

Министерство науки и высшего образования Российской Федерации

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТОМСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

САЕ «Институт человека цифровой эпохи»

Автономная магистерская программа

«Компьютерная и когнитивная лингвистика»

ДОПУСТИТЬ К ЗАЩИТЕ В ГЭК

Руководитель ООП

д-р филол. наук, профессор

З. И. Резанова

«18» июня 2020 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

ОТКЛОНЕНИЯ ОТ РЕЧЕВОГО СТАНДАРТА
В РУССКОЙ УСТНОЙ РЕЧИ ШОРСКО-РУССКИХ БИЛИНГВОВ:
К ПРОБЛЕМЕ ИНТЕРФЕРЕНТНЫХ ВЛИЯНИЙ

по основной образовательной программе подготовки магистров
направление подготовки 45.04.03 – Фундаментальная и прикладная
лингвистика

Степаненко Елена Борисовна

Научный руководитель,

д-р. филол. наук, профессор

З.И. Резанова

подпись

«18» июня 2020 г.

Автор работы

студент группы № 131883

Е.Б. Степаненко

подпись

Томск-2020

Министерство науки и высшего образования Российской Федерации
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (ТГУ)
САЕ «Институт человека цифровой эпохи»
Автономная магистерская программа «Компьютерная и когнитивная
лингвистика»

УТВЕРЖДАЮ
Руководитель АМП
доктор филол. наук, профессор,
зав.кафедрой

_____ З.И. Резанова
«_15_»_сентября_2019г.

ЗАДАНИЕ

по подготовке магистерской диссертации
студенту Степаненко Елене Борисовне, группы 131883

1. Тема магистерской диссертации: Отклонения от речевого стандарта в русской устной речи шорско-русских билингвов: к проблеме интерферентных влияний

2. Срок сдачи студентом выполненной магистерской диссертации:

а) на кафедре – 24.06.2020

б) в ГЭК – 28.06.2020

3. Исходные данные к работе (*цели и задачи исследования, объект и предмет, методы исследования*)

Объект исследования – устная форма русской речи билингвов, родным языком которых является шорский язык.

Предмет исследования – взаимодействие русского и шорского языков, проявляющееся в интерференции на лексическом и грамматическом уровнях.

Интерференция – «случаи отклонения от норм каждого языка, происходящие в речи билингвов в результате их знакомства с более чем одним языком» (У. Вайнрайх, «Языковые контакты», 1979 г.).

Цель работы – выявить основные элементы и типы интерференции русского и шорского языков, проявляющиеся в речи билингвов, а также условия и факторы, оказывающие влияние на существование интерферентных явлений в русской речи билингвов.

Для достижения цели поставлены следующие задачи:

1) Изучить теоретические основы и практику создания специализированных мультимодальных корпусов и корпусов неродной речи.

2) Изучить теоретические основания и практические решения создаваемого корпуса устной русской речи тюркско-русских билингвов (RuTuVic).

3) Изучить теоретические основания и принципы отклонений от речевого стандарта, разработанные в проекте создания корпуса.

4) Выявить отклонения от речевого стандарта в русской речи шорско-русских билингвов на грамматическом и лексическом уровнях

Материал исследования

Записи устной речи Расшифрованные записи корпуса русско-тюркской билингвальной речи Южной Сибири (RuTuViC): шорско-русские билингвы (шифры: ЧМН, ИМА, СОЕ), собранные в период экспедиций в период с 2017 по 2019 гг., длительностью 6 часов.

Теоретические основы исследования

Цель и задачи решаются на основе привлечения теоретических положений

- корпусной лингвистики
- теории интерференции – в сфере грамматики и лексики
- трудов в области русской лексикологии и грамматики

В том числе по русским сибирским говорам Южной Сибири, русской разговорной литературной речи

Методы исследования

Методы структурной лингвистики – элементы дистрибутивного метода.

Методы корпусной лингвистики – разметка отклонений от речевого стандарта.

4. Краткое содержание работы. Работа состоит из введения; двух глав, состоящих из 6 параграфов; заключения; списка использованных источников из 50 наименований; двух приложений; одной таблицы и четырех рисунков.

Введение содержит информацию об объекте и предмете исследования, о его цели и задачах, необходимых для её достижения, об актуальности и новизне работы, о её практической и теоретической значимости.

Первая глава посвящена истории и развитию корпусной лингвистики, ее становлению как науки. Проводится типологизация существующих корпусов.

Во второй главе отражены грамматические особенности тюркских языков и их возможное влияние на русский. Осуществляется анализ отклонения от речевого стандарта устной речи в текстах шорско-русских билингвов.

В заключении приводится общий итог работы.

5. Указать предприятие, организацию по заданию которого выполняется работа: Томский государственный университет.

6. Перечень графического материала: итоговая схема базы данных, графический интерфейс корпуса, пример аннотирования фрагментов текстов билингвов, частота отклонений от речевого стандарта в устной речи шорско-русских билингвов.

7. Дата выдачи задания «_15_»_сентября_2019 г.

Руководитель магистерской диссертации

доктор филол. наук,

профессор,

зав.кафедрой общего,

славяно-русского языкознания

и классической филологии



З.И. Резанова

Задание принял к исполнению



Е.Б. Степаненко

АННОТАЦИЯ

Данная работа посвящена выявлению интерференции в устной разговорной речи тюркских билингвов (шорский) на русский язык. Магистерская диссертация основывается на общих принципах и методах корпусной лингвистики и интерференции родного языка на русский на фонетическом, лексическом и синтаксическом уровнях. Цель данного исследования – выявить основные элементы и типы интерференции русского и шорского языков, проявляющиеся в речи билингвов, а также условия и факторы, оказывающие влияние на существование интерферентных явлений в русской речи билингвов. В результате проведенного исследования решены следующие задачи:

1) изучены теоретические основы и практику создания специализированных мультимодальных корпусов и корпусов неродной речи.

2) изучены теоретические основания и практические решения создаваемого корпуса устной русской речи тюркско-русских билингвов (RuTuVic).

3) изучены теоретические основания и принципы отклонений от речевого стандарта, разработанные в проекте создания корпуса.

4) выявлены отклонения от речевого стандарта в русской речи шорско-русских билингвов на грамматическом и лексическом уровнях

Оглавление

ВВЕДЕНИЕ	8
Глава 1.	11
Теоретические аспекты корпусного исследования речи билингвов.....	11
1.1 Место корпусной лингвистики в современной лингвистической парадигме	11
1.2 Место корпусов билингвальной речи в структуре современной корпусной лингвистики.....	19
1.1.1 История корпусной лингвистики. Направления корпусных исследований в России и за рубежом.	20
1.1.2 Типология корпусов	23
1.1.3 Особенности би- и полимодальных корпусов	27
1.2.4 Особенности корпусов неродной речи	31
2. Корпус речи тюркско-русских билингвов (RuTuVic).	36
2.1. Теоретические основания корпуса. Цель корпуса.	36
2.2 Структура корпуса RuTuVic	39
2.3 Особенности разметки корпуса RuTuVic	41
2.4. Особенности метаразметки корпуса RuTuVic	42
Глава 2. Отклонения от речевого стандарта в речи шорско-русских билингвов как проявление внутри и межязыковой интерференции	45
2.1. Языковая интерференция и ее типы.....	45
2.1.1 Интерференция и проблемы контактирования.....	45
2.1.2 Отношения индивида к языкам контактирования как фактор интерференции	46
2.1.3 Межгрупповое взаимодействие как фактор интерференции	46
2.1.4 Типы интерференции по отнесенности к уровням языковой системы	47
2.2. Межязыковая и внутриязыковая интерференция.....	52
2.2.1 Особенности языковых систем – источников внутриязыковой и межязыковой интерференции в речи шорско-русских билингвов.....	52

2.3 Отклонения от речевого стандарта в речи шорско-русских билингвов в аспекте интерферентных влияний.....	69
2.3.1 Отклонения от речевого стандарта в фонетике	70
2.3.2 Отклонения от речевого стандарта в грамматике	74
Заключение	83
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ И ЛИТЕРАТУРЫ.....	85
Приложение А	92
Приложение Б	93
Продолжение Приложения Б	94

ВВЕДЕНИЕ

Современная лингвистика, в частности, формальная, имеет тесную методологическую связь с исследованиями в области корпусов. Подобная интегративная тенденция технологий и методов обусловлена стремительным развитием информационных технологий, влекущим за собой упрощение и автоматизацию вычислительных возможностей в поле лингвистических исследований. Цель изучения языков детерминирует выбор соответствующих инструментов, подходов и методов создания корпусов и исследования языка как компонента языкового взаимодействия разных народов [13, 14] и представления языковой картины мира как целого языка [33], так и субъективного отражения языка в определенной этнической культуре [19, 25], направленных на практическое изучение языка, его действия в существующих лингвистических материалах.

