0%	100%
<b>Ch</b>	N 0	7	3	0	0	0	10
	% 0,0%	70,0%	30,0%	0,0%	0,0%	0,0%	100%

10 Отсутствует информация об образовании шести респондентов, принадлежащих к группе «основных говорящих»: к их числу относится еще один респондент из Октябрьского, а также два респондента из Шарковщины. Объем материала оставшихся 64 говорящих составляет в общей сложности 193181 словоформы, т.е. почти 91 процент от общего объема.

11 В вопросе о том, есть ли и в самом деле люди, способные говорить только на очень смешанной белорусско-русской речи, или же при этом они все-таки владеют «в приемлемой форме» русским (пусть и с белорусскими интерференциями в произношении и, в меньшей степени, в лексике, особенно типичными для белорусских реалий) или белорусским языком (пусть и с некоторыми заимствованиями из русского языка, особенно в лексике), мнения белорусистов, точнее языковедов в Беларуси, сильно расходятся.

<b>Mi</b>	<i>N</i>	0	3	6	1	1	0	11
	%	0,0%	27,3%	54,5%	9,1%	9,1%	0,0%	100%
<b>Ra</b>	<i>N</i>	0	0	2	8	0	0	10
	%	0,0%	0,0%	20,0%	80,0%	0,0%	0,0%	100%
<b>Sh</b>	<i>N</i>	3	1	1	0	0	2	7
	%	42,9%	14,3%	14,3%	0,0%	0,0%	28,6%	100%
<b>Sm</b>	<i>N</i>	2	2	7	1	0	0	12
	%	16,7%	16,7%	58,3%	8,3%	0,0%	0,0%	100%
<b>Bcero</b>	<i>N</i>	5	17	26	15	1	6	70
	%	7,1%	24,3%	37,1%	21,4%	1,4%	8,6%	100%

Большинство основных говорящих имеет либо обычное среднее (школьное) образование, либо средне-специальное. Однако присутствуют также и говорящие с более высоким уровнем образования, особенно в Рогачеве. Только у одного говорящего нет никакого законченного образования, а также пять человек еще учатся в школе. Кроме того, данные в таблице отчетливо показывают неоднородность семейных контекстов в разных городах относительно уровня образования. Тем не менее речь в собранных семейных разговорах была оценена С. Запрудским и другими сотрудниками проекта из Беларуси как смешанная, т.е. БРСР. Результаты проведенного в рамках представленного выше исследования определенно подтверждает данную оценку.<sup>12</sup>

Также и в этом отношении возникает вопрос о доле языкового материала разных групп говорящих, выделенных на основе уровня образования:

12 Библиография представлена здесь: <https://www.uni-oldenburg.de/slavistik/forschung/-sprachwissenschaft/schwerpunkt-mischvarieteten/oldenburger-publikationen-zur-weissrussisch-russischen-gemischten-rede/>.

Табл. 17 Доля языкового материала респондентов с разным уровнем образования (в количестве словоформ) из разных городов в семейном корпусе

Город	Еще незаконч. школьное	Средн. / школьн.	Средн. -спец.	Высш.	Незак. школьн.ое	Инф. от-сущств.	Всего
<b>Ak</b>	<i>n</i> 0	2491	9591	6645	0	3940	22667
	% 0,0%	11,0%	42,3%	29,3%	0,0%	17,4%	100%
<b>Ba</b>	<i>n</i> 0	8382	10863	5422	0	0	24667
	% 0,0%	34,0%	44,0%	22,0%	0,0%	0,0%	100%
<b>Ch</b>	<i>n</i> 0	27727	6183	0	0	0	33910
	% 0,0%	81,8%	18,2%	0,0%	0,0%	0,0%	100%
<b>Mi</b>	<i>n</i> 0	16742	12591	2329	1458	0	33120
	% 0,0%	50,5%	38,0%	7,0%	4,4%	0,0%	100%
<b>Ra</b>	<i>n</i> 0	0	5184	27839	0	0	33023
	% 0,0%	0,0%	15,7%	84,3%	0,0%	0,0%	100%
<b>Sh</b>	<i>n</i> 6295	7239	2169	0	0	9420	25123
	% 25,1%	28,8%	8,6%	0,0%	0,0%	37,5%	100%
<b>Sm</b>	<i>n</i> 1480	10477	20612	859	0	0	33428
	% 4,4%	31,3%	61,7%	2,6%	0,0%	0,0%	100%
<b>Всего</b>	<i>n</i> 7775	73058	67193	43094	1458	13360	205938
	% 3,8%	35,5%	32,6%	20,9%	,7%	6,5%	100%

Также и с этой точки зрения основные группы в корпусе составляют горячие с обычным школьным образованием и горячие со средне-специальным образованием. Только в Рогачеве значительно преобладает материал респондентов с высшим образованием. А в Шарковщине к тому же хорошо представлена младшая группа горячих, которые еще учатся в школе.

### 2.5.2 В корпусе интервью

Также и респонденты корпуса интервью ОК-БРСР сильно различаются по уровню образования, что представлено в следующей таблице.

Табл. 18 Уровень образования респондентов из разных городов в корпусе интервью

Город		Незак. школьное	Средн. / школьн.	Средн.- спец.	Выш.	Всего
<b>Ак</b>	<i>n</i>	0	3	3	1	7
	%	0,0%	42,9%	42,9%	14,3%	100,0%
<b>Ch</b>	<i>n</i>	0	1	10	2	13
	%	0,0%	7,7%	76,9%	15,4%	100,0%
<b>Ra</b>	<i>n</i>	2	2	6	1	11
	%	18,2%	18,2%	54,5%	9,1%	100,0%
<b>Sh</b>	<i>n</i>	0	1	5	3	9
	%	0,0%	11,1%	55,6%	33,3%	100,0%
<b>Sl</b>	<i>n</i>	1	0	5	1	7
	%	14,3%	0,0%	71,4%	14,3%	100,0%
<b>Sm</b>	<i>n</i>	1	4	2	0	7
	%	14,3%	57,1%	28,6%	0,0%	100,0%
<b>Всего</b>	<i>n</i>	4	11	31	8	54
	%	7,4%	20,4%	57,4%	14,8%	100,0%

Как видно из таблицы, четверо говорящих не окончили школу, а восемь говорящих имеют высшее образование. Самую же большую группу составляют говорящие со средне-специальным образованием. Следующая таблица показывает, что распределение говорящих по группам в зависимости от уровня образования и распределение языкового материала разных групп говорящих очень похожи.

Табл. 19 Доля языкового материала респондентов с разным уровнем образования (в количестве словоформ) из разных городов в корпусе интервью

Город		Незак. школьное	Средн. / школьн.	Средн.- спец.	Высш.	Всего
<b>Ак</b>	<i>n</i>	0	7950	12224	1602	21776
	%	0,0%	36,5%	56,1%	7,4%	100,0%
<b>Ch</b>	<i>n</i>	0	1565	20257	4979	26801
	%	0,0%	5,8%	75,6%	18,6%	100,0%
<b>Ra</b>	<i>n</i>	1016	3516	13613	3381	21526
	%	4,7%	16,3%	63,2%	15,7%	100,0%
<b>Sh</b>	<i>n</i>	0	2036	12602	5765	20403
	%	0,0%	10,0%	61,8%	28,3%	100,0%
<b>Sl</b>	<i>n</i>	1527	0	17866	5729	25122
	%	6,1%	0,0%	71,1%	22,8%	100,0%
<b>Sm</b>	<i>n</i>	7939	13395	4608	0	25942
	%	30,6%	51,6%	17,8%	0,0%	100,0%
<b>Всего</b>	<i>n</i>	10482	28462	81170	21456	141570
	%	7,4%	20,1%	57,3%	15,2%	100,0%

Однако наблюдаются значительные различия в распределении материала между городами.

## 2.6 О миграции из деревни в город

### 2.6.1 В семейном корпусе

Принято считать, что БРСР, или «трасянка», является характерной чертой деревенских жителей, переехавших в города. Особенный же интерес вызывают, естественно, дети таких мигрантов из деревни в город. Следующая таблица показывает разделение респондентов на группы по критерию «внутренней миграции».

Табл. 20 Мигранты из деревни в город и горожане среди респондентов семейного корпуса из разных городов

		Поздние мигранты и деревенские жители	Типичные мигранты из деревни в город	Горожане (рожденные в городах)	Информация отсутствует	Всего
<b>Ак</b>	<i>n</i>	1	6	3	4	14
	%	7,1%	42,9%	21,4%	28,6%	100,0%
<b>Вa</b>	<i>n</i>	0	4	2	0	6
	%	0,0%	66,7%	33,3%	0,0%	100,0%
<b>Ch</b>	<i>n</i>	1	4	5	0	10
	%	10,0%	40,0%	50,0%	0,0%	100,0%
<b>Mi</b>	<i>n</i>	1	7	3	0	11
	%	9,1%	63,6%	27,3%	0,0%	100,0%
<b>Ra</b>	<i>n</i>	2	4	4	0	10
	%	20,0%	40,0%	40,0%	0,0%	100,0%
<b>Sh</b>	<i>n</i>	0	0	0	7	7
	%	0,0%	0,0%	0,0%	100,0%	100,0%
<b>Sm</b>	<i>n</i>	0	8	4	0	12
	%	0,0%	66,7%	33,3%	0,0%	100,0%
<b>Всего</b>	<i>n</i>	5	33	21	11	70
	%	7,1%	47,1%	30,0%	15,7%	100,0%

Самую большую часть всех говорящих (почти половину) составляют «типичные мигранты из деревни в город», которые провели свою молодость преимущественно в сельской местности и переехали в города на учебу или работу, где и остались как минимум на 10, а то и намного больше лет. Ко второй по численности группе относятся говорящие, которые родились и выросли уже в городах; чаще всего это дети представителей первой группы. Третью группу составляют деревенские жители, которые во время проведения записей находились в городе в гостях, а также респонденты, переехавшие в город уже в пожилом возрасте. Как правило, речь здесь идет о бабушках и дедушках в рассматриваемых семьях. В следующей таблице представлены доли материала каждой из этих групп:



Табл. 21 Доля языкового материала мигрантов из деревни в город и горожан (в количестве словоформ) из разных городов в семейном корпусе

		Поздние мигранты и деревенские жители	Типичные мигранты из деревни в город	Горожане (рожденные в городах)	Информация отсутствует	Всего
<b>Ak</b>	<i>n</i>	1432	13421	3186	4628	22667
	%	6,3%	59,2%	14,1%	20,4%	100,0%
<b>Ba</b>	<i>n</i>	0	17455	7212	0	24667
	%	0,0%	70,8%	29,2%	0,0%	100,0%
<b>Ch</b>	<i>n</i>	1543	18366	14001	0	33910
	%	4,6%	54,2%	41,3%	0,0%	100,0%
<b>Mi</b>	<i>n</i>	1458	27279	4383	0	33120
	%	4,4%	82,4%	13,2%	0,0%	100,0%
<b>Ra</b>	<i>n</i>	3880	20075	9068	0	33023
	%	11,7%	60,8%	27,5%	0,0%	100,0%
<b>Sh</b>	<i>n</i>	0	0	0	25123	25123
	%	0,0%	0,0%	0,0%	100,0%	100,0%
<b>Sm</b>	<i>n</i>	0	21471	11957	0	33428
	%	0,0%	64,2%	35,8%	0,0%	100,0%
<b>Всего</b>	<i>n</i>	8313	118067	49807	29751	205938
	%	4,0%	57,3%	24,2%	14,4%	100,0%

Если рассматривать долю языкового материала (в количестве словоформ) каждой из названных групп, то значимость группы типичных мигрантов из деревни в город возрастает еще на 10 процентов, а доля материала двух других групп, соответственно, уменьшается.

### 2.6.2 В корпусе интервью

В следующей таблице представлено распределение респондентов по группам в зависимости от критерия внутренней миграции.

Табл. 22 Мигранты из деревни в город и горожане среди респондентов корпуса интервью из разных городов

		Поздние мигранты и деревенские жители	Типичные мигранты из деревни в город	Горожане (рожденные в городах)	Информация отсутствует	Всего
<b>Ak</b>	<i>n</i>	0	5	2	0	7
	%	0,0%	71,4%	28,6%	0,0%	100,0%
<b>Ch</b>	<i>n</i>	0	8	4	1	13
	%	0,0%	61,5%	30,8%	7,7%	100,0%
<b>Ra</b>	<i>n</i>	0	6	1	4	11
	%	0,0%	54,5%	9,1%	36,4%	100,0%
<b>Sh</b>	<i>n</i>	0	8	1	0	9
	%	0,0%	88,9%	11,1%	0,0%	100,0%
<b>Sl</b>	<i>n</i>	0	3	3	1	7
	%	0,0%	42,9%	42,9%	14,3%	100,0%
<b>Sm</b>	<i>n</i>	1	4	2	0	7
	%	14,3%	57,1%	28,6%	0,0%	100,0%
<b>Всего</b>	<i>n</i>	1	34	13	6	54
	%	1,9%	63,0%	24,1%	11,1%	100,0%

Приблизительно две трети всех респондентов являются мигрантами из деревни в город. Только одного говорящего можно отнести к группе поздних мигрантов и деревенских жителей. Каждый же четвертый респондент родился и вырос в городе.