Актуальность исследования

Важным вектором языкового развития сегодня является изучение в мультимедийных, мультимодальных текстовых корпусов или текстов устной речи, где обозначена транскрипция текстов устной речи (*это аудиозаписи, видеозаписи, диалоговая, монологическая речь и др.*). Мультимодальные корпусы перспективны, так как при их помощи происходит взаимодействие языковых и неязыковых элементов речи процессе речевых диалогов. Эти корпусы призваны решить различные лингвистические и экстралингвистические цели и задачи, к примеру, фиксация речевых явлений фонетики, грамматики изучаемого языка, а также взаимодействие между вербальными и невербальными средствами общения: речью говорящего и мимикой, жестами, телодвижениями и т.п.

Не менее важным направлением в современной корпусной лингвистике является проблема влияния материнских языков на речевые практики русскоязычных билингов. Существует ряд проектов [2, 5, 7], изучающих интерференцию родных языков, функционирующих в субъектах Российской Федерации.

Работа тесно связана с созданием корпуса русско-тюркской билингвальной речи Южной Сибири (RuTuBiC). Актуальность работы заключается в том, что небольшие исследования речи билингвов являются составляющей частью большой исследовательской работы по изучению эволюции языков малых народов Южной Сибири, а также их оцифровке и сохранению.

Создание корпуса русско-тюркской билингвальной речи Южной Сибири стало возможным благодаря развитию теории практики корпусной лингвистики.

Объект исследования – устная форма русской речи билингвов, родным языком которых является шорский язык.

Предмет исследования – взаимодействие русского и шорского языков, проявляющееся в интерференции на лексическом и грамматическом уровнях..

Интерференция это есть «случаи отклонения от норм каждого языка, происходящие в речи билингвов в результате их знакомства с более чем одним языком [12].

Цель работы – выявить основные элементы и типы интерференции русского и шорского языков, проявляющиеся в речи билингвов, а также условия и факторы, оказывающие влияние на существование интерферентных явлений в русской речи билингвов.

Для достижения цели поставлены следующие задачи:

1) изучить теоретические основы и практику создания специализированных мультимодальных корпусов и корпусов неродной речи.

2) изучить теоретические основания и практические решения создаваемого корпуса устной русской речи тюркско-русских билингвов (RuTuBiC).

3) изучить теоретические основания и принципы отклонений от речевого стандарта, разработанные в проекте создания корпуса.

4) выявить отклонения от речевого стандарта в русской речи шорско-русских билингвов на грамматическом и лексическом уровнях.

Материал исследования

Расшифрованные записи корпуса русско-тюркской билингвальной речи Южной Сибири (RuTuViC): шорско-русские билингвы (шифры: ЧМН, ИМА, СОЕ), собранные в период экспедиций в период с 2017 по 2019гг., длительностью 6 часов.

Теоретические основы исследования

Цель и задачи решаются на сове привлечения теоретических положений - корпусной лингвистики. Теоретические аспекты корпусной лингвистики: А.А. Кибрик, Ю.Д. Апресян, В.А. Плунгян, В.П. Захаров.

- теории интерференции – в сфере грамматики и лексики: А.В. Дыбо, З.И. Резанова, Е.В. Рахилина, A. Ralston, M.A. Daniel.

- труды в области русской лексикологии и грамматики: Э.Р. Тенишев, А.С. Выренкова, Н.П. Дыренкова, Н.З. Гаджиева.

- по русским сибирским говорам Южной Сибири, русской разговорной литературной речи

Методы исследования

Методы структурной лингвистики – элементы дистрибутивного метода.

Методы корпусной лингвистики – разметка отклонений от речевого стандарта.

Новизна исследования

В лингвистический оборот вносится новый эмпирический материал – расшифровка устной речи шорско-русских билингвов.

Впервые системно описаны отклонения от речевого стандарта в речи шорско-русских билингвов.

Структура работы. Работа состоит из введения; двух глав, состоящих из 6 параграфов; заключения; списка использованных источников из 63 наименований; двух приложений; одной таблицы и четырех рисунков.

Глава 1.

Теоретические аспекты корпусного исследования речи билингвов.

1.1 Место корпусной лингвистики в современной лингвистической парадигме

Становление корпусной лингвистики способствовало развитию новых активных инструментов анализа разных языковых явлений и изменению в некотором роде теоретических воззрений на само понятие языка, его систему и, соответственно, на основные вопросы по его дальнейшему изучению и исследованию.

В языковой практике всего мира определены контуры новой структуры языка, у неё есть отличия от предыдущих моделей, существовавших в конце двадцатого столетия.

Как это ни парадоксально, исторически сложилось так, что теоретическое развитие языка позволило абстрагироваться от различий, обозначившихся в середине 20 века определённых научных школ, и обнаружить общие теоретические тенденции противоположных во взглядах лингвистических направлений, например, структурной и генеративной школ. Именно поэтому стало возможным «...в некоторых случаях апеллировать к некой обобщенной лингвистической идеологии «середины XX века», противопоставленной идеологии нынешнего рубежа столетий (следует подчеркнуть, однако, что в таких случаях всегда имеется в виду именно господствующая или преобладающая идеология, поскольку «теоретические диссиденты» встречались на каждом этапе существования лингвистики - и иногда их взгляды становились преобладающими позднее)...» [32, С. 8]

Как отмечает В.А. Плунгян, основной важной отличительной особенностью современной языковой модели от предыдущей является научно обоснованный переход понятий от «языка» к «речи» и от «системы» к «узусу». Современные лингвисты стремятся изучать языковые факты, а не

конструкции, исследовать, размышлять о языке как природном явлении, а не о свойствах его структурных моделей. Но это вызывает трудности, так как язык в некотором смысле абстрактен, прямое наблюдение за ним невозможно. Наблюдать можно за речью и порождаемым ею текстом.

Это и есть новое направление в науке о языке, которое сформировалось в конце 20 и начале 21 века. Учеными назван «список идеологических предпочтений», определяющий новые цели и задачи современной лингвистической теории: 1) активное исследование дискурса как инструмента коммуникации (в противопоставлении исследованию слов и предложений как языковых единиц), 2) выявление количественных отношений в языке; 3) описание синхронной вариативности языка, ее обусловленности психологическими, биологическими и социальными факторами; описание синхронной и диахронной вариативности и изменчивости языка в их взаимодействии; 4) анализ языковой нормы и языковой правильности благодаря появлению новых теоретических взглядов. («Граница между «ошибкой» и «маргинальным вариантом», а также между маргинальным вариантом и полноценным (а впоследствии, может быть, и единственным) признается гораздо более подвижной и зыбкой» [32, С. 2]); 5) формирование подхода, ориентированного на узус.

Исследователи отмечают, что сегодня корпусная лингвистика перевернула представление о языке, подчеркивая значимость технических средств, возможностей программирования, которые способствуют быстрому и, вместе с тем, эффективному доступу к достаточно значительной по объёму совокупности текстов, сегодня эти возможности реально существуют и способствуют более детальному и углублённому изучению языков.

Объём текстов любого живого языка на много порядков превосходит возможности одного учёного-исследователя. Исследователям, научная деятельность которых направлена на анализ большого массива данных с целью структурирования теоретических моделей, корпус - основной инструмент, способный в несколько раз повысить качество, надёжность и

скорость исследований в области языков. Исследователи, владеющие техническими инструментами, способны быстрее и эффективнее исследовать массивы текстового материала. Раньше на обработку того же материала вручную, на создание картотек уходили месяцы, годы. С появлением машин ситуация меняется, и гарантия правильности анализа языковых данных становится шире и глубже [32, С. 2].

Важной особенностью теоретиков и практиков мировой корпусной лингвистики, что находит отражение и в современной российской лингвистике, является принцип «нелитературоцентричности». Как отмечает, В.А. Плуноян, «лингвистический корпус вернул исследователям главное – объёмные тексты естественного языка» [32, С. 2].

Поэтому в современной теоретической науке о языке создание и изучение лингвистического корпуса – это не только важное и глобальное средство изучения и описания языковой теории и практического применения теории, но и новый взгляд, можно сказать, новое мировоззрение, которое направило лингвистов рассматривать текст в качестве основного объекта теоретических изысканий.

Современные исследователи-лингвисты считают, что фонетическое грамматическое изучение языков и их анализ в скором времени станет «корпусно-ориентированным» - и это значит, что изложенные о конкретном языке факты, представленные или заявленные учёными, можно проверить при помощи существующих данных корпуса исследуемого языка. Исследовательский статус корпусной грамматики станет больше, так как произойдёт увеличение проверяемых наблюдением фактов и текстовых файлов резко вырастет в сравнении с нынешним имеющимся речевым материалом, где теория языка в большинстве случаев ограничивается небольшим количеством приведённых примеров, выделенных из лингвистического материала: иллюстрация явления, а именно того, что оно существует в данном языке, но не может быть применено с целью доказать частотность употребления какой-то особенности в речи говорящего. Также в

этом случае невозможно изучение искусственно созданных исследователем конструктивных языковых высказываний.

Также тесным образом связана с использованием лингвистического корпуса, а именно с положениями «иметься в корпусе» и «иметься в данном языке», которые не одинаковы в полном объёме - это проблема, которая призвана доказать присутствие того или иного не часто повторяющегося явления в языке. Всегда существовали и существуют в языке потенциально возможные явления, и не всегда их теоретическое описание есть в определённом корпусе лингвистических материалов. Но в объёмном, большом, представительном корпусе отсутствие некоторого языкового явления (хотя и предполагаемого или реально существующего) имеет большое значение. Кроме того, в лингвистическом корпусе можно определить окказионализмы или ошибки, не принятые говорящими на данном языке при показе соответствующих языковых фреймов с целью оценить их правильность или уместность употребления. Активность и частота таких проявлений, по мнению учёных, если встретятся так называемые не часто употребляемые в речи окказиональные слова – не очень большая. В том случае, если языковая «ошибка» определяется часто, то, по всей видимости, это не уже не «ошибка, нарушение», а из этого следует, что важно тщательнее исследовать возможные формы использования изучаемой языковой единицы и после уже устанавливать её статус: «норма» или нет. Таким образом, при применении корпусной лингвистики, учёные-лингвисты не имеют возможности основываться на собственной интроспекции, и, следовательно, выдавать желаемое за реальное и заниматься описанием не самого языка, а своих собственных убеждений и представлений о том, каковым он должен существовать.

Можно, «отождествить понятия «существующего в языке» и «надежно засвидетельствованного в корпусе»» [32, С. 16]. В реально имеющемся делении системы и узуса лингвистический корпус помогает учёным концентрироваться на языковом материале в связи с тем, что прежде, до

времен становления корпусной лингвистики, языковой материал не был достаточно глубоко, а, самое важное, объективно изучен. И тем не менее, именно узус является подлинным и главным доказательством исследований о естественном языке, т. е. объектом, эмпирически доступным любому исследователю.

Положения и исследуемые текстовые материалы лингвистических корпусов дают возможность понимания того, что при равных условиях, преимущество отдано описанию языка, распространённого на явления, объёмнее и глубже исследованные и обозначенные в языковом корпусе. Таким образом, явление, являющееся основным, главенствующим в корпусе, определяется языковой грамматикой. Смыслы слов, не являющиеся необходимыми для реальных говорящих, владеющих данным языком, не должны подлежать их описанию. Они не столь важны для реального языка. И бесполезность их исследования ясна. Полезнее, по мнению исследователей, вначале остановиться на правилах и способах выражения смыслов, являющихся основными и главными для существующего языкового коллектива людей: а именно, легко обнаруживаемые в речи носителей и описанные, отражённые в соответствующем языковом корпусе. «Язык, прежде всего, дает нам возможность выразить и понять то, что уже было многократно сказано в данном национальном пространстве; может ли язык, без существенного насилия над его структурой и механизмами, быть успешно использован как-то еще - вопрос, вообще говоря, открытый. Может быть, и может - и если так, то такие смещенные способы использования языка тоже должны быть описаны» [32, С. 17]. Важны здесь и основные направления, а корпус, по мнению исследователей, способен описать и распределить эти направления достаточно правильно. Корпусно-ориентированная лингвистика способствует исследованию реально представленных в языке явлений, а не искусственному их описанию и созданию.

Реально существующих в языковых материалах различных грамматических средств в некотором смысле менее, нежели в

лингвистических литературных нормированных источниках. Используемые системы автоматизированной морфологической обработки по большей части не могут определить частоту употребления той или иной единицы языка: если единица реально присутствует в текстах, то существует она и в автоматическом словаре. В результате всего этого появляются в большом количестве формы, не встречающиеся в связи с существующими нормами и правилами морфологии в естественных текстах лингвистических материалах, но существующих в грамматике родного языка. Возникшее «гиперпорождение» специалистам по автоматическому морфологическому анализу только мешает, являясь так называемым «эффектом паразитической омонимии», затрудняя и «сбивая с толку» автоматические программы. Проявляется он в том случае, когда у имеющейся распространенной словоформы определяется схожая пара, существующая формально, но тем не менее определяемая в естественном тексте. *К примеру, морфологический разбор союза «хотя», который образован от деепричастия «хотя», а «хотя», соответственно, от глагола «хотеть» морфолого-синтаксическим способом. Или анализ формы презенса «сеем, решаем» в качестве краткой формы страдательного причастия сеемый, решаемый, так как полная форма данного причастия встречается в корпусе один единственный раз; является проблемой и существование формы множественного числа у слов, к примеру, ежедневность, белила и т.д.*

Как писал В.А. Плунгян: «грамматика, имеющаяся в данных корпуса, можно сказать, и больше, и меньше реальной, обычной, потому что может включать в себя «в качестве легитимных многие не предусматриваемые стандартной грамматикой варианты» [32, С. 19], но и, что является плюсом, реальную возможность объяснения «потенциальных», существующих лишь в теории, форм. Описанная корпусом грамматика русского языка отличается уже сегодня от существующих аналогичных исследований, описанных в так называемую «докорпусную» эпоху.