Следующая таблица показывает доли языкового материала этих групп:

Табл. 23 Доля языкового материала мигрантов из деревни в город и горожан (в количестве словоформ) из разных городов в корпусе интервью

		Поздние мигранты и деревенские жители	Типичные мигранты из деревни в город	Горожане (рожденные в городах)	Информация отсутствует	Всего
<b>Ак</b>	<i>n</i>	0	14923	6853	0	21776
	%	0,0%	68,5%	31,5%	0,0%	100,0%
<b>Сh</b>	<i>n</i>	0	17278	7958	1565	26801
	%	0,0%	64,5%	29,7%	5,8%	100,0%
<b>Ра</b>	<i>n</i>	0	12113	858	8555	21526
	%	0,0%	56,3%	4,0%	39,7%	100,0%
<b>Sh</b>		0	18367	2036	0	20403
	%	0,0%	90,0%	10,0%	0,0%	100,0%
<b>Sl</b>		0	10709	11267	3146	25122
	%	0,0%	42,6%	44,8%	12,5%	100,0%
<b>Sm</b>	<i>n</i>	7939	13395	4608	0	25942
	%	30,6%	51,6%	17,8%	0,0%	100,0%
<b>Всего</b>	<i>n</i>	7939	86785	33580	13266	141570
	%	5,6%	61,3%	23,7%	9,4%	100,0%

И здесь распределение языкового материала разных групп напоминает распределение говорящих по этим группам. Только представителям группы поздних мигрантов принадлежит объем материала больше среднего.

## 2.7 Вывод

Подобные расхождения, которых невозможно избежать при проведении записей с, как правило, ограниченными практическими, в том числе финансовыми, возможностями и при этом необходимостью получить обширный материал, конечно же, ограничивают «прямую» сопоставимость частей корпуса с помощью методов простой описательной статистики. Однако такую «ассиметрию» можно преодолеть, используя определенные статистические методы.

### 3 Структура транскрипции в системе СНАТ

ОК-БРСР был затранскрибирован и аннотирован согласно правилам СНАТ, разработанным в рамках проекта CHILDES (по состоянию на 2013 г.; ср. McWHINNEY 2013a), что дает возможность анализировать корпус с помощью программы CLAN (ср. McWHINNEY 2013b, см. главу 4). Речь идет о текстовых документах с кодировкой UTF8, информация в которых разбита на отдельные строки (так называемые «tiers»).

Далее внимание будет обращено в первую очередь на некоторые расхождения с общепринятой системой СНАТ, а также на принципы транскрипции, определения «языковой афинности» словоформ (в общих чертах: можно ли определенную словоформу отнести к белорусскому или русскому языку или она является „общей“ либо „гибридной“), а также грамматической индексации. Для ознакомления с другими правилами системы СНАТ пользователю рекомендуется обратиться к руководству по ней (McWHINNEY 2013a).<sup>13</sup>

#### 3.1 Названия файлов

Выбор заголовков для транскриптов происходил по следующей схеме: сокращенное (трехзначное) обозначение части корпуса (fam для *семейного корпуса*, int для *корпуса интервью*), двузначное сокращение названия города, в котором была сделана определенная запись, и номер разговора внутри каждого города, отделенный знаком нижнего подчеркивания. В следующей таблице представлены сокращенные названия городов:

Сокращение	Город
ak	Октябрьский
ba	Барановичи (только в семейном корпусе)
ch	Хотимск

---

13 Следующее описание методов работы с ОК-БРСР не исключает возможности появления единичных случаев расхождения с принятыми принципами. По причине большого объема материала, который частично обрабатывался студентами (хотя и под руководством опытных сотрудников), невозможно полностью избежать ошибок и противоречий. Возможность проверки транскрипции более 400000 словоформ в обеих частях корпуса с помощью акустического анализа совершенно исключается.

sl	Слоним (только в корпусе интервью)
mi	Минск (только в семейном корпусе)
ra	Рогачев
sm	Сморгонь
sh	Шарковщина

### 3.2 Основные черты структуры транскрипции

Следующий пример наглядно иллюстрирует структуру транскрипции:

*Пример:*

```
@UTF8
@Begin
@Languages:   hybe
@Participants: RFB Village
@ID:   hybe|fam|RFB|72;|female|central||Village|higher||
@L1 of RFB:   be
@MT of RFB:   be
@FL of RFB:   brms
@Situation:   Conversation at home
*RFB:   pryvet@r@pb , dzetkâ@c@pb .
%mor:   r:pb#n:inanim|_0_privet_&NOM-M:SG
        c:pb#n:anim|_dzetka_detka_&NOM-GC:SG .
*RFB:   nu@c@pc , bjažy@b@pb , a@c@pc to@c@pc (.) ja@c@pc s@c@pc
        xolâdu@b@pc .
%mor:   c:pc#ptl|_nu_nu_ b:pb#v:ipf|_behčy_0_&IMPER-2:SG
        c:pc#conj|_a_a_ c:pc#ptl|_to_to_
        c:pc#pro:pers:dei|_ja_ja_&NOM-1:SG c:pc#prep|_z_s_
        b:pb#n:inanim|_cholad_cholod_&GEN-M:SG .
*RFB:   xalodnâ@c@pc .
%mor:   c:pc#adv|_choladna_cholodno_ .
*RFB:   zâstužu@r@pc cjabe@b@pb .
%mor:   r:pc#v:pf|_0_zastudit'_&FUT-1:SG
        b:pb#pro:pers:dei|_ty_ty_&ACC-2:SG .
@End
```

Каждый транскрипт начинается с указания кодировки **@UTF8**, после которого следует строка **@Begin**, обозначающая начало транскрипта.

Далее, от строки **@Languages:** до строки **@Situation:**, представлена основная информация о разговоре и о принимающих в нем участие респондентах. Затем следует непосредственная транскрипция разговора, содержащая также дополнительную информацию о словоформах. Каждый транскрипт заканчивается строкой **@End**.

### 3.3 Основная информация о разговорах

В самом начале транскрипта приводится основная информация о разговоре и об участвующих в нем респондентах. Какая именно информация имеется в виду, будет пояснено ниже.

#### 3.3.1 *@Languages: Типы говорящих, принимающие участие в разговорах*

Как правило, в системе CHAT строка **@Languages:** содержит информацию об используемых в разговоре языках. В корпусе же ОК-БРСР здесь указываются *типы говорящих*, под которыми мы понимаем следующее: HENTSCHEL & ZELLER (2013) показывают, на основании проведенного ими кластерного анализа, что основные говорящие семейного корпуса – представленные в корпусе достаточно большим количеством высказываний – в зависимости от доли „белорусских“, „русских“ и „гибридных“ высказываний в их речи (абстрагируясь от чисто фонетико-фонологических характеристик) могут быть разделены на четыре группы.<sup>14</sup> Таким образом, тип говорящего является константной характеристикой респондентов (это значит, что определенный говорящий во всех разговорах относится к одному и тому же типу) и не дает никаких сведений о языковом поведении говорящего в том или ином конкретном разговоре.

Респонденты корпуса интервью – это «признавшиеся» носители смешанной речи, т.е. в опросе, проведенном в Беларуси среди 1400 респондентов, который предшествовал интервью, они ответили, что БРСР является их регулярным, в большинстве случаев основным средством коммуникации. Кроме того, для корпуса интервью – в отличие от семейного корпуса – целенаправленно были выбраны только такие фрагменты из интервью, речь в которых можно охарактеризовать

---

14 Как показывают HENTSCHEL & ZELLER (2014), наблюдаются различия также и в фонетико-фонологических особенностях речи данных типов говорящих.

как БРСР. Поэтому провести основанную на статистических данных дедуктивную классификацию респондентов интервью, проведенную в случае с говорящими семейного корпуса, невозможно. Поэтому все респонденты корпуса интервью получают сокращение **brms**.

<i>Корпус</i>	<i>Сокращ.</i>	<i>Расшифровка</i>	<i>Количество говорящих в ОК-БРСР</i>
Семейный корпус	hybe	Высокая частотность как «гибридных», так и «белорусских» высказываний, низкая частотность «русских» высказываний (тип говорящих ГБ в классификации HENTSCHEL & ZELLER (2013))	10
Семейный корпус	hyb	Однозначное преобладание «гибридных» высказываний над «русскими» и «белорусскими» (тип говорящих Г в классификации HENTSCHEL & ZELLER (2013))	23
Семейный корпус	hyru	Высокая частотность прежде всего «гибридных», а также «русских» высказываний, низкая частотность «белорусских» высказываний (тип говорящих ГР в классификации HENTSCHEL & ZELLER (2013))	25
Семейный корпус	rus	Однозначное преобладание «русских» высказываний над «гибридными» и – еще сильнее – над «белорусским» (тип говорящих Р в классификации HENTSCHEL & ZELLER (2013))	12
Семейный корпус	nan	«Информация отсутствует»: говорящие с небольшим количеством высказываний	59
Корпус интервью	brms	„Belarusian-Russian mixed speech“ (белорусско-русская смешанная речь): все респонденты корпуса интервью	54
Корпус интервью	nan	«Информация отсутствует»: интервьюеры и говорящие, представленные только небольшими отрывками, т.е. не «целевые» говорящие	не различаются

*Пример:*

@Languages: hyb, hyru

Таким образом, в данном разговоре участвуют говорящие, в речи которых преобладают «гибридные» высказывания, а также говорящие, для которых характерно большое количество как «гибридных», так и «русских» высказываний.

### 3.3.2 @Participants: Роль или миграционный статус говорящих

Как правило, в строке @Participants: в системе СНАТ содержится информация о «роли» участвующих в разговоре респондентов. А в ОК-БРСР здесь подана информация о миграционном статусе тех говорящих, для которых такая информация является релевантной (см. пункт 3.3.3.5).

На основе критерия миграции можно выделить три группы прототипных носителей смешанной речи:

- К первой группе относятся *мигранты из деревни в город*, т.е. люди, в 1960-х – 1970-х гг., возможно, также в 1980-х, в молодом возрасте переехавшие на работу из деревень, в которых еще сохранялись белорусские диалекты, в быстро развивающиеся города. Их школьное обучение проходило еще преимущественно на белорусском языке.
- Вторую группу составляют *горожане* (т.е. рожденные в городе). Как правило, это дети представителей первой группы. Они были рождены уже в городах и провели в них большую часть своего детства. Их первичная языковая социализация происходила на смешанной речи их родителей, т.е. на смешанном монолекте, без дифференциации языков. Однако по сравнению со своими родителями представители данного поколения намного раньше, еще в школе или даже в дошкольных учреждениях, прошли интенсивную социализацию на русском языке. Поэтому они, как правило, в состоянии говорить на литературном русском языке без глубинных интерференций со стороны белорусского языка. Контакт же с диалектным белорусским языком у представителей данного поколения значительно слабее.
- В первую очередь в семейном корпусе встречаются говорящие, которых можно отнести к третьей группе, а именно группе *поздних мигрантов и деревенских жителей*. Это, как правило, поколение родителей мигрантов. Представители этого поколения провели всю свою жизнь в сельской местности или переехали к своим детям в города уже в пожилом возрасте.



В центре внимания проекта находились две первые группы. Третью группу составляют говорящие, случайно принявшие участие в разговорах; объем их материала небольшой (ср. 2.6).

Таким образом, в ОК-БРСР в строке **@Participants**: содержится информация о том, к какой из перечисленных групп принадлежат респонденты как семейного корпуса, так и корпуса интервью, если, конечно, такая информация известна. Интервьюеры получают роль **Interviewer**, другие говорящие корпуса интервью – роль **Other**.

Сокращение	Значение
Village	Поздние мигранты и деревенские жители
Migrant	Мигранты из деревни в город
Town	Горожане (рожденные в городах)
Na	Информация отсутствует
Interviewer	Интервьюеры в корпусе интервью
Other	Случайно присутствующие при интервью

*Пример:*

@Participants: BFA Migrant, BFB Town

Говорящий BFA является представителем группы мигрантов из деревни в город, а говорящий BFB – группы рожденных в городе.

### 3.3.3 @ID: Информация о говорящих

В строке **@ID**: содержится следующая информация о каждом из говорящих:

@ID: тип говорящих|Корпус|Сокращенное обозначение говорящего/  
 респондента|Возраст|Пол|Представленность в  
 корпусе||Роль/Миграционный статус|Образование||

Показатели *возраста* (количество полных лет; по состоянию на 2008 г.) и *пола* (male/female) не требуют дальнейшего объяснения, другие же параметры будут вкратце оговорены ниже.

### 3.3.3.1 Тип говорящих

В позиции, где в системе СНАТ обычно представлена информация о языке, используемом определенным говорящим в разговоре, в ОК-БРСР будет представлена информация о языковой ориентации говорящего, т.е. *тип говорящих* (см. пункт 3.3.1).

### 3.3.3.2 Корпус

Как было отмечено вначале, ОК-БРСР состоит из двух частей корпуса: из семейного корпуса и корпуса интервью. Таким образом, в данной позиции будет отмечено, к какому из двух корпусов принадлежит определенный разговор.

<i>Сокращение</i>	<i>Значение</i>
fam	Семейный корпус
int	Корпус интервью

### 3.3.3.3 Сокращенное обозначение говорящего / респондента

Каждый говорящий семейного корпуса и каждый респондент корпуса интервью имеет свое постоянное, т.е. не зависящее от отдельных разговоров, сокращенное обозначение, содержащее три или четыре знака. Первая буква данной характеристики обозначает город.

<i>Сокращение</i>	<i>Город</i>
A	Октябрьский
B	Барановичи (только в семейном корпусе)
C	Хотимск
L	Слоним (только в корпусе интервью)
M	Минск (только в семейном корпусе)
R	Рогачев
S	Сморгонь
H	Шарковщина

Второй знак в семейном корпусе всегда **F**, а в корпусе интервью – **I**. Далее следует сокращенное обозначение говорящего внутри каждого

города, как правило, однозначное (в семейном корпусе представленное одной буквой либо цифрой, в корпусе интервью всегда цифрой), в некоторых случаях также двузначное (если в каком-то из городов говорящих больше, чем используемых букв или цифр).

*Примеры:*

С112: Говорящий 12 из Хотимска в корпусе интервью

НFN: Говорящий N из Шарковщины в семейном корпусе

Все интервьюеры в корпусе интервью получают обозначение **INT**. Говорящие, которые случайно присутствовали при интервью и представлены в корпусе очень незначительным количеством высказываний, обозначены сокращением **OTH** („Other“). Если же таких говорящих в одном разговоре несколько, то они пронумерованы (**OT1**, **OT2** и т.д.).

#### 3.3.3.4 Представленность в корпусе

Относительно каждого говорящего в семейном корпусе указано, принадлежит ли он к группе «основных» респондентов. Как было сказано вначале, к «основным» респондентам относятся 70 говорящих, представленных в корпусе 500 и более словоформами (что соответствует примерно 100 высказываниям); такие говорящие получают обозначение **central**, а все остальные говорящие – обозначение **peripher**.

В корпусе интервью все респонденты обозначены как **Informant**, а интервьюеры, а также случайные участники разговора, не имеют специального обозначения; информация же о том, что речь идет об интервьюере, приводится позже (ср. 3.3.3.5).

#### 3.3.3.5 Миграционный статус / роль в интервью

В данной позиции в системе СНАТ обычно приводятся сведения о «роли» участвующих в разговоре информантов. В ОК-БРСР же здесь содержится информация о миграционном статусе (**Village**, **Migrant** или **Town**) тех респондентов, для которых такая информация релевантна (говорящие семейного корпуса и респонденты корпуса интервью), интервьюеры получают обозначение **Interviewer**, другие говорящие корпуса интервью – обозначение **Other** (см. главу 3.3.2).

### 3.3.3.6 Образование

В зависимости от уровня образования выделяются следующие группы:

<i>Сокращение</i>	<i>Значение</i>
in education	Школьники
incomplete/primary	8-9 или менее классов
secondary	10-11 классов
professional	Училище, колледж, техникум
higher	Высшее образование
Na	Информация отсутствует

### 3.3.4 @L1: Первый язык

При наличии соответствующей информации в строке **@L1 of #**: указан язык, который определенный говорящий назвал своим первым языком (при ответе на вопрос: «На каком языке Вы начали разговаривать?»). Возможны следующие показатели:

<i>Сокращение</i>	<i>Значение</i>
be	белорусский
ru	русский
brms	белорусско-русская смешанная речь

В единичных случаях возможны также другие варианты, например, **p1** (*польский*). Т.к. у респондентов была возможность выбрать сразу несколько вариантов, то встречаются также и комбинации из этих языков.

### 3.3.5 @FL: Язык семьи в детстве

При условии наличия соответствующей информации в строке **@FL of #**: указываются языки, названные определенным говорящим в качестве языков семейного общения в детские годы (FL: „Family language“).

### 3.3.6 @MT: Родной язык

При наличии соответствующей информации в строке **@MT of #:** указываются языки, называемые говорящими в качестве родного языка (MT: „Mother tongue“).

### 3.3.7 @Comment: Комментарии к транскрипции

Строка **@Comment:** содержит общие комментарии к транскрипции. Так, например, в некоторых редких случаях в разговорах не отмечены паузы либо наложения речи одного говорящего на речь другого, что в таком случае отмечается здесь.

### 3.3.8 @Situation: Речевая ситуация

Строка **@Situation:** содержит информацию о речевой ситуации. Типичными ситуациями являются **Conversation at home** либо **x talks on the phone** и т.д.

## 3.4 Воспроизведение разговоров

Строки, начинающиеся со звездочки (\*), содержат высказывания одного говорящего в хронологической последовательности. За звездочкой следует сокращенное обозначение говорящего (ср. 3.3.3.3). Далее, в строке **%mor:**, содержится аннотация словоформ данного высказывания (см. главу 3.5).

*Пример:*

```
*BFA: jana@b@pb ne@c@pc máladaĵê@c@pc užé@r@pc takajê@c@pc
pârikmaxêršý@r@pr .
%mor: b:pb#pro:pers:ana|_jana_0_&NOM-F:SG c:pc#ptl|_ne_ne_
c:pc#adj|_malady_molodoj_&NOM-F:SG r:pc#adv|_0_uže_
c:pc#pro:dem|_taki_takoj_&NOM-F:SG
r:pr#n:anim|_0_parikmacherša_&NOM-F:SG .
*BFA: ne@c@pc dzevâčkâ@r@pb .
%mor: c:pc#ptl|_ne_ne_ r:pb#n:anim|_0_devočka_&NOM-F:SG .
```

Паузы и прерывания записи, а также незатранскрибированные фрагменты (например, по причине защиты данных) отмечены при помощи новой строки **@New Episode:**. Опущенные фрагменты могут быть очень короткими; иногда они могут представлять собой только одно высказывание.

*Пример:*

```
*AFD: on@r@pc taková@r@pc slová@c@pc ni@c@pr znajêt@r@pc .
%mor: r:pc#pro:pers:ana|_0_on_&NOM-M:SG
      r:pc#pro:dem|_taki_takoj_&GEN-N:SG
      c:pc#n:inanim|_slova_slovo_&GEN-N:SG c:pr#ptl|_ne_ne_
      r:pc#v:ipf|_znanac'_znanat'_&PRAES-3:SG .
@New Episode
*AFJ: nu@c@pc tak@c@pc nâbiri@c@pr jemu@c@pc .
%mor: c:pc#ptl|_nu_nu_ c:pc#ptl|_tak_tak_
      c:pr#v:pf|_nabrac'_nabrat'_&IMPER-2:SG
      c:pc#pro:pers:ana|_ën_on_&DAT-M:SG .
```

### 3.4.1 Деление речи на высказывания

Как правило, деление потока речи на высказывания соответствует синтаксическому делению.

- В сомнительных случаях при принятии решения о том, является ли определенный отрывок речи одним высказыванием, или же его следует разделить на два высказывания, предпочтение отдается синтаксическому членению: как правило, высказывание соответствует простому предложению (причем, как и следовало ожидать, очень часто встречаются эллиптические структуры). Высказывания из всего лишь одной части предложения (с или без предиката, как эллиптические, так и нет) составляют почти две трети всех высказываний в семейном корпусе; приблизительно одна четверть состоит из сложных предложений (как сложносочиненных, так и сложноподчиненных). Очень часто встречаются также безпредикатные высказывания или их части, состоящие из всего лишь одной (или нескольких) словоформы, как, например, утвердительная и отрицательная частицы и / или дискурсивные

маркеры (например, *Да, Нет, Конечно, Ага*) и т.д. Подобные конструкции составляют приблизительно 8% семейного корпуса.

- Только при наличии определенной интонации (долгая пауза и / или свидетельствующая о завершенности интонация главного предложения) синтаксическое придаточное предложение рассматривается как самостоятельное высказывание.
- Только при наличии определенной интонации (отсутствие паузы и / или однозначно свидетельствующая о незавершенности интонация первой части предложения) два синтаксических предложения объединяются в одно высказывание.

### 3.4.2 Транскрипция словоформ

Используемая в ОК-БРСР транскрипция представляет собой компромисс между общепринятой транслитерацией (белорусского и / или русского кириллического алфавита на латинский) и транскрипцией (т.е. преобразованием звуков в письменный текст). При этом обращается внимание на соответствующие, наиболее важные различия в фонетике и фонологии белорусского и русского языков.<sup>15</sup>

#### 3.4.2.1 Гласные в ударной позиции

Для обозначения гласных в ударных слогах используются следующие знаки: ( **a** ), ( **e** ), ( **i** ), ( **o** ), ( **ě** ), ( **u** ), ( **y** ) и ( **é** ). Что следует принимать во внимание:

- /i/ ([i]) и /e/ ([ɛ]) после твердых согласных передаются знаками ( **y** ) и ( **é** ).

---

15 Речь идет о фонетической классификации, проведенной на слух, большей частью белорусскими студентами, принимавшими участие в проекте. Ввиду фонетической градуальности некоторых релевантных явлений, например, в случае безударного вокализма, присутствует, естественно, определенная ненадежность. Однако, как показывает ZELLER (2015) на основе акустического анализа материала ОК-БРСР, в общей сложности транскрипции являются достоверными.

<i>Транскрипция:</i>	
mašyna	бел. <i>машына</i> ; рус. <i>машина</i>
šélest	рус. <i>шелест</i>

- ( **і** ) и ( **е** ), а также ( **ё** ) указывают также на мягкость предыдущего согласного, а ( **у** ) и ( **э** ) – на твердость. В случае других гласных на мягкость предыдущего согласного указывает либо ( **ј** ), либо – в случае предыдущих ( **ч** ) или ( **щ** ) – знак ( **´** ). Таким же образом передается мягкость / твердость согласных, являющихся в русском языке мягкими, а в белорусском твердыми (см. далее).

<i>Транскрипция словоформ с твердым согласным</i>	<i>Транскрипция словоформ с мягким согласным</i>	
radâm	rjadâm	рус. <i>рядом</i>
trox	trëx	бел. <i>трох</i> ; рус. <i>трех</i>
učorâ	učërâ	бел. <i>учора</i>
častâ	č´astâ	бел. <i>часта</i> ; рус. <i>часто</i>
xaču	xač´u	бел. <i>хачу</i> ; рус. <i>хочу</i>

- Ударная /o/ после мягких согласных передается через ( **ё** ) (ASCII 235). А /jo/ транскрибируется как ( **јо** ):

<i>Транскрипция</i>	
vsë	рус. <i>все</i>
trëx	бел. <i>трох</i> ; рус. <i>трех</i>
jon	бел. <i>ён</i>
vtrajom	рус. <i>втроем</i>
pap´´jom	бел. <i>пап'ём</i> ; рус. <i>попьем</i>

#### 3.4.2.2 Гласные в первом предударном слоге

В данной позиции могут стоять знаки ( **а** ), ( **е** ), ( **і** ), ( **у** ), ( **у** ) и ( **э** ). ( **і** ) и ( **е** ) указывают на мягкость предыдущего согласного, ( **у** ) и ( **э** ) – на твердость. В случае других гласных на мягкость предыдущего согласного указывает либо ( **ј** ), либо – в случае предыдущих ( **ч** ) или ( **щ** ) – знак ( **´** ).



- после мягких согласных могут встречаться знаки ( **i** ), ( **e** ), а также – после ( **j** ) – ( **a** ) и ( **u** ). Т.е. делается различие между [i], [e], [a] и [u]. В случае /a/, /e/, /o/ в первых предударных слогах соответственно проводится различие между тождественными с «русским» иканьем [i]-подобными реализациями (в определенных случаях также тождественными с «белорусским диалектным» диссимилятивным яканьем), тождественными с «белорусским» яканьем [a]-подобными реализациями и «промежуточными» [e]-подобными реализациями.<sup>16</sup>

Транскрипция (в зависимости от фонетической реализации)	
ciɣer / seɣer / cɣaper	бел. цяпер; рус. теперь
nima / nema / njama	бел. німа/няма

- после твердых (не затвердевших) согласных могут, как правило, встречаться знаки ( **a** ), ( **u** ) и ( **y** ) (исключения указаны ниже). Аканье, т.е. нейтрализация оппозиции /a/ и /o/ и их совпадение в [a]-подобном звуке, отражается в транскрипции. Однако не различаются более редуцированные реализации типа [ɐ], [ʌ] или [ə], характерные для русского языка, и более открытые [a]-подобные реализации, типичные для белорусского языка.

Транскрипция	
vada	бел. в[а]да; рус. в[ɐ]да

- По аналогии с позицией после мягких согласных в случае /a/, /e/, /o/, после затвердевших согласных проводится различие между [i, ɛ, a] при помощи знаков ( **y** ), ( **é** ) и ( **a** ). То же касается гласных после твердого [tʃ], а также после твердого [r], когда ему в русском языке соответствует мягкий /rʲ/.

16 Анализ фонетико-фонологической вариативности на основе транскрипций ОК-БРСР, проведенных на слух, можно найти в HENTSCHEL & ZELLER (2014). А акустический анализ данного явления представлен в ХЕНШЕЛЬ & ЦЕЛЛЕР (2011), а также в ZELLER (2015).

<i>Транскрипция (в зависимости от фонетической реализации)</i>	
žylanêj / žélanêj / žalanêj	рус. <i>желаний</i>
ryka / réka / raka	бел. <i>рака</i> ; рус. <i>река</i>
učyra / učéra / učara	рус. <i>вчера</i>

### 3.4.2.3 Гласные в других предударных и послеударных слогах

В начале слова, а также при сочетании гласных могут стоять следующие знаки: ( **â** ), ( **i** ), ( **u** ), ( **y** ), ( **é** ).

<i>Транскрипция:</i>	
inânda	рус. <i>иногда</i>
nêinteresnâ	рус. <i>неинтересно</i>
ykânamist / ékanamist	бел. <i>эканаміст</i> ; рус. <i>экономист</i>
ânanas	бел. <i>ананас</i> ; рус. <i>ананас</i>

В других случаях (т.е. не в начале слова и не при сочетании гласных) могут встречаться следующие знаки: ( **ê** ), ( **u** ), ( **y** ), ( **ÿ** ), ( **â** ).