Идеи корпусной лингвистики были определены и обозначены как основные ещё в 60-е годы XX века, а активно развиваться корпуса начали в 80-е годы. Связано это было с новой технической революцией и сопутствующим ей развивающимся непрерывно информационным технологиям, программированию и компьютерным вычислительным машинам. Сегодня существуют размеченные на основе теоретических положений, обоснованные существенными принципами корпуса текстов языков, которые являются значимыми источниками филологических исследований, не уступая, а даже успешно конкурируя с имеющимися в науке о языке словарями и всевозможными справочниками

В лингвистический корпус обычно включаются следующие виды разметок: морфологическая, лексико-семантическая и стилистическая. Последний вид разметки как основа любого лингвистического исследования находится в соотношении с различными видами лексических словарей и справочников, которые информируют о языковых единицах и описывают их квалификацию.

Вызывает интерес учёных и противопоставление принципов отбора единиц, включающихся в словарь и корпус – словоцентричность одного источника и текстоцентричность другого. Поэтому текстоцентричный корпус предоставляет большие возможности для изучения немалого текстового объёма имеющейся сегодня информации и в частности «контекстной реализации языковых единиц, их форм, вариантов» [35]. А это напрямую повышает качество лингвистических исследований.

В настоящее время в рамках реализации различных лингвистических научных программ ведётся большая работа по созданию как больших национальных проектов, так и частных лингвистических проектов. Один из частных национальных проектов – проект фронтального исследования билингвальной речи, связанной с изменениями в речи билингвов. Интенференции - активные процессы, происходящие как в русском языке, так и в других языках всего мира. Для их более точного исследования необходимы

взаимодействие и совместная работа учёных с целью определения общих и частных закономерностей, связанных с языковой эволюцией.

Россия - многонациональная страна. Поэтому именно здесь, особенно на территории Сибири, возможно изучение разных языков, установление взаимовлияния языков. В Томском государственном университете ведётся работа по исследованию интерференции в русский язык речи тюркских билингвов.

«В современной интенсивно развиваемой корпусной лингвистике все большее развитие получают корпуса текстов, собранные в качестве эмпирической основы разнообразных лингвистических проектов, что мотивирует разработку особых принципов разметки и метаразметки текстового материала, которые дополняют выработанные ранее стандарты его обработки» [35, С. 108].

Большая часть имеющихся на сегодняшний день наиболее описанных и реально обоснованных корпусов национальных языков состоит по большей части из письменной речи, и чтобы, так сказать, их поддержать, исследователи начали создавать, на сегодняшний день пока ещё не такие представительные как мультимедийные, письменные тексты, устные подкорпусы.

Теоретическая направленность представляемого корпуса русской речи билингвов, концепция, касающаяся в частности подкорпуса устной речи билингвов, считается основополагающей. При создании текстовой базы устной речи билингвов необходима максимально полная фиксация значимых психолингвистических и социолингвистических особенностей речи тюркских билингвов, и это стало необходимым при исследовании когнитивных «эффектов билингвального взаимодействия». Корпус русской речи билингвов, созданный учёными, станет первым представительным сборником устных текстов, где будет описана речь людей, говорящих по-русски, и которая находится под интерферентным влиянием родных языков тюрков, которые по причине сложившихся исторических особенностей языковой ситуации региона сейчас являются средством обыденного общения. Таким образом,

созданный корпус устной речи билингвов станет основой для дальнейших научных исследований, описывающих социолингвистические, психолингвистические и когнитивные направления, их взаимодействие и влияние на совершенствование и развитие языковой системы.

1.2 Место корпусов билингвальной речи в структуре современной корпусной лингвистики

Корпусная лингвистика – раздел языкознания, прикладная наука, но, по мнению М. Копотева, развивающаяся в новое научное направление и «предлагающая в настоящее время как новые теоретические решения, так и конкретные исследовательские и педагогические инструменты для работы с языком». М. Копотев выделяет основные приоритеты корпусной лингвистики [24, С. 32]:

- переход исследовательской стратегии с исследования нормированного языка на изучение узуса («как говорят/пишут»);
- использование квантитативных методов, позволяющих учитывать частотные характеристики исследуемых единиц, и замена интроспективных оценок материала точными количественными данными об употреблении;
- внимание к контексту в широком смысле (исследование коллокаций, ключевых слов, конструкций);
- автоматическое извлечение информации с помощью поисковых запросов, что может приводить к получению объемного, но не всегда релевантного материала;
- опора на автоматическое аннотирование, не лишенное с точки зрения традиционной лингвистики определенных неточностей и упрощений;
- распространенность «формально-морфологического» подхода, при котором поиск примеров часто основывается на морфологической (или просто на буквенной) форме.

В.А. Плуноян даёт следующее определение корпусной лингвистике: «...это наука о том, как сделать хорошую разметку (аннотацию) текста.... Корпус – это...собрание текстов на данном языке, представленное в электронной форме и снабжённое научным аппаратом (разметкой)» [32, С. 6]

Несмотря на различия в определениях, можно выделить универсальные характерные черты для лингвистических корпусов: 1) корпус – совокупность текстов; 2) определённая цель и методология создания; 3) репрезентативность; 4) сбалансированность; 5) машиночитаемость; 6) наличие разметки и/или мета-информации.

1.1.1 История корпусной лингвистики. Направления корпусных исследований в России и за рубежом.

Впервые методология корпусной лингвистики зародилась на рубежах конца XIX - начала XX с целью сопоставления и анализа контекстов. В указанном периоде преимуществами обладал так называемый рациональный подход, основанный «на лингвистической интуиции, проводящей различие между правильными и неправильными конструкциями» [47, С. 14]. Этому взгляду был противопоставлен другой, связанный с языковым наблюдением и изучающий язык в качестве «ресурса, обеспечивающего набор возможности для коммуникации» [47, С. 14]. Формирование научных взглядов стало основой и началом развития современных корпусных исследований. Не менее важным ключевым «идеологом» развития современной корпусной лингвистики был Захаров В.П. В своем учебнике «Корпусная лингвистика» он описывает ключевые технологии. В своих трудах [20, С. 25], он отмечает три основные области лингвистических исследований: 1) сравнительно-историческое языкознание; 2) составление грамматик и словарей и обучение языку; 3) социальная лингвистика.

С середины 19 века учёными лингвистами начаты исследования по построению диалектных карт и созданию сборников текстов и теоретических

материалов, связанных с диалектными отношениями [2, С. 26]. Соответствующие научные разработки стали основой современной корпусной лингвистики, но они обладали основным недостатком – низкой скоростью обработки анализа и обработки текстов. И лишь в середине XX века появились технические возможности решить подобную проблему, а также упростить доступ за счет интернет-технологий.

Первый лингвистический корпус (в современном понимании) появился в 1963 г. в Брауновском университете (США) под руководством исследователей У. Френсиса и Г. Кучера, который включал в себя набор текстов из газет, журналов, художественные тексты известных писателей и поэтов с целью изучения жанровых особенностей языка и создания частотного словаря. На тот период Brown Corpus включал 1 млн. слов. Разработанные технологии начали активно привлекать внимание исследователей к подобной проблематике [51]. Позднее в 1971-1978 годах, европейские ученые вели разработку Ланкастерско–Осло–Бергенского корпуса британского аналога современного английского языка (*The Lancaster–Oslo/Bergen Corpus of British English, LOB*) [52]. Во времена с начала создания в нём было около пятисот текстов, относящихся к разным жанрам, а объём составил около 2000 слов.

Вышеуказанные корпуса были ориентированы на изучение определенных жанров (специализированные) и неразмеченные. Однако развитие информационных технологий, способных работать «дешевле и быстрее» позволило изменить объект исследований. В 90-е годы XX века стали развиваться, так называемые, национальные корпуса. Сегодня более значимым считается Британский национальный корпус (*British National Corpus, BNC*), созданный и записанный в 1991-1994 гг. учёными Оксфордского Университета, а также Университета Ланкастер. Более ста миллионов слов составляет общий объем корпуса, он намного массивнее, чем предыдущие. Составляющая часть корпуса тоже изменилась:

девяносто процентов корпуса содержит письменные языковые материалы и всего десять процентов составляют устные тексты. Языковые материалы относятся к концу двадцатого века и представлены они разными речевыми жанрами, дискурсами: газетная статья, научно-популярная литература, деловая переписка, статьи и материалы, включающие религиозную тематику, транскрипция записей неформальной речи, шоу, правительственная речь и т. д [51].