- После мягких согласных стоят знаки ( **ê** ) и – после ( **j** ) – ( **u** ). Т.е. более [i]-, [e]- или [a]-подобные реализации на месте /a/, /e/, /o/ в других безударных слогах, кроме первых предударных, не различаются. Таким же образом не проводится различие между /i/, с одной стороны, и /a/, /e/, /o/ с другой стороны. Везде стоит знак ( **ê** ). Даже не принимая во внимание явлений смешивания, реализация гласных в данных позициях вариативна как в белорусском, так и в русском языках, а восприятие на слух сильно ограничено.

<i>Транскрипция:</i>	
étê / écê	рус. <i>эти</i>
mamê	бел. <i>маме</i> ; рус. <i>маме</i>
vamê	бел. <i>вамi</i> ; рус. <i>вами</i>
cjanê	бел. <i>цягне</i> ; рус. <i>тянет</i>
xodêt	бел. <i>ходзіць</i> ; рус. <i>ходит</i>
sinêm	бел. <i>сiнiм</i> ; рус. <i>синем</i>

- После твердых (не затвердевших) согласных могут, как правило, встречаться знаки ( **â** ), ( **у** ) и ( **u** ). Знак ( **у** ) стоит на месте /i/; а знак ( **â** ) – /a/ и /o/.

<i>Транскрипция:</i>	
mâlako	бел. <i>малако</i> ; рус. <i>молоко</i>
horâdâ	бел. <i>горада</i> ; рус. <i>города</i>
krasnyjê	рус. <i>красные</i>

- После затвердевших согласных, по аналогии с позицией после мягких согласных в случае /a/, /e/, /o/, не проводится различие между фонетически разными реализациями. Везде стоит знак ( **ý** ). Это же касается твердого [ʃ] и твердого [r], когда ему в русском языке соответствует мягкий /r'/. После твердого [ʃ] и твердого [r] также и /i/ передается знаком ( **ý** ).

<i>Транскрипция:</i>	
bol'šý	бел. <i>большы</i>
čatyrý	бел. <i>чатыры</i> ; рус. <i>четыре</i>
čýlavec	бел. <i>чалавек</i> ; рус. <i>человек</i>
хо́чý	бел. <i>хоча</i>
vyhrýlâ	бел. <i>выйграла</i> ; рус. <i>выиграла</i>
zakoněylâ	бел. <i>закончыла</i> ; рус. <i>закончила</i>

#### 3.4.2.4 Особые случаи транскрипции гласных

- Случаи так называемого «неполного аканья» будут передаваться при помощи знака ( **ô** ) или, после мягких согласных, при помощи знака ( **ǰô** ) (*Примеры: maslô, pãčakajcǰô*).
- В заимствованных словах знак ( **é** ) может встречаться также и в безударных слогах после твердых согласных (*Примеры: dékanat, kamp' jutér*).
- При дополнительном ударении качество гласных различается при помощи следующих знаков: ( **î** ), ( **у** ), ( **е** ), ( **é** ), ( **а** ), ( **о** ) и ( **ë** ) (*Примеры: kilahram, diagânal', fotââparat, šestičýsâvoj*).

### 3.4.2.5 Транскрипция согласных

Транскрипция согласных осуществляется в соответствии с правилами научной транслитерации, за исключением следующих пунктов:

- [x] передается знаком ( **ж** ).
- Наличие /j/ в начале слов либо слогов отражается в транскрипции. Также отражается и протетический [j] перед /i/ (*Примеры: māladaǰê, jêdinêcâ, list´jêv, jix*).
- Неслоговой [ɟ] передается знаком ( **џ** ).

### 3.4.2.6 Палатализация

Палатализация согласных передается в транскрипции разными способами:

- Следующие знаки гласных указывают на мягкость предыдущих согласных: ( **і** ), ( **е** ), ( **ё** ), ( **ѐ** ).
- Перед остальными гласными на мягкость согласного указывает следующий за ним знак ( **ј** ). Исключениями в данном случае являются ( **щџ** ) и ( **џ** ), на мягкость которых указывает следующий за ними знак апострофа ( **'** ) (*Примеры: lǰudzê, šč´aslivâǰê*).
- Перед согласными, а также в конце слова на мягкость согласного указывает следующий за ним апостроф ( **'** ) (ASCII 146) (*Примеры: tol´kê, byvajuc´*).

### 3.4.2.7 Различия согласных в белорусском и русском языках

Следующие фонетические и фонологические различия между белорусским языком и русским отражены в транскрипции:<sup>17</sup>

- рус. [d´], [t´] vs. бел. [dʒ´], [tʃ´] (так называемые белорусские дзеканье и цеканье):

---

<sup>17</sup> Анализ фонетико-фонологической вариативности на основе транскрипций ОК-БРСР, проведенных на слух, можно найти в HENTSCHEL & ZELLER (2014). А акустический анализ данного явления представлен в HENTSCHEL & ZELLER (2011), а также в ZELLER (2015).

Транскрипция ,белорусских‘ аффрикат:	Транскрипция ,русских‘ взрывных звуков:	
dzecê	detê	бел. дзеці; рус. дети

- рус. взрывной [g] vs. бел. фрикативный [ɣ] или [ɦ] (причем белорусские фрикативные реализации не различаются. Также не передаются отвердение конечных согласных и ассимиляция по глухости / звонкости):

Транскрипция ,белорусского‘ фрикативного звука:	Транскрипция ,русского‘ взрывного звука:	
hlaz	glaz	рус. глаз
moh	mog	бел. мог; рус. мог

- рус. [l] vs. бел. [ɫ] в глагольных формах прошедшего времени мужского рода, а также в некоторых других лексемах типа рус. волк – бел. воўк:

Транскрипция ,белорусского‘ [ɫ]	Транскрипция ,русского‘ [l]	
daŭ	dal	бел. даў; рус. дал
voŭk	volk	бел. воўк; рус. волк

- рус. /v/ vs. бел. [u] или [ɯ], как, например, в рус. лавка vs. бел. лаўка; рус. все vs. бел. ўсё/усё. Реализации же как [v] или как [f] не различаются:

Транскрипция ,белорусского‘ гласного или полугласного	Транскрипция ,русского‘ фрикативного звука	
uremê, úremê	vremê	рус. время
laŭkâ	lavka	бел. лаўка; рус. лавка
usê, úsê	vsê	бел. ўсё/усё; рус. все

- рус. [u] vs. бел. [ɯ] на месте начального /u/, если предыдущее слово оканчивалось гласной:

Транскрипция ,белорусского‘ неслогового [ɯ]	Транскрипция ,русского‘ слогового [u]	
úžé	užé	рус. уже

- рус. мягкие [ʃʲ] и [ʃʲ:] vs. бел. [ʃ] и [ʃʲ], а также рус. мягкая /гʲ/ vs. бел. твердая /г/. Мягкость согласного в конце слова и перед другими согласными передается знаком апострофа ( ´ ) (ASCII 146), а отсутствие данного знака отражает, соответственно, твердость согласного. Также на мягкость согласных указывают следующие за ними знаки гласных ( *i* ), ( *e* ), ( *ë* ), ( *ê* ) или – в случае /гʲ/ – знак ( *j* ), либо – в случае [ʃʲ] и [ʃʲ:] – знак апострофа ( ´ ). А на твердость согласных указывают знаки ( *a* ), ( *é* ), ( *у* ), ( *o* ), ( *u* ), ( *ŷ* ):

Транскрипция ,белорусского´ твердого согласного	Транскрипция ,русского´ мягкого согласного	
uĉorâ	učërâ	бел. <i>учора</i>
ĉastâ	ĉ´astâ	рус. <i>часто</i> ; бел. <i>часта</i>
pamoĉ	pamoĉ´	рус. <i>помочь</i>
radâm	rjadâm	рус. <i>рядом</i>
trox	trëx	бел. <i>трох</i> ; рус. <i>трех</i>
ciper	ciper´	бел. <i>цяпер</i> ; рус. <i>теперь</i>

#### 3.4.2.8 Передача результатов некоторых фонологических процессов и другие явления

- Суффиксы *-чск-* и *-ц-* различаются (*Примеры: baranâvêĉskêj vs. baranâvêĉkê*).
- В большинстве случаев, однако не всегда, транскрипторы передали ассимиляцию сибилантов. Часто определить наличие ассимиляции на слух очень сложно. Таким образом, в данном отношении транскрипции не всегда можно доверять, например, в случае возвратных глаголов второго лица единственного числа (*Примеры: naješšê vs. najessê* от рус. *наешься*) или в случае существительных женского рода с основой на заднеязычный и уменьшительно-ласкательным суффиксом *-ч-* в предложном или дательном падеже (*Примеры: réĉcŷ vs. réccŷ* от бел. *рэчцы*).
- Ассимиляция по мягкости типа [sʲ] *ліва* или [zʲ] *бег* в транскрипции не отражена (*Примеры: slivâ, zbeh*).

- Ассимиляция по глухости / звонкости и отвердение конечных согласных, как, например, в *но[ʃ]ка* или *коро[f]* в транскрипции не передается (*Примеры: nožkâ, karov*).
- Комбинация из инфинитивного окончания и возвратного постфикса (на письме: рус. *-ться*, бел. *-цца*) передается как ( **сцý** ) (*Пример: dâgâvariccý*).
- Звук [ʦ] всегда передается знаком ( **с** ), независимо от того, возможно ли его дальнейшее фонологическое (или даже морфологическое) членение (*Примеры: saldackê* от рус. *солдатский* и бел. *салдацкі*; *dvaccýc´* от рус. *двадцать* и бел. *дваццаць*).

#### 3.4.2.9 Особые случаи

- Наблюдаются некоторые различия в транскрипции аббревиатур, выполненной разными транскрипторами. Так, рядом с примерами типа *ésémés, béémvé* встречаются также аббревиатуры типа *bgu*.
- Клитки, которые в русском/белорусском языке пишутся через дефис, так же передаются и в ОК-БРСР (*Пример: dze-tâ*).
- Также и другие словоформы с дефисом передаются с дефисом в транскрипции ОК-БРСР (*Примеры: sinê-sinê, kojê-xto*).
- Сложные междометия также передаются в транскрипции с дефисом (*Примеры: aj-aj-aj*).
- Сложные слова, пишущиеся в русском/белорусском языке отдельно, соединяются в транскрипции, как правило, при помощи нижнего подчеркивания (*Примеры: sjuda\_tuda*).
- Неразборчивый материал отмечен знаком ( **жж** ).

*Пример:*

\*RFK: za@c@pc жж@u jim@b@pc zâpîsalê@c@pc , vylêčýúsê@b@pb  
toj@b@pc čêlavêk@c@pr .

- Личные имена, если это не имена известных общественных деятелей, по причине защиты данных не транскрибируются, на их месте стоит символ-заполнитель (**persname**).

*Пример:*

\*S3B: +" vухadzi@r@pb за@с@рс persname persname !

- Вульгаризмы не транскрибируются, вместо них стоит символ-заполнитель (**swearword**).

*Пример:*

\*RFL: to@с@рс sâvišćanêjê@r@ph , to@с@рс pêdsavet@с@рс ,  
to@с@рс išćé@b@ph jakajê@b@pb swearword .

### 3.4.3 Языковая афинность

При помощи так называемых тегов («tags») в корпусе отмечена языковая афинность каждой отдельной словоформы. Следует особо подчеркнуть, что вопрос о том, что в смешанной речи, появившейся в результате контакта двух родственных и структурно очень похожих языков, можно классифицировать как принадлежащее одному или другому языку, а что как общее или гибридное, является в меньшей степени эмпирическим вопросом, а скорее теоретическим. Другими словами, здесь возможно несколько способов описания. Классификация словоформ в корпусе основана на строгом алгоритме, подробно представленном в HENTSCHEL (2008). Определение языковой афинности словоформ происходит раздельно на двух уровнях: на фонетико-фонологическом уровне и на структурно более глубоком уровне, в котором рассматриваются лексические, морфонологические и морфосинтаксические явления (далее для краткости данный уровень будет называться *морфологическим*).

Под термином «афинность» понимается прежде всего схожесть либо идентичность определенной словоформы с белорусской или русской литературной формой, или же, если словоформу невозможно отнести к одному или другому языку, классификация ее как ‚общей‘ или ‚гибридной‘ в соответствии с выше упомянутым алгоритмом. Как известно, в белорусских диалектах имеются лексические единицы, совпадающие скорее с функционально или семантически эквивалентными единицами русского литературного языка, а не с соответствующими эквивалентами русского литературного языка (ср.



RAMZA 2008). На подобные обстоятельства внимание обращать не будет, чему имеется целый ряд причин. Во-первых, в справочной диалектологической литературе отмечена только часть представленной в корпусе лексики. Кстати, диалектного словаря, который охватывал бы весь белорусский языковой ареал, на сегодняшний день не существует. Во-вторых, внимание в любом случае следовало бы обращать только на диалекты, распространенные в местах проживания респондентов корпуса. В-третьих, и это самая главная причина, следует учитывать следующее: если в БРСР определенного говорящего в одном из высказываний встречается лексема, характерная как для его диалекта, так и для русского литературного языка, но не для белорусского, то в подобных конкретных случаях в рамках одного высказывания невозможно определить, откуда этот респондент «взял» данную лексему: из своего диалекта или же из русского языка. Только в тех случаях, если определенная языковая единица отсутствует в литературных языках, проверяется ее наличие также в белорусских диалектах и в русском просторечии. Классификация по категориям ‚белорусский‘ и ‚русский‘, ориентирующаяся в первую очередь на аффиность с литературными языками и обращающаяся к аффиности с субстандартами только при условии отсутствия определенной единицы в литературных языках, представляет собой научно проверяемый (фальсифицируемый) метод. Конечно, многие словоформы невозможно однозначно отнести к белорусскому или русскому языку: речь здесь идет об ‚общих‘ или ‚гибридных‘ языковых единицах, что будет подробнее рассмотрено дальше.

#### 3.4.3.1 Фонетико-фонологический уровень

На данном уровне будут рассматриваться шесть чисто фонетико-фонологических различий между белорусским и русским языками, которые, как известно, относятся к числу наиболее значимых различий в произношении обоих языков.

- рус. [dʲ], [tʲ] vs. бел. [dʒʲ], [tʂʲ] (так называемые дзеканье и цеканье в белорусском языке); ср. рус. *дети* vs. бел. *дзеці*
- рус. взрывной [g] vs. бел. фрикативный [ɣ] или [ɦ]; ср. рус. *нога* vs. бел. *нага*

- рус. мягкий [rʲ] vs. бел. твердый [r], ср. рус. *кричать* vs. бел. *крычаць*, рус. *резать* vs. бел. *рэзаць*
- рус. [v] в конце слова или перед согласным vs. бел. [ũ] или, в зависимости от позиции, [u]; ср. рус. *врач* vs. бел. *урач* или *ўрач*. Звук [ũ] классифицируется как ‚белорусский‘ и тогда, когда он соответствует рус. [ɪ] или [u], как, например, в *даў* (бел. *даў*, рус. *дал*) или *ўжэ* (рус. *уже*). Звук же [ɪ] или [u], соответствующий бел. [ũ], рассматривается как ‚русский‘ не только на фонетико-фонологическом уровне, т.к. в сравнимых фонологических контекстах также и в белорусском языке возможно употребление [ɪ] или [u].
- рус. иканье vs. бел. яканье, т.е. [a]-подобная реализация /a/, /e/, /o/ в первом предударном слоге после мягких согласных в белорусском языке или [i]-подобная реализация в русском языке; ср. рус. *ле[ɪ]жать* vs. бел. *ляжаць*. Межъязыковые различия в редукации гласных после мягких и твердых согласных в других безударных слогах не рассматриваются.
- рус. мягкие [ʃʲ] и [ʃʲ:] vs. бел. твердые [ʃ] и [ʃʲ]; ср. рус. *ч[ʃʲ]ерт* vs. бел. *ч[ʃ]орт*.

Если словоформа содержит одну или несколько ‚белорусских‘ черт, то на фонетико-фонологическом уровне она классифицируется как фонетически ‚белорусская‘. Если она содержит ‚русские‘ черты, то она классифицируется как фонетически ‚русская‘. Если же она содержит хотя бы одну ‚русскую‘ черту и хотя бы одну ‚белорусскую‘, то она классифицируется как фонетически ‚гибридная‘. А если она не содержит ни одной из перечисленных черт, то она классифицируется как фонетически ‚общая‘.

Сокращение	Расшифровка	Примеры
@pb	фонетически ‚белорусская‘	jamu@c@pr; try@c@pb
@pr	фонетически ‚русская‘	jimu@c@pr; tri@c@pr
@pc	фонетически ‚общая‘	jemu@c@pc; da@r@pc
@ph	фонетически ‚гибридная‘	jišćé@b@ph; havorêc´@b@ph

### 3.4.3.2 Морфологический уровень

На так называемом морфологическом уровне, охватывающем – как было отмечено ранее – лексические, морфонологические, морфологические и морфосинтаксические явления, фонетико-фонологические различия не принимаются во внимание. Так, например, словоформа *dzelâ* (рус. *дело*, бел. *справа*), несмотря на ‚белорусское‘ дзеканье, на морфологическом уровне квалифицируется не как ‚белорусская‘ или ‚гибридная‘, а как ‚русская‘. Таким же образом словоформа *sibry*, несмотря на считающееся ‚русским‘ иканье (ср. NENTSCHEL & ZELLER 2014) вместо ‚белорусского‘ яканья (рус. *друзья*, бел. *сябры*), определяется на морфологическом уровне как ‚белорусская‘, а не как ‚русская‘ или ‚гибридная‘.

Основной единицей при определении языковой афинности на морфологическом уровне является морф, т.е., прежде всего, корень, а также – если имеются – приставки, суффиксы окончания и постфиксы словоформ. Относительно каждого морфа задается вопрос, является ли он в данной комбинации формы и значения составным элементом современного белорусского или русского языка либо же обоих языков. Т.е. существуют следующие возможности определить языковую афинность морфов. (Для данного анализа используются определенные словари<sup>18</sup>: ТСБМ 1977, БРС 2003 и РБС 2005.)

---

18 В некоторых случаях белорусским словарям свойственна определенная степень пуризма в отношении принятия русизмов в качестве лексем белорусского литературного языка. Таким образом, существуют разные подходы к кодификации, выходящие за пределы различий между «тарашкевицей» и «наркомовкой». Т.к. число «активных» носителей белорусского литературного языка очень ограничено, даже среди белорусских белорусистов иногда наблюдаются (как показала работа в рамках названного выше проекта) значительные различия во взглядах относительно того, можно ли считать определенную лексему приемлемым заимствованием из русского языка или же ее следует рассматривать как русский «варваризм». Единственным методом в таком случае остается классификация, ориентированная на определенные словари. Здесь следует отметить, что во многих случаях даже наши коллеги-белорусисты из Белорусского государственного университета высказывали различные мнения относительно того, является ли определенное слово русизмом, недопускаемым литературной нормой белорусского языка, или же приемлемым заимствованием из русского языка. Данная неуверенность относительно норм белорусского языка не вызывает удивления, если принять во внимание языковые реалии в стране.

- Специфически ‚русский‘, как, например, корень существительного *jazuk* (рус. *язык*; бел. *мова*) или союз *no* (рус. *но*; бел. *але*).
- Специфически ‚белорусский‘, как, например, корень существительного *bas'kâ* (рус. *отец*; бел. *бацька*) или предлог *dzelê* (рус. *ради*; бел. *дзеля*).
- ‚общий‘ (имеющийся в обоих языках), например, частица *tak* (рус. *так*; бел. *так*; однако *tak* в значении «да» было бы определено как ‚белорусское‘, а *da* в том же значении – как ‚русское‘).
- ‚гибридный‘, как, например, корень указательного местоимения *êtyx* (рус. *этих*; бел. *этых*).

Относительно последнего примера: в корпусе встречаются также ‚гибридные‘ морфы, гибридность которых состоит в том, что их морфонематическая репрезентация имеет как белорусские, так и русские черты. Таким образом, корень словоформы *êtyx* является гибридным, т.к. его русский эквивалент *этих* имеет мягкий конечный звук основы, а для белорусского эквивалента *этых* характерен протетический звук.

Если же рассматриваемые словоформы в совокупности своих морфов не соответствуют ни одному из языков-источников в их литературной форме, то в таком случае проверяется их наличие в белорусских диалектах. Если словоформы имеются в диалектах, то они, точнее их морфы, дополнительно к характеристике ‚белорусские‘ получают характеристику ‚диалектные‘. Если же соответствий нет и в белорусских диалектах, то их наличие проверяется в русском просторечии. Если словоформа, или морф, встречается в просторечии, то к характеристике ‚русская‘ добавляется характеристика ‚просторечная‘. Т.е. характеристика ‚просторечная‘ встречается только в том случае, если определенной формы нет ни в одном из литературных языков (ни в белорусском, ни в русском), а также ее нет в белорусских диалектах. Если же словоформы, т.е. морфа нет и в просторечии, то он классифицируется как ‚неясный‘.

Афинность отдельных словоформ «подсчитывается», как правило (исключения приведены ниже), на основе афинности морфов, из

которых они состоят. «Подсчет» проводится по следующим правилам: словоформа классифицируется как ,русская‘, если она состоит либо из исключительно ,русских‘ морфов, либо ,русских‘ и ,общих‘ морфов. Соответственно, ,белорусскими‘ считаются такие словоформы, которые содержат либо только ,белорусские‘, либо ,белорусские‘ и ,общие‘ морфы. ,Общие‘ же словоформы содержат только ,общие‘ морфы. Последнюю группу составляют ,гибридные‘ словоформы, содержащие как ,белорусские‘, так и ,русские‘ (возможно, также и ,общие‘) морфы или же содержащие хотя бы один ,гибридный‘ морф. Это значит, что при определении афинности ,общие‘ морфы не играют никакой роли, если кроме них словоформа содержит хотя бы один специфический, т.е. ,белорусский‘ или ,русский‘, морф, или же хотя бы один ,гибридный‘ морф.

Сокращение	Значение	Возможные комбинации морфов
@b	,белорусская‘	,белорусский/-ие‘ (+ ,общий/-ие‘)
@r	,русская‘	,русский/-ие‘ (+ ,общий/-ие‘)
@c	,общая‘	,общий/-ие‘
@h	,гибридная‘	,белорусский/-ие‘ + ,русский/-ие‘ (+ ,общий/-ие‘); ,гибридный/-ие‘ (+ ,русский/-ие‘ / ,белорусский/-ие‘ / ,общий/-ие‘)
@u	,неясная‘	,неясный‘ (+,общий/-ие‘ / ,русский/-ие‘ / ,белорусский/-ие‘ / ,гибридный/-ие‘)

Словоформы могут также получить дополнительное сокращенное обозначение ,диалектная‘ или ,просторечная‘.

Сокращение	Значение
@pros	,(русско-)просторечная‘
@dial	,(белорусско-)диалектная‘

Также могут встречаться словоформы, состоящие из исключительно ,общих‘ морфов, которые в данном значении характерны только для одного из языков-источников. Например, словоформа *žonkâ* (рус. *жена*, мн. ч. *жены*, бел. *жонка*): корень *жон/жен-*, суффикс *-к-* и окончание *-a* встречаются как в русском, так и в белорусском языке, однако комбинация этих морфов, в качестве нейтрального обозначения

жены, характерна только для белорусского языка. В таком случае словоформа *žonkâ* в целом классифицируется как ‚белорусская‘.

К похожей ситуации может привести также ударение. Как в русском, так и в белорусском языке ударение имеет смысловозначительную функцию. В обоих языках оно является морфологизированным, т.е. его нужно заучивать, причем ориентируясь как на лексемы: например, бел.=рус. *мука* vs. *мука*, так и на отдельные словоформы: например, рус. *города* (род. п., ед. ч.) vs. *города* (имен. п., мн. ч.). Ударение в белорусском и русском языках не может рассматриваться как исключительно фонетическое явление, т.к. известно, что в обоих языках оно является лексически или морфологически мотивированным. Так, словоформа *vesci* с ударением на первом слоге (ср. рус. *вести* с ударением на последнем слоге и бел. *vesci* с ударением на первом слоге) классифицируется как ‚белорусская‘, несмотря на то, что все ее морфологические составные части, т.е. корень *ves-* и окончание *-ти / -ці* являются ‚общими‘.<sup>19</sup>

Свободные элементы, как, например, корни в словоформах рус. *бумаг-а* vs. бел. *папер-а*, классифицируются парадигматически, это значит, что корни определяются как ‚белорусский‘ или ‚русский‘ без синтагматического контекста. Многие же связанные элементы: окончания, приставки, а также большинство предлогов, – присутствуют в обоих языках, и их можно отнести к одному или другому языку только на основании синтагматических критериев. По этой причине в подобных случаях рассматривается также контекст, в котором находится морф: так, например, предлог *na/no* «как таковой», т.е. парадигматически, есть и в русском, и в белорусском языке, и в большинстве случаев его употребление в обоих языках совпадает. Однако в конструкции типа *найсці на хлеб* имеет место специфически белорусское употребление предлога *na*, поэтому в подобных случаях он определяется как ‚белорусский‘. В русском языке данное значение выражается при помощи предлога *за*: *пойти за хлебом*. В таком предложении предлог *за* был бы, соответственно, определен как ‚русский‘. Подобным образом рассматриваются и окончания: при определении окончания, имеющегося в обоих языках, как ‚белорусского‘, ‚русского‘ или ‚общего‘ внимание обращается на синтагматический контекст. Решающим является то, возможно ли

---

19 В опубликованной версии ОК-БРСР ударение указано только в PDF-версии .

сочетание определенной основы и определенного окончания только в русском, только в белорусском или же в обоих языках. Так, существительные в именительном падеже, множественном числе могут иметь окончание *-ы* (или *-и/-і*) как в русском, так и в белорусском языке: рус. *студент-ы* и бел. *студэнт-ы*. Однако в словоформе *prafesârŭ* окончание *-у* определяется как ‚белорусское‘, ‚русским‘ соответствием было бы окончание *-а*; ср. рус. *профессор-а*; бел. *прафесар-ы*.

Принципиально различные элементы, как, например, бел. *каля* и рус. *около*, определяются как принадлежащие одному или другому языку, конечно же, без принятия во внимание контекста.

Т.к. языковая афинность на фонетико-фонологическом и на «более глубоком», морфологическо-морфосинтаксическом уровнях является основной для дальнейшего анализа белорусско-русской смешанной речи, она указывается в ОК-БРСР два раза, чтобы обеспечить по возможности легкий поиск в корпусе. Во-первых, она представлена в строке **%mor** (строка с аннотациями), о чем говорится в пункте 3.5. А во-вторых, в строке с затранскрибированными высказываниями она отмечена после каждой словоформы, причем сначала указывается морфологическая афинность, затем фонетическая, а затем, в некоторых случаях, является ли словоформа кроме того белорусско-‚диалектной‘ или ‚просторечной‘.