Необходимо отметить, что Британский корпус стал первым национальным корпусом. Плунгян В.А. пишет «для британцев слово «национальный» значило в своей основе «характеризующий британский национальный вариант английского языка», он по разным стилевым и жанровым признакам, лексическому составу противоречит иным языковым вариантам: американскому и австралийскому. Исследователь Плунгян В.А. делает вывод о том, что национальный корпус должен быть не только большим по объему, но и должен иметь тексты самых разных жанров, многообразных в определённую историческую эпоху, и при этом важно соотносить их в нужной, необходимой пропорции [32, С. 7].

Взяв пример у Британского Национального корпуса, страны всего мира осуществляют сегодня разработку своих национальных корпусов. Так уже созданы корпуса французского, немецкого, испанского языков. В России активные работы по созданию корпусов начались на базе Института русского языка им. В. В. Виноградова РАН в 1985г. К началу 1990-х годов были собраны коллекции документов художественной литературы и программа для их обработки «UNILEX». Однако проект вскоре был завершён, но на его методологической базе были разработаны в начале 2000-х годов проекты БОКР (Большой корпус русского языка, руководитель — С. А. Шаров) и «Русский стандарт» (руководитель — В. А. Плунгян) [2, С. 25]. На основании полученных технологий и разработке теоретических аспектов корпусной лингвистики, в период с 2003 по 2010 гг. под руководством российской академии наук и группой лингвистов из Москвы, Петербурга, Воронежа

велась разработка русского национального корпуса. Во взаимодействии с компанией «Яндекс» в 2004 году разработали сайт «Национальный корпус русского языка». Кроме того, было принято решение о создании своих национальных корпусов, в которые сегодня выложено более 288 миллионов словоупотреблений начиная с середины 18 века до начала 21 века самых разных жанров. В языковом корпусе можно найти примеры как современных, так и исторических текстов, благодаря существованию в них текстов различных эпох. При всём этом нужно сказать, что содержащиеся в корпусе тексты соответствуют литературному, разговорному, официально-деловому и научному стилям. Кроме перечисленного, Национальный корпус русского языка содержит много разных примеров диалектов, просторечий, жаргонизмов.

С момента определения корпусов среди теоретиков-лингвистов начинается демаркация границы лингвистических исследований «докорпусной» эпохи и эпохи создания корпусов [32, С. 8]. Получили свой вектор развития и являются обсуждаемыми на специальных семинарах разных научных конференций по прикладной и компьютерной лингвистике теоретические и практические проблемы корпусов. Это и ежегодная международная конференция по компьютерной лингвистике «Диалог», конференция «Корпусная лингвистика» при кафедре математической лингвистики СПбГУ, конференция «Мегалинг». Ведется постоянная работа над созданием и исследованиями в области корпусной лингвистики и по созданию формальных стандартов разработки и развитию прикладных особенностей составления корпусов.

1.1.2 Типология корпусов

Отечественная и зарубежная корпусная лингвистика обладает сегодня большим объемом корпусов различных типов и размеров. Подобное разнообразие обуславливается различными целями и задачами

исследовательских и прикладных задач проблемной области, особенностями языкового материала. В связи с этим сегодня нет единого мнения, касающегося общепринятой системы типологии корпусов. По мнению Т. Макенри, Р. Сяо, классификация корпусов зависит от определённой исследовательской цели и задач, которые составляют основу предложенной учёными классификации: 1) общие и специализированные корпуса; 2) корпуса письменной и устной речи; 3) синхронические и диахронические корпуса; 4) ученические и мониторинговые корпуса [10, С. 67].

В классификации корпусов М. Коптев выделял следующие типы (в зависимости от цели исследования): 1. По форме хранения: – в звуковой форме, письменные, смешанные; 2. По языку представления текстов: одноязычные, многоязычные; 3. По жанровой принадлежности: литературные, диалектные, разговорные, публицистические, смешанные; 4. По способам доступа: свободно доступные, коммерческие, закрытые; 5. По назначению: исследовательские, иллюстративные; 6. По динамичности: динамические (мониторные), статические; 7. По наличию дополнительной информации: аннотированные (размеченные), неразмеченные [24, С. 31].

Исследования В.В. Рыкова позволяют выделить такую типологию лингвистических корпусов: 1. По степени организации и структурированности: электронный архив; электронная библиотека; корпус; подкорпус. 2. По хронологическому признаку: синхронический; мониторинговый; диахронический. 3. По типу разметки: простой; аннотированный. 4. По языковому признаку: одноязычный; двуязычный; многоязычный. 5. По способу применения и использования корпуса: исследовательский; иллюстративный; параллельный. Выделены типы: 1. По степени организации и структурированности: электронный архив; электронная библиотека; корпус; подкорпус. 2. По хронологическому признаку: синхронический; мониторинговый; диахронический. 3. По типу разметки: простой; аннотированный. 4. По языковому признаку: одноязычный; двуязычный; многоязычный. 5. По способу применения и

использования корпуса: исследовательский; иллюстративный; параллельный. Автором российского руководства «Корпусная лингвистика» [41] В.П. Захаровым предлагается разделение существующих корпусов на два определяющих класса:

1. Противопоставление корпусов, относящихся ко всему языку (часто к языку определенного периода), корпусам, относящимся к какому-либо подязыку (жанр, стиль, язык определенной возрастной или социальной группы, язык писателя или ученого и т.п.).

2. Разделение корпусов по типу лингвистической разметки. В зависимости от поставленных целей и классифицирующих признаков: 1) тип данных (письменные, речевые, смешанные); 2) язык текстов (русский, английский и т.д.); 3) «параллельности» (одноязычные, двуязычные, многоязычные); 4) «литературности», специфичности (литературные, диалектные, разговорные, терминологические, смешанные); 5) жанр (литературные, фольклорные, драматургические, публицистические); 6) доступность (свободно доступные, коммерческие, закрытые); 7) назначение (исследовательские, иллюстративные); 8) динамичность (динамические (мониторные), статические); 9) разметка (размеченные, неразмеченные); 10) характер разметки (морфологические, синтаксические, семантические, просодические и т.д.); 11) объем текстов (полнотекстовые, «фрагментнотекстовые»); 12) хронологический аспект (синхронические, диахронические); 13) «общность» (общие, одного писателя); 14) структура (центральные и архивные, ядерные и периферийные) [20, С. 12].

Все вышесказанные классификации так или иначе рекомендуются национальным корпусам. Однако есть корпуса, которые противопоставляются национальным и обладают специфичной разметкой и были созданы с целью решения узких лингвистических целей и задач. Это так называемые «специализированные» корпуса. Специализированный корпус представляет собой жанровый специфический корпус, целью которого является описать некий подязык. К примеру, в НКРЯ существует фрагмент письменного

подкорпуса под названием «СинТагРус» или «Синтаксически размеченный корпус русского языка». «СинТагРус» состоит из структур со снятой морфологической и синтаксической омонимией. Это означает, что каждому слову текста противопоставляется единственная морфологическая структура, а каждому предложению соответствует единственная синтаксическая структура. Наряду с морфосинтаксической разметкой корпус содержит данные о реализованных в текстах лексических функциях. Эти функции описывают идиоматичные словосочетания, компоненты которых связаны определенными семантическими отношениями. В корпусе «СинТагРус» представлено более 100 лексических функций, которым в настоящее время соответствует около 21 тысячи словосочетаний [53]. Не менее известен корпус «The Corpus of Professional Spoken American English» (CPSA) [54], состоящий из коммуникативных ситуаций академической и политической профессиональных областей, а именно из их транскриптов. В The Michigan Corpus of Academic Spoken English (MICASE) содержится около 1700000 слов [55], около 200 часов записей университетской устной современной речи. В этих типах корпусов выделены корпуса конкретно по автору или жанру к примеру, The Epistolary Corpus of Victorian Women Writers 'Letters [56], The Shakespeare Corpus [57], Корпус словаря языка Достоевского [58]. В русской лингвистике одним из наиболее известных корпусов устной речи обозначился корпус под названием «Рассказы о сновидениях и другие корпуса звучащей речи». Цель корпуса - описание и стилистический анализ живой устной речи носителей языка.