Пример	Расшифровка
ja@c@pc	‚общая‘; фонетически ‚общая‘
dumâjêcê@c@pb	‚общая‘; фонетически ‚белорусская‘
kudy@b@pc	‚белорусская‘; фонетически ‚общая‘
vot@r@pc	‚русская‘; фонетически ‚общая‘
pakudâvâ@r@pc@pros	‚(русско-)просторечная‘; фонетически ‚общая‘
brynčéc´@b@pb@diał	‚(белорусско-)диалектная‘; фонетически ‚белорусская‘

#### 3.4.4 Дополнительная информация по словоформам

- Стяженные формы типа *tada* (рус. *тогда*) или *tokâ* (рус. *только*) дополнительно к обозначению языковой афинности получают обозначение ( **@w** ) (Пример: *tada@r@pc@w*<sup>20</sup>).

20 Известно, что в случае форм типа *tada* в белорусских или русских субстандартах существует определенная неуверенность относительно того, стоит ли говорить о «все

- Метаязыковые названия звуков и букв дополнительно к обозначению языковой афинности получают обозначение (**@1s**). (Пример: o@c@pc@1s; также см. пункт 3.4.5.3)
- Неполные или частично непонятные словоформы получают обозначение (**@k**) (Пример: dastu@k@pc).

### 3.4.5 Пунктуация

#### 3.4.5.1 Простые знаки препинания

- В соответствии с общепринятыми правилами пунктуации употребляются следующие знаки: ( , ), ( . ), ( ! ) и ( ? ).
- Двоеточие ( : ) употребляется перед перечислением (не перед прямой речью, см. ниже):

*Пример:*

```
*SFA: my@c@pc što@c@pc kupilê@c@pc sa@b@pc svadzêbnyx@r@pb
      dzenêh@r@pb : škaf@r@pc , tryl'jaž@c@pb i@c@pc
      kravac'@r@pb .
```

- Вместе с распространенными в Chat знаками препинания, в предложениях, в которых белорусская и русская орфография требуют постановки тире (например, в предложениях с глаголом-связкой или в неполных предложениях), употребляется знак ( [-] ).

*Пример:*

```
*AFP: tam@c@pc balotâ@c@pc , paétâmu@r@pc , v@r@pr
      pryncýrê@c@pb , samâ@b@pc âpcimal'nâ@r@pb [-]
      izmel'čenêjë@r@pr i@c@pc razbrasyvânêjë@r@pc .
```

---

еще» фонетически слабой форме либо об «уже» закрепившейся (морфологизированной) фонологической форме (ср. Генчэль 2013). Т.к. согласно представленной выше методике сравнение проводится в первую очередь с литературными нормами, то предпочтение отдается первому варианту.



### 3.4.5.2 Прямая речь

В корпусе используется двойное обозначение прямой речи: во-первых, в предложении, вводящем прямую речь (так называемых «словах автора», если они имеются), а во-вторых, в самой прямой речи:

- Если слова автора стоят перед прямой речью: (+"/. )
- Если слова автора стоят после прямой речи: (+". )
- В начале прямой речи, а также в начале каждого предложения внутри прямой речи ставится знак (+". )

*Примеры:*

\*S3B: i@c@pc vsë@c@pr ja@c@pc (.) na@c@pc bac´ku@b@pb  
hâvaru@c@pb +"/.

\*S3B: +". bac´kâ@b@pb , vaz´mi@c@pc ty@c@pc âtruhaj@r@pb .

\*S3B: +". i@c@pc kak-tâ@r@pc vot@r@pc vyžýlê@c@pc .

\*S3B: hâvaryt@r@pb +".

- Отличие от правил, принятых в СНАТ: если в прямую речь вставлен короткий элемент типа dumâju, kažý, havorýc´ и под., то он не рассматривается как отдельное высказывание. Вместо этого данный вводный элемент внутри прямой речи отмечается следующим образом: ( ["-" ] )

*Пример:*

\*SAF: +". ja@c@pc haryt@r@pb@w ["-"] iščo@r@ph sutkê@c@pc  
pabudu@c@pc .

### 3.4.5.3 Метаязыковые элементы

Метаязыковые элементы обозначаются знаком ( [" ] ). Этот знак указывает как на непосредственно метаязыковые элементы, так и на различные названия, приложения и т.д. Если метаязыковые элементы, названия и т.д. состоят из более чем одной словоформы, то они берутся в угловые скобки.

*Примеры:*

\*S3B: nu@c@pc , slovâ@c@pc bêlaruskâjê@c@pc , zarâz@b@pc [" ] .

\*L2B: u@c@pc nix@r@pc boléjê@r@pc tako@c@pc o@c@pc@ls ["  
 vydzêlênâ@c@pb , v@r@pr râsian@c@pc .

\*SI2: nu@c@pc my@c@pc ix@r@pc zavëm@c@pc ix@r@pc <kléckê@c@pc  
 z@b@pc mjasâm@c@pc>["] .

\*INT: a@c@pc vot@r@pc znajêcê@c@pb , možý@b@pc , takuju@c@pc  
 hrupu@c@pb krâmbambulê@c@pc ["] ?

#### 3.4.5.4 Повторения, исправления, переформулировки, переводы

В системе СНАТ используются следующие знаки: ([/]) для обозначения повторений, ([//]) для исправлений и ([///]) для переформулировок. Однако СНАТ не предусматривает никакой аннотации (грамматических и т.д.) для повторяющегося языкового материала. Поэтому данные знаки не употребляются в ОК-БРСР, что, с одной стороны, препятствует исключению повторений и т.д. при использовании некоторых инструментов программы CLAN, с другой же стороны, делает возможной грамматическую аннотацию рассматриваемого материала. В ОК-БРСР используются следующие обозначения:

- Повторения: ([\_/1])
- Исправления: ([\_//1])
- Переформулировки: ([\_///1])
- Переводы: ([\_////1])

«Перевод» здесь понимается в широком смысле: знак перевода ставится в том случае, если рядом стоящие словоформы (или комбинации словоформ) имеют одно и то же значение, но разную языковую афинность. При этом дословный и свободный перевод не различаются.

Названные знаки ставятся после элементов, которые повторяются, исправляются, переформулируются или переводятся, т.е. перед повторениями, исправлениями, переформулировками и переводами.

Если речь идет о более чем одной словоформе, то все словосочетание заключается в угловые скобки.

*Примеры:*

\*CFK: jany@b@pb šcé@b@pb@w tut@c@pc máladéjê@b@pc@dial [\_/]  
máladéjê@b@pc@dial saúsem@r@pb .

\*CI4: synâ@c@pc vydava@k@pc zam@k@pc [\_//] žynilâ@c@pc ,  
nu@c@pc .

\*AI2: da@r@pc , étym@h@pc ja@c@pc [\_///] bez@c@pc étâvâ@r@pc  
ne@c@pc magu@c@pcr .

\*AFA: nu@c@pc jano@b@pb , znajêš@c@pc , jak@b@pc što@c@pc  
<pad@c@pc škurâju@b@pc>[\_////] pad@c@pc kožýju@h@pc  
jak@b@pc što@c@pc tam@c@pc , o@c@pc tak@c@pc  
pêrýcirajê@b@pb .

### 3.4.5.5 Неполные предложения

В данном пункте нет отличий от правил, принятых в СНАТ. Здесь проводится различие между предложениями, оборванными самим говорящим и предложениями, когда говорящего перебивает его собеседник.

- После высказываний, оборванных самим говорящим, ставится знак ( + . . . )
- После вопросительных высказываний, оборванных самим говорящим, ставится знак ( + . . ? )
- После высказываний, перебитых собеседником, ставится знак ( + / . )
- После вопросов, перебитых собеседником, ставится знак ( + / ? )
- Если высказывание, перебитое собеседником, продолжается дальше, то перед его продолжением ставится знак ( + , )

*Примеры:*

\*S7B: budut@r@pc učyccý@r@pb , moží@b@pc , kto-tâ@r@pc ú@b@pb  
minsku@b@pc astanêccý@r@pc , moží@b@pc +...

\*S5B: pajedzê@b@pb pa@b@pc rasijë@b@pc , za@b@pc hranicâj@b@pb  
+/.  
\*INT: nu@b@pc , vot@r@pc râskažycê@b@pb .  
\*S5B: +, pajez´dzêú@b@pb .

### 3.4.5.6 Паузы и наложения речи одного говорящего на речь другого

В данном пункте нет отличий от правил, принятых в СНАТ. Следует отметить только то, что паузы и наложения речи нотируются достаточно свободно.

- Паузы различаются по продолжительности. Короткие паузы отмечаются знаком ( . ) , длинные знаком ( . . ) . Только в случае чрезвычайно длинных пауз употребляется знак ( . . . ) .
- При наложениях речи говорящих, т.е. когда второй говорящий начинает свое высказывание, не дав договорить первому, перед вторым высказыванием ставится ( +< ) .

*Примеры:*

\*RFN: ja@b@pc pašla@r@pc (..) v@r@pc adzinâccýty@h@pb klas@b@pc .

\*CFA: nja@b@pb budzêc´@b@pb@dial žý@b@pc jna@b@pc úsem@r@pb  
xadzic´@b@pb raskazyvâc´@r@pb .

\*CFC: +< a@b@pc mamkâ@b@pc nikomu@b@pc ne@b@pc havorêc´@b@pb .

### 3.5 Аннотации в строке %mor

Строка %mor содержит белорусское и / или русское заголовочное слово, информацию о морфологической и фонетической языковой афинности словоформы, а также о ее части речи и грамматические показатели.

Для указания информации о „suffixes“ СНАТ предусматривает дефис ( - ), для „fusional or infixed morphology“ – знак ( & ), для приставок знак

( # ) (ср. McWHINNEY 2013a, 109). Данные знаки используются в ОК-БРСР для указания языковой афинности и грамматических категорий. В отличие от системы СНАТ, целью ОК-БРСР является не морфологический анализ, а указание грамматических показателей. Поэтому различия между „fusional suffixes“ и „suffixes“, т.е. различия в способе кодировки грамматической информации не проводится.

Используемые в ОК-БРСР сокращенные обозначения частей речи и грамматических показателей ориентируются на правила системы СНАТ, однако в некоторых аспектах отличаются от них.

Схема аннотаций выглядит следующим образом:

МорфологическаяАфинность : ФонетическаяАфинность #ЧастьРечи |\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_ & (Падеж/Время) -  
(Род/Лицо) : Число

Т.е. последовательность подачи информации выглядит так:

- сначала информация о морфологической афинности;
- затем, отделенная двоеточием ( : ), информация о фонетической афинности;
- затем, после знака ( # ), указывается часть речи;
- затем, после знака ( | ), приводится белорусское или русское (или оба) заголовочное слово;
- затем (в случае изменяемых частей речи), после знака ( & ), указывается падеж либо время;
- затем, после знака ( - ), указывается род или лицо;
- затем, после знака ( : ), указывается число.

В случае неизменяемых частей речи указываются только языковая афинность, часть речи и заголовочное слово (либо заголовочные слова).

### 3.5.1 Языковая афинность

Для указания языковой афинности используется знак ( # ), предусмотренный системой СНАТ для указания приставок. Сначала

указывается морфологическая афинность, затем, через двоеточие, фонетическая афинность.

**МорфологическаяАфинность : ФонетическаяАфинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_& (Падеж/Время) - (Род/Лицо) : Число**

Используются такие же сокращенные обозначения, как и в строках респондентов (см. 3.4.3.).

*Пример:* **r:pg#pro:pers:ana|\_0\_oni\_&NOM-0:PL**

### 3.5.2 Часть речи

Часть речи указывается после языковой афинности перед заголовочным словом, от которого она отделяется знаком ( | ).<sup>21</sup>

Афинность#**ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_& (Падеж/Время) - (Род/Лицо) : Число**

В случае прилагательных и наречий здесь также указываются степени сравнения.

Сокращение	Значение
adj:abrev	Аббревиатура от прилагательного (напр., рус. <i>žedě</i> как краткая форма от рус. <i>железнодорожный</i> )
adj:brev	Краткая форма прилагательного
adj:cp	Прилагательное в сравнительной степени
adj:indecl	Несклоняемое прилагательное
adj:loc	Прилагательное, образованное от названия населенного пункта
adj:poss	Притяжательное прилагательное
adj:prop	Прилагательное, образованное от имени собственного
adv	Наречие

21 Как известно, достаточно проблематичным является разграничение разных частей речи среди неизменяемых слов. Т.к. по причине большого объема материала большая часть работы по определению частей речи выполнялась студентами, в ОК-БРСР применялась достаточно простая, не совсем лишенная противоречий схема. Поэтому в отдельных случаях при классификации словоформ может встречаться варьирование.

adv:conj	Наречие/союз
adv:cp	Наречие в сравнительной степени
adv:pro	Наречие/местоимение
adv:ptl	Наречие/частица
co	Междометие
conj	Союз
conj:ptl	Союз/частица
conj:pro	Союз/местоимение
disc	Дискурсный маркер
n:abrev	Аббревиатура
n:adj	Субстантивированное прилагательное
n:adj:anim	Одушевленное субстантивированное прилагательное
n:adj:inanim	Неодушевленное субстантивированное прилагательное
n:adj:prop	Субстантивированное прилагательное, образованное от собственного имени
n:anim	Одушевленное существительное
n:inanim	Неодушевленное существительное
n:loc	Название населенного пункта
n:prod	Название продукта
n:prop	Имя собственное
na	Информация отсутствует
num:kard	Количественное числительное
num:koll	Собирательное числительное
num:ord	Порядковое числительное
num:pro	Числительное/местоимение
part:akt	Активное причастие
part:pass	Пассивное причастие
part:pass:brev	Краткая форма пассивного причастия
pred	Предикатив
prep	Предлог
pro:dem	Указательное местоимение
pro:dem:kop	<i>это / эта</i> в предложениях с глаголом-связкой
pro:det	Определительное местоимение (напр., рус. <i>всякий, каждый</i> )
pro:ind	Неопределенное местоимение
pro:int	Вопросительное местоимение
pro:neg	Отрицательное местоимение

pro:pers:ana	Личное местоимение 3-го лица
pro:pers:dei	Личное местоимение 1-го и 2-го лица
pro:poss:ana	Притяжательное местоимение 3-го лица
pro:poss:dei	Притяжательное местоимение 1-го и 2-го лица
pro:poss:refl	Возвратное притяжательное местоимение
pro:ptl	Местоимение/частица
pro:refl	Возвратное местоимение
pro:rez	Реципрокальное местоимение
pro:sonst	Другое местоимение
ptl	Частица
v:comasp	Двувидовый глагол
v:ger	Деепричастие
v:ipf	Глагол несовершенного вида
v:pf	Глагол совершенного вида

### 3.5.3 Заголовочное слово

Там, где в системе СНАТ предусматривается указание «основы», в ОК-БРСР указывается начальная форма слова, т.е. инфинитив в случае глаголов, именительный падеж в случае существительных (более подробная информация представлена ниже).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_ & (Падеж/Время) - (Род/Лицо) :Число

Некоторые словоформы имеют как белорусское, так и русское заголовочное слово, другие же только белорусское или только русское, что зависит от языковой афинности основы; классификация окончаний не играет никакой роли. При этом рассматривается только морфологический уровень.

- Если основа определена как ‚русская‘, то указывается только русское заголовочное слово. Так, например, словоформа *jizyk*, основа которой определена как ‚русская‘, имеет только русское заголовочное слово *jazuk*.
- Если основа определена как ‚белорусская‘, то указывается только белорусское заголовочное слово. Так, например, основа



словоформы *мова* определена как ‚белорусская‘, поэтому словоформа имеет только белорусское заголовочное слово *мова*.

- Если основа словоформы ‚общая‘, то указывается как белорусское, так и русское заголовочные слова. Так, например, основа словоформы *syn* определена как ‚общая‘, поэтому для данной словоформы фиксируется как белорусское, так и русское заголовочное слово *syn*. Если же основа определена как ‚общая‘, однако вся словоформа в целом классифицирована как ‚русская‘, то для словоформы указывается только русское заголовочное слово. Так, в словоформе *âb'jasnilâ* все морфы, и, соответственно, основа, определены как ‚общие‘, однако вся словоформа в целом классифицирована как ‚русская‘, что означает, что для нее фиксируется только русское заголовочное слово *ob'jasnit'*. Если же основа определена как ‚общая‘, а вся словоформа в целом как ‚белорусская‘, то она получает только белорусское заголовочное слово. Так, например, словоформа *žonkâ* имеет только белорусское заголовочное слово *žonka*.
- Если основа словоформы ‚гибридная‘, то действует принцип наибольшей схожести: в зависимости от того, эквиваленту из какого языка (белорусского или русского) рассматриваемая словоформа «ближе», указывается, соответственно, белорусское или русское заголовочное слово. Так, например, словоформа *âbižaúsê* имеет русское заголовочное слово *obižat'sja*, а словоформа *sënnëšnêj* – белорусское заголовочное слово *sënnjašni*. В неоднозначных случаях для словоформы фиксируются и белорусское, и русское заголовочные слова. Так, например, словоформа *hétix* имеет как русское заголовочное слово *étot*, так и белорусское заголовочное слово *hëty*.
- Ни белорусское, ни русское заголовочное слово не указывается в том случае, если основа классифицирована как ‚неясная‘.

Заголовочные слова представляются в ОК-БРСР в строке **mor%** следующим образом: если словоформа имеет два заголовочных слова, то сначала указывается белорусское, а затем русское, причем между собой заголовочные слова разделены знаком нижнего подчеркивания (  ). Перед белорусским и после русского заголовочного слова также

ставится знак нижнего подчеркивания. Если же словоформа имеет только белорусское или только русское заголовочное слово, то на месте отсутствующего заголовочного слова стоит ( 0 ). (Если отсутствуют оба заголовочных слова, то ( 0 ) ставится два раза: на месте белорусского и на месте русского заголовочных слов.)

Словоформа	Афинность основы	заголовочное слово
ja	,общая‘	_ja ja_
paperâ	,белорусская‘	papera 0
bumahâ	,русская‘	0 bumaga
étyjê	,гибридная‘	hétý étot

Замечания к нотированию заголовочных слов:

- В основном заголовочные слова транслитерируются в соответствии с правилами научной транслитерации, однако следует обратить внимание на следующие дополнения и отклонения:

кириллица	Транслитерация в ОК-БРСР:
х	ch (отличается от транскрипции словоформ)
э	é
ь	˘ (ASCII 180)
ъ	˘˘
’	˘
й	ú
бел. з	h
рус. з	g
бел. ě	ě (ASCII 235)
рус. ě	e

- Все заголовочные слова пишутся с маленькой буквы.
- Части сложных слов, пишущихся в русском / белорусском языке через дефис, соединяются при помощи знака ( / ) (Пример: \_0\_sjuda/tuda\_).
- То же касается и клитиков, пишущихся в русском / белорусском языке через дефис (Пример: \_dzes˘ci\_gde/to\_).

- Имена людей на уровне заголовочного слова получают также указание `_persname_`.
- В качестве заголовочных слов кратких прилагательных (или кратких причастий) указываются, как правило, полные формы, за исключением случаев, когда полных форм не существует (как, например, рус. *должен* или бел. *павінен*). В таких случаях в качестве заголовочного слова указывается краткая форма.
- В качестве заголовочных слов деепричастий указываются не инфинитивы, а деепричастия (*Пример: \_pryvykšy\_0\_*).
- В случае анафорических местоимений сохраняется различие по роду и числу: для словоформы *jejo* в качестве заголовочного слова указывается `_0_ona_`, *jevo* – `_0_on_` или `_0_ono_`, а для *ix* указывается заголовочное слово `_0_oni_`.
- В случае ‚русской‘ сравнительной степени в качестве заголовочного слова указывается сравнительная форма, причем с суффиксом *-ee*, а не *-ей*, например, `_0_dobree_`.
- Если словоформа начинается с *pol-*, то при указании заголовочного слова на *pol-* внимание не обращается. Так, например, для словоформы *polstala* в качестве заголовочного слова фиксируется `_0_stol_`.
- В качестве заголовочных слов кратких вариантов частиц указываются их полные варианты. Так, например, частица *b* в качестве заголовочного слова получает `_by_by_`.
- В качестве заголовочных слов вариантов предлогов указываются их краткие варианты. Так, например, для предлог *va* в качестве заголовочного слова фиксируется `_0_v_`.
- В случае белорусских диалектных словоформ предпочтение отдается их литературным соответствиям. Так, например, в качестве заголовочного слова диалектной словоформы *dak* указывается литературная форма `_dyk_0_`, а в качестве заголовочного слова диалектной словоформы *ina* – литературная форма `_jana_0_`. Если же диалектная и литературная формы сильно различаются, то в качестве заголовочного слова указывается диалектная форма.

- В случае омонимии между разными частями речи заголовочные слова могут различаться: если *étâ* (рус. *это*) или *hétâ* (бел. *héta*) употребляется в роли частицы, то в качестве заголовочного слова указывается *\_0\_éto\_* или *\_héta\_0\_*, а если в роли указательного местоимения, то *\_0\_étot\_* или *\_héty\_0\_*.
- Если в случае окказионализма, не имеющего соответствия в словарях, можно установить его начальную форму, то она указывается в качестве заголовочного слова. Так, например, заголовочным словом словоформы *âtžalelâ* является *\_0\_otžalet'\_*.
- Если из контекста можно определить начальную форму неполной словоформы, то она указывается в качестве заголовочного слова. Так, например, заголовочным словом неполной словоформы *topč@k* (вместо рус. *тончеться*) является *\_0\_toptat'sja\_*.
- Если начальную форму неполной словоформы нельзя определить из контекста (а также в других неясных случаях), заголовочное слово не указывается, т.е. на месте белорусского и русского заголовочных слов ставится *\_0\_0\_*.
- Заголовочные слова составных частей идиоматических выражений конструируются. Так, например, в случае выражения *z buchty barachty* для словоформы *buchty* фиксируется в качестве заголовочного слова *\_0\_buchta\_*, а для словоформы *barachty* – *\_0\_barachta\_*.
- Русские диалектные формы указываются также в качестве заголовочных слов. Например, заголовочным словом русской диалектной словоформы *ichnêj* является *\_0\_ichnij\_*.
- В случае существительных, употребляемых только во множественном числе, в качестве заголовочного слова указывается форма множественного числа.
- В случае реципрокальных местоимений (*drug\_drugâ*) в качестве заголовочного слова указывается форма родительного падежа: *\_0\_drug\_druga\_*, *\_adna\_adnoj\_0\_*.

### 3.5.4 Склонение

Грамматические категории именных частей речи указываются в следующем порядке:

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ПАДЕЖ-РОД:ЧИСЛО

В неоднозначных случаях рассматривается контекст. Например, в случае омонимии падежей внимание обращается на синтагматику или контекст словоформы, а для определения рода прилагательных во множественном числе – на род элементов, к которым они относятся. Только в том случае, если и контекст ясности не вносит, для соответствующей категории вносится показатель «ambiguous» (неоднозначный).

#### 3.5.4.1 Падеж

Падеж указывается после заголовочных слов, от которых он отделяется знаком ( & ).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ПАДЕЖ-РОД:ЧИСЛО

Сокращение	Значение
NOM	Именительный падеж
GEN	Родительный падеж
DAT	Дательный падеж
ACC	Винительный падеж
INS	Творительный падеж
LOC	Местный/предложный падеж
VOC	Звательный падеж
CAMB	Падеж невозможно определить однозначно
0	Категория отсутствует

### 3.5.4.2 Род

Род указывается после указания падежа, от которого он отделяется знаком ( - ).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ПАДЕЖ-РОД: ЧИСЛО

Сокращение	Значение
M	Мужской род
F	Женский род
N	Средний род
GC	Общий род
GAMB	Род невозможно определить однозначно
0	Категория отсутствует

### 3.5.4.3 Число

Число указывается после указания рода, от которого он отделяется знаком ( : ).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ПАДЕЖ-РОД: ЧИСЛО

Сокращение	Значение
SG	Единственное число
PL	Множественное число
0	Категория отсутствует

### 3.5.5 Спряжение

Грамматические категории глагола указываются в следующей последовательности:

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ВРЕМЯ-ЛИЦО\_или\_РОД: ЧИСЛО

### 3.5.5.1 Время и повелительное наклонение

Время глагола указывается после заголовочных слов, от которых оно отделяется знаком ( & ).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ВРЕМЯ-ЛИЦО\_или\_РОД:ЧИСЛО

Сокращение	Значение
IMPER	Повелительное наклонение
PRAES	Настоящее время
PRAET	Прошедшее время
INF	Инфинитив
0	Категория отсутствует

### 3.5.5.2 Лицо

Лицо глагола указывается после указания времени, от которого оно отделяется знаком ( - ).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ВРЕМЯ-ЛИЦО:ЧИСЛО

Сокращение	Значение
1	1-е лицо
2	2-е лицо
3	3-е лицо
0	Категория отсутствует

### 3.5.5.3 Род

Род указывается после указания времени, от которого он отделяется знаком ( - ).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ВРЕМЯ-РОД:ЧИСЛО

<i>Сокращение</i>	<i>Значение</i>
M	Мужской род
F	Женский род
N	Средний род
GAMB	Род невозможно определить однозначно
0	Категория отсутствует

#### 3.5.5.4 Число

Число указывается после указания рода или лица, от которого оно отделяется знаком ( : ).