Корпуса, изучающие отклонения от речевого стандарта представляют собой особый тип специальных корпусов. Например, «Корпус русских учебных текстов» (КРУТ), который был разработан под руководством Е.В. Рахилиной [59]. Этот корпус является производным из многочисленных текстов, написанных на русском языке студентами высших учебных заведений. Объем корпуса - более 2,6 млн. слов. Тексты сопровождаются несколькими видами разметки: метатекстовая, морфологическая и разметка по

ошибкам. Всё это даёт хорошую возможность поиска данных по корпусу. Задачи, решаемые в данном исследовании, связаны со следующей гипотезой: ошибки показывают направление возможных языковых изменений. Для того, чтобы определить это направление, необходимо описать пределы вариативности для нестабильных языковых явлений.

Корпуса устной речи значительно меньше по объему, что в положительной мере сказывается на их репрезентативности, хотя трудоемкость и напряженность сбора устных данных выше по сравнению с письменными данными. Однако процесс создания корпуса сложен с лингвистической и технической стороны.

Делая выводы по обзорному анализу существующей типологии корпусной лингвистики, можно отметить то, что разные, существующие ныне корпуса не ограничены приведенными нами примерами. Средства, виды, типы, методы исследования корпусной лингвистики будут в системе расширяться, развиваться и пополняться новыми, ещё более эффективными приёмами совершенствования ИКТ, а также путём увеличения спектра лингвистических идей, поставленных и решаемых исследователями языка.

1.1.3 Особенности би- и полимодальных корпусов

Стремительные темпы развития корпусной лингвистики обусловлены широким спектром лингвистических целей и задач. Уже к концу 1990 года было известно более 600 компьютерных корпусов. Как было сказано ранее (см. 1.1.2), существуют национальные и специализированные корпуса. Каждый из них обладает определенным набором черт, характерных для корпуса: 1) цель и методология создания; 2) наличие разметки и/или мета-информации; 3) машиночитаемость; 4) репрезентативность. Именно первые два пункта сопоставляют между собой специализированные и национальные корпуса. Национальный корпус русского языка разработан, к примеру, для базовых исследований в различных областях русского языка, для изучения процессов

языковых изменений и для составления справок, важных для исследований [45]. Все тексты имеют общую лингвистическую (часть речи, род, число, падеж и т.п.) и экстралингвистическую разметку (жанр, время создания, тип общения и т.д.). К разметке относится составление к текстам и их элементам специальных меток (внешних, экстралингвистических, структурных и собственно лингвистических, которые занимаются описанием лексических, грамматических и иных характеристик единиц, составляющих языковые материалы). В большинстве национальных корпусов применяются пять типов разметок, а именно: акцентная и семантическая, метатекстовая, морфологическая (словоизменительная), синтаксическая. Такая общность разметки ограничивается ее глубиной аннотирования, [11] дает возможность проводить исследования в широком спектре диапазона исследовательских задач, однако не учитывает специфичные для того или иного направления требования разметки. Именно поэтому разрабатываются специальные корпуса, в которых создаётся специфическая разметка текстовой информации. Выбор «нестандартной» разметки обусловлен появлением новых парадигм исследований, отличающихся от ранее создаваемых письменных корпусов: например, исследования У. Чейза в области анализа «Рассказы о грушах». Таким образом стало понятно, что реальное общение не ограничивается одной только передачей информации, хранящейся в языковом варианте — большая часть опирается на неязыковые средства, в том числе на просодическую и кинетическую (жестовую) сущности языка. Эти исследования подтверждают, что естественная речь по своей природе мультимодальна и не проста в анализе. Следовательно, для таких исследований стали создаваться корпуса с особой иерархией разметки.

Подобной специфичной разметкой обладают бимодальные и полимодальные корпуса, так как разметка, создаваемая в этих корпусах, требует дополнительной теоретизации и разработки дополнительных алгоритмов, которые не предусмотрены в рамках национальных корпусов. К числу таких корпусов, например, можно отнести мультимодальный корпус

«Русский мультимодальный дискурс» [60], целью которого является комплексное изучение мультимодального дискурса на основе единого теоретического подхода и методологии. Обозначенное направление, где разные типы коммуникации рассмотрены вместе, в последнее время стал именоваться мультимодальным или мультимодальным. Базируется он на важном понятии модальности, которое было принято в физиологии и психологии, что является типом сенсорного сигнала [23]. Последовательное данное научное направление активно развивается в течение двух последних десятилетий. [1] [3] [6][9]. Как правило, дополнительно аннотирование осуществляется в программе ELAN.

Схожим типом и стратегией разметки обладают бимодальные корпуса, которые чаще всего ограничены двумя изучаемыми модальностями. Чаще всего – зрительной и слуховой. К бимодальным корпусам можно отнести корпус «Рассказы о сновидениях и другие корпуса звучащей речи», в котором представлены и размечены в программе ELAN тексты устной и письменной речи.

Помимо схожего технологического процесса создания бимультимодальных корпусов их общая черта заключается в междисциплинарном взаимодействии. В таких исследованиях активно принимают участие не только лингвисты, но и психологи, нейрофизиологи, исследователи, изучающие искусственный интеллект и разные виды дискурсивного, в том числе и делового, общения.

1.1.3.1 Русский мультимодальный корпус

Рассмотрим подробнее устный мультимодальный корпус, который связан с аннотированием мануальных жестов и других движений рук в мультимодальном (мультимодальном) дискурсе и основе которого материал создающегося русского корпуса «Рассказы и разговоры о грушах» [60].

Цель проекта заключатся в исследовании различных коммуникативных каналов, используемых в дискурсе и изучении их влияния на процесс коммуникации посредством методов мультимодальной лингвистики. Основой экспериментальных исследований является «Фильм о грушах» У. Чейфа [4], после просмотра которого ведется обсуждение в группах из четырех участников. Всякое обсуждение, или «запись», состоит из фрагментов монологов и общей коммуникации. Объем корпуса составляет двадцать четыре записи длительностью около девяти часов и содержит около ста тысяч слов. Каждая запись в корпусе состоит из комплекта Аннотирование мультиканального корпуса, куда входят мануальные жесты (пятьсот семнадцать медиафайлов): видео- и аудиофайлы, а также данные айтрекинга с комплектом аннотаций.

Разметка корпуса осуществляются в программе ELAN и включает в себя два основных типа аннотаций: основная и дополнительная. Основной разметкой учёные считают изучение и описание вокализма: вербального и просодического; а также окуломоторного канала, жестов рук, головы. Вокализм для его изучения и понимания разделяется на значимые фрагменты речевого потока (элементарные институциональные единицы, слова, с паузами и без них, неречевые звуки). К этому же можно отнести и придание свойств небольшим речевым элементам и непосредственно связанных с ними компонентам. Аннотация жестикуляции основывается на уровне линейного анализа передаваемого мануального кинетического поведения: это и моменты, когда нет никакого движения и отдельные двигательные моменты, формирующие отдельные функциональные единицы, а именно: жесты, адапторы, смена позы, а также их изменения в речевом общении. Для этих элементов позже определяются их данные: фазовая структура, функциональный тип и т.д.