Афинность#ЧастьРечи|\_БелорусскоеЗаголовочноеСлово\_РусскоеЗаголовочноеСлово\_&ВРЕМЯ-ЛИЦО\_или\_РОД:**ЧИСЛО**

<i>Сокращение</i>	<i>Значение</i>
SG	Единственное число
PL	Множественное число
0	Категория отсутствует



## 4 Работа с ОК-БРСР в программе CLAN

### 4.1 Введение

ОК-БРСР можно анализировать при помощи разработанных в рамках проекта CHILDES инструментов программы CLAN. Далее будут представлены необходимые для этого приготовления, а также показаны некоторые примеры анализа, а именно посредством команд `kwal` и `freq`. Данная глава не является введением в программу CLAN или его заменой. Руководство по применению программы CLAN (McWHINNEY 2013b) доступно на странице <http://childes.psy.cmu.edu/manuals/CLAN.pdf>

### 4.2 Подготовка корпуса

*Установка программы CLAN:*

- Установите CLAN. Файл для установки вместе с инструкцией можно найти на сайте <http://childes.psy.cmu.edu/clan/>

*Подготовка ОК-БРСР:*

- Со страницы <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/> скачайте интересующие Вас протоколы (например, папку *ok-wrgr\_fam\_alles.zip*, содержащую все разговоры семейного корпуса). Распакуйте скачанную папку / скачанные папки.
- Со страницы <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/> скачайте файл *depfile.cut*, это контрольный файл программы CLAN. В нем определено, какая информация (сокращения и т.д.) может содержаться в каких строках протокола. Версия файла *depfile.cut*, которую можно скачать из ОК-БРСР, представляет собой расширенную версию соответствующего файла, доступного для скачивания на сайте CHILDES, т.к. в ней учтены особенности транскрипции ОК-БРСР:

- Строка **@Participants:** содержит дополнительно следующую информацию:  
Village Migrant Town na Other
- Строка **\*:**, отвечающая за строки с высказываниями в протоколах, содержит дополнительно следующую информацию:  
[-] [\_/] [\_//] [\_///] [\_////] ["] ["-"] \*@r \*@h  
\*@pb \*@pc \*@ph \*@dial \*@pros \*@w \*@ls \*-\*
- Строка **%mor:** содержит дополнительно следующую информацию:  
[-] [\_/] [\_//] [\_///] [\_////] ["] ["-"] :
- Также в файле определено, что в протоколах кроме обычных строк могут содержаться и следующие строки:  
@MT of #: \* и @FL of #: \*

Сохраните файл *depfile.cut* в той же папке, что и распакованные протоколы.<sup>22</sup>

- При помощи текстового редактора или программы CLAN откройте файл *ISO-639.cut*, находящийся в папке ...\*CLAN\lib\fixes* (там же, где Вы установили программу CLAN). Файл *ISO-639.cut* содержит список сокращенных языковых обозначений, используемый программой CLAN. Вручную добавьте следующие строки (части текста должны быть разделены при помощи табуляции):

```
hb  hybe  HybBel
yr  hyru  HybRus
ru  rus   Rus
hy  hyb   Hyb
na  nan   NoAnswer
brms brms  BelarusianRussianMixedSpeech
```

---

22 К моменту публикации ОК-БРСР в CLAN уже не было возможности обратиться к файлу *depadd.cut*, используемому для дополнительного определения строк протоколов. Поэтому должны быть внесены необходимые изменения в файл *depfile.cut*.

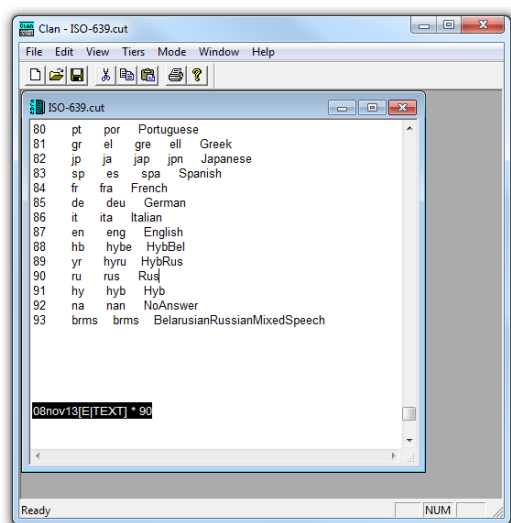


Рис. 4 Файл *ISO-639.cut*

Возможен также и другой вариант: со страницы <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgt/> скачайте файл *ISO-639.cut*. Этот файл идентичен файлу, возникающему при установке CLAN, однако он содержит упомянутые выше используемые в ОК-БРСР языковые обозначения. Замените уже существующий файл *ISO-639.cut* в папке ...\*CLAN*\lib\fixes на файл *ISO-639.cut*, скачанный Вами.

- После запуска CLAN установите папку с протоколами разговоров в качестве рабочей директории (*working directory*).

### 4.3 Поиск словоформ

Т.к. после транскрипции словоформ при помощи тегов (tags) указана их языковая афинность, то в поисковом запросе должен использоваться символ подстановки ( \* ). Так, например, запрос `freq +s"tak"` не найдет ни одной словоформы.

*Пример поиска:*

```
freq fam_sm_9.cha +s"tak*"
```

*Результат поиска:*

Speaker: \*SFA:

2 tak@c@pc

1 takojê@c@pc

-----

2 Total number of different item types used

3 Total number of items (tokens)

0.667 Type/Token ratio

Знак ( @ ) можно использовать в поисковом запросе для обозначения конца словоформы:

*Пример поиска:*

```
freq fam_sm_9.cha +s"tak@"
```

*Результат поиска:*

Speaker: \*SFA:

2 tak@c@pc

-----

1 Total number of different item types used

2 Total number of items (tokens)

0.500 Type/Token ratio

#### **4.4 Поиск аннотаций в строке %mor**

Представленная в главе 3.5 структура аннотаций в строке %mor дает возможность использовать функцию поиска программы CLAN для поиска грамматических показателей или словоформ с определенной языковой афинностью. Так, часть речи можно найти через *part of speech*, заголовочное слово через *stem*, языковую афинность через *prefix*, падеж и время через *suffix*, род, лицо и число через *fusion*. На следующем изображении показано, как искать словоформы с ‚русской‘ афинностью и ‚белорусским‘ произношением:

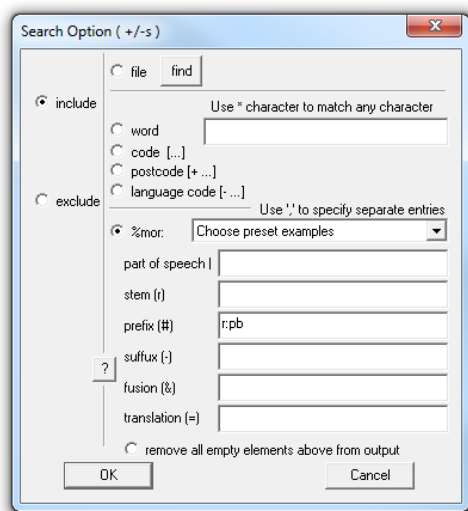


Рис. 5 Поиск ,русских' словоформ с ,белорусским' произношением при помощи функции поиска программы CLAN

Соответствующий поисковый запрос в командном окне программы CLAN выглядит следующим образом (для команды **freq**):

```
freq *.cha +s"@#-r:pb"
```

Данные поисковые возможности можно также комбинировать. Например, следующий поисковый запрос найдет все наречия с ,русской' афинностью:

```
freq *.cha +s"@|-adv,#-r:*
```

#### 4.5 Поиск знаков препинания

Можно искать также и знаки препинания. Так, следующий поисковый запрос найдет все повторения (для команды **kwal**):

```
kwal *.cha +s'[_/]'
```

Здесь тоже можно использовать символы подстановки. Следующий запрос найдет повторения, исправления, переформулировки и переводы:

```
kwal @ +s'[_/*]'
```

#### 4.6 Ограничение поиска по отдельным частям корпуса / городам / миграционному статусу

Называние файлов (ср. 3.1) дает возможность ограничить поиск определенной частью корпуса. Например, следующий запрос следует использовать для поиска только в разговорах корпуса интервью:

```
freq int*.cha +s"@|-adv,#-r:pc"
```

А следующий запрос используется для поиска только в разговорах из города *Октябрьский*:

```
freq *ak*.cha +s"@|-adv,#-r:pc"
```

Также возможно ограничение поиска говорящими с определенным миграционным статусом. Следующий запрос следует использовать для поиска только в высказываниях мигрантов из деревни в город:

```
freq *.cha +s"@|-adv,#-r:pc" +t#Migrant
```

Другие поисковые возможности описаны в руководстве по CLAN (McWhinney 2013b).

## 5 Список літературы

- БРС 2003: *Беларуска-рускі слоўнік (у 3 т.)*. Мінск
- ГЕНЧЭЛЬ, Г. 2013: Размеркаванне беларускіх і рускіх варыянтаў структурных слоў у беларуска-рускім мяшаным маўленні. // *Веснік БДУ*. 2/2013, 52-61
- ЛИСКОВЕЦ, И. В. 2005: *Русский и белорусский языки в Минске: проблемы билингвизма и отношения к языку*. Диссертация ... кандидата филологических наук. Европейский университет, Санкт-Петербург
- МЕЧКОВСКАЯ, Н. Б. 2006: Беларуская трасянка і украінскі суржык: суррогаты этнічнага субстандарта в іх адношэннях к масовай культуры і літаратурным мовам. // *Проблемы зіставной семантыкі* 7, 109-115
- МЯЧКОЎСКАЯ, Н. Б. 2007: Трасянка ў кантынууме беларуска-рускіх ідыялектаў: Хто і калі размаўляе на трасянцы? // *Веснік БДУ*. Сер. 4 (2007), 91-97
- РБС 2005: *Русско-белорусский словарь (в 3 т.)*. Мінск
- ТСБМ 1977-1984: *Тлумачальны слоўнік беларускай мовы (у 5 т.)*. Мінск
- ХЕНТШЕЛЬ, Г. 2013: Беларускі, рускі і беларуска-руская смешанная мова. // *Вопросы языкознания* (2013), 53-76
- ХЕНТШЕЛЬ, Г. 2015: Беларуско-руская смешанная мова («трасянка»): восем пытанняў і адказаў. // *Языковой контакт. Сборник научных статей*. Мінск, 172-186
- ХЕНШЕЛЬ, Г. & ЦЕЛЛЕР, Я. П. 2011: Якань, екань, ікань в беларуска-рускай смешаннай мове: назіранні на аснове эксперыментальна-акустычнага аналізу. // Шевченко, Г. (ред.), *Актуальныя праблемы філалогіі: антычная культура і славянскі свет*. Мінск, 228-234
- ЦЫХУН, Н. 2013 [1998]: Трасянка як аб'ект лінгвістычнага даследавання. // ЦЫХУН, Н. 2013: *Studien zur „Trasjanka“ – Штудыі пра „трасянку“*. Herausgegeben von Gerd Hentschel. Oldenburg, 19-30 [= *Studia Slavica Oldenburgensia* 22]
- HENTSCHEL, G. 2013: Zwischen Variabilität und Regularität, „Chaos“ und Usus: Zu Lautung und Lexik der weißrussisch-russischen gemischten Rede. In: G. Hentschel (Hrsg.): *Variation und Stabilität in Kontaktvarietäten: Beobachtungen zu gemischten Formen der*

*Rede in Weißrussland, der Ukraine und Schlesien.* (= *Studia Slavica Oldenburgensia* 21), Oldenburg, 63-99

- HENTSCHEL, G. & KITTEL, B. 2011a: Zur weißrussisch-russischen Zweisprachigkeit in Weißrussland – nicht zuletzt aus der Sicht der Weißrussen. // Bohn, T. (Hrsg.): *Ein weißer Fleck in Europa. Die Imagination der Belarus als Kontaktzone zwischen Ost und West*. Bielefeld, 49-67
- HENTSCHEL, G. & KITTEL, B. 2011b: Weißrussische Dreisprachigkeit? Zur sprachlichen Situation in Weißrussland auf der Basis von Urteilen von Weißrussen über die Verbreitung „ihrer Sprachen“ im Lande. // *Wiener Slawistischer Almanach* 67, 107-135
- HENTSCHEL, G., TARANENKO, O. & ZAPRUDSKI, S. (Hrg.) 2014 *Trasjanka und Suržyk – gemischte weißrussisch-russische und ukrainisch-russische Rede. Sprachlicher Inzest in Weißrussland und der Ukraine?* Frankfurt/M.
- HENTSCHEL, G. & ZELLER, J. P. 2013: Gemischte Rede, gemischter Diskurs, Sprechertypen: Weißrussisch, Russisch und gemischte Rede in der Kommunikation weißrussischer Familien. // *Wiener Slawistischer Almanach* 70 (2012), 127-155
- HENTSCHEL, G. & ZELLER, J. P. 2014: Belarusians' pronunciation: Belarusian or Russian? Evidence from Belarusian-Russian mixed speech. // *Russian Linguistics* 38, 229-255. DOI: 10.1007/s11185-014-9126-1 [<http://link.springer.com/article/10.1007/s11185-014-9126-1>]
- HENTSCHEL, G. 2008: Zur weißrussisch-russischen Hybridität in der weißrussischen „Trasjanka“. // Kosta, P. & Weiss, D. (Hrsg.): *Slavistische Linguistik 2006/2007*. [= Slavistische Beiträge 464], 169-219
- KITTEL, B. & LINDNER, D. 2011: Der soziale Hintergrund von Sprachwahlen in Belarus. Eine sprachsoziologische Analyse der „gemischten Rede“. // *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 63, 623-647
- KITTEL, B., LINDNER, D., TESCH, S. & HENTSCHEL, G. 2010: Mixed Language Usage in Belarus. The Sociostructural Background of Language Choice. // *International Journal of the Sociology of Language* 206, 47-71
- MACWHINNEY, B. 2013a: *The CHILDES Project. Tools for Analyzing Talk – Electronic Edition. Part 1: The CHAT Transcription Format*. Car-



- negie Mellon University [<http://childes.psy.cmu.edu/manuals/chat.pdf>]
- MACWHINNEY, B. 2013b: *The CHILDES Project. Tools for Analyzing Talk – Electronic Edition. Part 2: The CLAN Programs*. Carnegie Mellon University [<http://childes.psy.cmu.edu/manuals/clan.pdf>]
- RAMZA, T. 2008: Die Evolution der Trasjanka in literarischen Texten. // *Zeitschrift für Slawistik* 53, 305–325
- ZELLER, J. P. 2015: *Phonische Variation in der weißrussischen „Trasjanka“: Sprachwandel und Sprachwechsel im weißrussisch-russischen Sprachkontakt*. Oldenburg [= *Studia Slavica Oldenburgensia*; 27]