Аннотирование жестов головы основано на принципах, разработанных в рамках проекта с целью аннотирования мануальных жестов. В ходе аннотации окуломоторного канала происходит экспорт данных айтрекинга на видеосцену

(в виде наложения точки). При помощи программы Tobii Pro Glasses Analyzer можно извлечь данные о непродолжительном открытии всех фиксированных движений длиной свыше ста миллисекунд, на них потом вручную кладётся аннотационная модель, указывающая на конкретный взгляд.

К дополнительной разметке относится и аннотирование жестов тела, мимики, проксемики, фонетической транскрипции, а также референциальная аннотация, в основу которой положена разметка разных языковых высказываний с более конкретной референцией.

Подобная разметка помогает проводить исследования в области дискурса, вербальной и кинетической коммуникации человека [22][44][29], сравнивать коммуникативные каналы и находить их взаимодействие друг с другом.

1.2.4 Особенности корпусов неродной речи

Особое место в корпусной лингвистике занимают корпуса неоднородной речи, целью которых является изучение отклонений от речевого стандарта языка у респондента. В таком случае корпус выступает инструментом автоматического анализа подобных отклонений и предоставляет исследователю возможность эффективно производить анализ речевых явлений. Подобные корпуса получили название «Корпус ошибок», и они представляют собой коллекции текстов нестандартных носителей языка: студентов, которые изучают русский, эмигрантов, билингвов и т.п. [8]. Идея поиска ошибок зародилась в исследованиях Анри Фрейя в 1929 [45]. Подобные работы стали доступны для русского языка на базе института ИРЯ им. В.В. Виноградова, они направлены на изучение особенностей ошибок разговорной формы русского языка зарубежья [26]. Только в последнее десятилетие это направление исследования стало активно развиваться в отечественной лингвистике на разных уровнях языка [15][21][27][34]. В период проведения подобных исследований у учёных была возможность

разграничить два понятия: русский язык как иностранный» и «русский язык как неродной» [28]. Однако и на этот счет нет единого мнения. В направленности корпусной лингвистики и в разработке ею вопросов применения языковых корпусов было определено деление в зависимости от разных групп носителей языка, а также от их компетенций в успешном овладении речью. Исследователь А. Мустайоки пишет о важной необходимости различения групп своеобразных носителей языка: студенты, изучающие русский язык в качестве иностранного, носители регионального русского, носители так называемого *эритажного*, или унаследованного русского, дети и подростки, взрослые носители стандартного русского в определенных ситуациях [8]. Как видно из вышеизложенного, большая часть работ противопоставляется по типу отношения к языку респондента. Одни корпуса связаны с поиском ошибок в устной или письменной речи респондентов, которые начали изучение языка. Другие корпуса определяют ошибки только у респондентов, для которых русский язык является родным: к ним можно отнести следующие корпуса: «Корпус русских учебных текстов» [59], «Русский учебный корпус» («Russian learning corpus») [61].

Таким корпусам противопоставляются корпуса контактных языков, цель которых проявляется в отражении специфики русского языка в условиях его существования в билингвальных ситуациях. Подобные корпуса обладают спецификой учебных и эритажных корпусов, но дополняют их материалом исследования, где материалом выступают русские тексты устной и письменной коммуникации и изучается интерференция родного языка на русский. Характерной чертой контактных корпусов является дополнительная лингвистическая и экстралингвистическая разметка, в которой учитывается тип билингва, его ведущий язык, территория проживания и специфичные морфологические, синтаксические, фонетические и т.п. теги отклонения от речевого стандарта. Наиболее известными подобными корпусами сегодня являются следующие корпуса: «Исследование контактно-обусловленной

русской речи малых народов России» (ВШЭ) [62], «DagRus» (Дагестанский корпус - ВШЭ) [63].

1.2.4.1 Корпус русских учебных текстов

Корпус русских учебных текстов (КРУТ) существует с 2013 г как один из проектов Лингвистической лаборатории по корпусным технологиям на факультете филологии в НИУ ВШЭ под руководством Е.В.Рахилиной. КРУТ –открытый бесплатный интернет-ресурс, содержащий учебные тексты студентов НИУ ВШЭ и других вузов, носителей русского языка. КРУТ интересен тем, что содержит тексты носителей языка на родном языке вместо ожидаемых в учебных корпусах текстов на изучаемом языке . Объем корпуса составляет более 2,5 миллионов слов, часть корпуса –500 тысяч слов размечена на предмет ошибок. Всего на данный момент выделено более 10 тыс. ошибок, разметка по ошибкам производится вручную [59].

Тексты в корпусе снабжены метатекстовой, морфологической разметкой и разметкой по ошибкам. Разметка по ошибкам производится вручную с помощью программы Crocodiles2.6. Классификация ошибок включает тип ошибки (лексический, грамматический, дискурсивный) и указание на возможную причину ошибки (опечатка, контаминация конструкций). Классификация ошибок содержит 39 помет, 20 верхнего и 19 нижнего уровней. Главной проблемой разметки по ошибкам является субъективность – зачастую одно и то же явление разными экспертами классифицируется по-разному, более того, в некоторых случаях возникают разногласия: выделять ли ошибку как таковую. Снизить эффект субъективности помогает многомерная классификация – возможность отнесения одной и той же ошибки к разным типам. Создание учебных корпусов преследует как педагогические, так и исследовательские цели. С одной стороны, на базе таких корпусов могут создаваться лингвистические тренажеры, тестовые и контрольные задания (с учетом самых распространенных ошибок), а кроме того – индивидуальные

учебные подкорпусы студентов. С другой стороны, ошибки служат маркерами языковых изменений.

Составители корпуса КРУТ утверждают то, что, работая с данным корпусом, исследователь или педагог может проводить работу по коррекции и предупреждению ошибок на предметных занятиях со школьниками и студентами, используя возможности корпуса. К примеру, на таких курсах как «Литературное редактирование», «Стилистика», практикумах по письменной и устной речи, а также с целью для изучения направления развития русского языка, что является научной исследовательской работой.

1.2.4.2 Корпус «Русских учебных текстов»

В 2013 г был составлен корпус «Русских учебных текстов» (RLC) под руководством Е.В. Рахилиной в лингвистической лаборатории по корпусным технологиям в высшей школе экономики [61]. Основа корпуса - устные и письменные тексты русского языка, относящиеся к двум категориям говорящих, владеющим русским языком: иностранцам, изучающим русский язык в качестве неродного, и «эритажных» (региональных) носителей родного русского языка. Для последних, для которых русский язык является родным, но по определённым ситуациям (чаще всего это эмиграция) основным средством коммуникации для них стал другой язык. Метатекстовой разметкой включается сообщение об авторе (пол, возраст, тип носителя: эритажный или иностранец), доминирующий язык (степень владения русским языком), а также информация о самом тексте: письменный или устный и его жанровая особенность. В тексте размечаются типы речевых отклонений от стандартного русского языка. Разработчики корпуса пишут о том, что «Русский учебный корпус – это ценный ресурс для исследований в области усвоения иностранных языков (Second Language Acquisition), преподавания иностранных языков (Second Language teaching), языковой интерференции, а также других лингвистических исследований». [Е.В. Рахилина].

Лингвистические материалы корпуса, поисковая система являются основой исследований, сравнивающих речь эритажных носителей и, в сравнении, носителей, которые русский язык изучают как иностранный. Используя в практике данные представленного корпуса возможно более глубокое освоение сложных языковых явлений, а именно: нестандартное использование русского вида, падежа, предлогов, неправильное или неточное употребление слов или неполное понимание смысла в многосложных образцах.

1.2.4.3 Корпус русской речи носителей автохтонных языков Севера Сибири и Дальнего Востока

Корпус носителей автохтонных языков Севера Сибири и Дальнего Востока и русской речи был разработан в 2019 г. на факультете филологии в НИУ ВШЭ под руководством Н.М. Стойновой и является частью большого проекта «Динамика языковых контактов в циркумполярном регионе», целью которого является изучение интерференции родных языков на русский [62].

Технически корпус приближен к созданному ранее корпусу «Linguistic Convergence Laboratory» и схож с корпусом высшей школы экономики «Russian Learner Corpus created in Linguistic Laboratory for Corpus Studies». Корпус контактной речи содержит записи текстов устной бытовой коммуникации русской речи билингвов (L2) малых народов, говорящих на родном языке (L1: чукотский, ульчский, нанайский, мокшанский, горномарийский языки и т.п.). Объем корпуса составил более 30 часов звучания устной речи билингвов на русском языке из 100 часов полевой записи. В частности, ульчский подкорпус (11334 словоупотреблений), нанайский подкорпус (27211 словоупотреблений), мокшанский подкорпус (17578 словоупотреблений), горномарийский подкорпус (15901 словоупотреблений). Общий объем корпуса сегодня составляет более 72 тысяч словоупотреблений.

Корпус содержит метаразметку респондентов, которая включена в веб-корпуса. Корпус расположен по адресу: http://web-corpora.net/tsakorpus_russian_nonst/corpus.html. Эта разметка состоит из: 1) технической информации о записи (имя файла, дата записи, место записи, длительность записи); 2) информации о тексте (тип, жанр, содержание); 3) информации о рассказчике (имя, шифр, родной язык, дата и место рождения, уровень образования, возраст усвоения языков); 4) транскрипции речи респондента.

В качестве отклонений от стандартов русского языка авторы выделяют следующие уровни, которые также ими были включены в разметку: 1) phonetics; 2) morphology; 3) syntax; 4) complex sentences & discourse & prosody 5) substandard.

2. Корпус речи тюркско-русских билингвов (RuTuVic).

2.1. Теоретические основания корпуса. Цель корпуса.

Важной особенностью Корпуса речи тюркско-русских билингвов (RuTuVic) является интеграция в программу междисциплинарных исследований процессов когнитивных и языковых эффектов, являющихся результатом взаимодействия языков народов, живших и живущих на территории Южной Сибири. Интеграция в межнаучную программу осуществляется с помощью созданного учёными проекта «Языковое и этнокультурное разнообразие Южной Сибири в синхронии и диахронии: взаимодействие языков и культур». Особенностью проекта является, во-первых, то, что динамические процессы языковых взаимодействий изучаются не только в аспекте изменений, протекающих в языках малых народностей Южной Сибири, но и в процессе их интерференции в русский язык как язык-макропосредник, являющийся средством коммуникации народов в регионе. Во-вторых, в проекте языковое взаимодействие исследуется не только в

проявлении на речевом уровне, но и также одновременно проводится анализ когнитивных, психолингвистических оснований таких проявлений. Такая постановка проблемы обуславливает применение в проекте корпусных технологий, технологий баз данных и экспериментальных аппаратных психолингвистических исследований. Таким образом, корпус посвящён нескольким трекам исследований: 1) фиксации типов проявлений русско-тюркской интерференции в системе отклонений от речевого стандарта в речи билингвов, соотнесению их с другими источниками отклонений, социолингвистическому портретированию билингва, фиксации его языкового опыта, что позволяет не только фиксировать языковые явления, но даёт возможность проследить влияние социальных и языковых факторов на появление отклонений от речевого стандарта.; 2) исследовании когнитивных особенностей и языковой картины мира билингвов. Так как исследование находится на стыке наук, для достижения целей применяются методы из смежных наук: корпусная лингвистика, когнитивная лингвистика, базы данных, статистика, психология, социология, экспериментальные методы исследования и т.п. [38] [40].

Исследования корпуса основаны на изучении форм проявления билингвальной или мультилингвальной интерференции, связанной с психолингвистическими, лингвокогнитивными и лингвистическими функциями речевой деятельности. Чтобы решить ряд поставленных исследователями целей и задач, создание корпуса билингвальной русскоязычной речи имеет, во-первых, самостоятельную научную ценность, так как результат работы будет представлен в виде фиксации типов речевых отклонений на всех ярусах языка и в соотнесении с типами языковых контактов. И второе: рассматриваются собранные учёными двуязычные тексты в соответствии с предложенной исследователями специальной языковой разметкой как основной источник составления «материалов психолингвистических экспериментальных исследований механизмов контактирования языков в когнитивных структурах билингвов» [35, С. 875].

Теоретическая и практическая стратегия разработки корпуса коррелирует со структурой и принципами «родительского» корпуса «Lingvodoc» и выбирает из него главные принципы определения текстов корпуса, его разметку и метаразметку, а их структура определена стремлением показать полноту функционального взаимодействия территориального варианта национального языка. К таким принципам учёными отнесены важные положения: при отборе текстов – отражение устной и письменной речи всех дискурсов, стремление к разнообразию жанров и тематики, а при планировании разметки – введение стилистически важных маркеров [36].

Текстовая база данных создаваемого корпуса включает подкорпуса записей устной русской речи шорцев, татар и хакасов, владеющих двумя языками. Русский язык может занимать различные функциональные позиции, в большинстве записей он представлен как осознаваемый информантом в качестве второго изучаемого языка (L2). Однако русский язык может выступать в качестве L1, тогда как шорский, татарский, хакасский – L2.

Корпус включает три подкорпуса, организуемых на основании своеобразной комбинаций языков – русско-шорское, русско-татарское, русско-хакасское двуязычие. В настоящее время база записей звучащей речи татарского подкорпуса включает – часов звучания, хакасского -, шорского -. Общее время звучания составляет более 400 часов.

Во всех подкорпусах представлена речь билингвов со сформированной языковой способностью представителей разных возрастных групп от 18 до 84 лет. Детская речь не записывалась. Записи собирались при помощи полевых исследований в местах исконного компактного проживания билингвов (Кемеровская область, п. Шерегеш, п. Таштагол, Большая суета (шорцы), Республика Хакасия (хакасы), Томская область (татары)), а также в ходе практики в Лаборатории лингвистической антропологии ТГУ (привлекались студенты-билингвы, получающие высшее образование в ТГУ)

2.2 Структура корпуса RuTuVic

Структура главного подкорпуса русской речи билингов по мнению исследователей состоит из определяющих её структуру особенностей:

1) основу подкорпуса составили отличия контактирующих языков – речевой подкорпус русско-татарских, русско-хакасских, русско-шорских билингов;

2) в отдельно взятом подкорпусе противопоставляются более частные подкорпуса, и к тому же с противопоставлением основных модусов речи – устной и письменной. Большая часть из созданных в 21 веке представительных корпусов национальных языков считаются совокупностью письменных текстов, и с ними объединяются подкорпуса мультимедийные или устные. В теоретической направленности концепции созданного исследователями корпуса русской речи тюркских билингов именно подкорпус устной речи стал определяющим, ядерным, поскольку только благодаря созданию его текстовой структуры возможна максимально полная фиксация социолингвистически, психолингвистически значимых особенностей билинга, а это, в свою очередь, очень важно учёным для изучения когнитивных особенностей в билингвальном взаимодействии. Не менее важной считается и возможность привлечения одних и тех же респондентов как к соавторству текстов, так и привлечение «испытуемых при проведении поведенческих психолингвистических исследований билингвизма» [35].

Технически корпус реализуется на базе языка PostgreSQL и работает по следующей структуре (Рисунок 1) [39]:

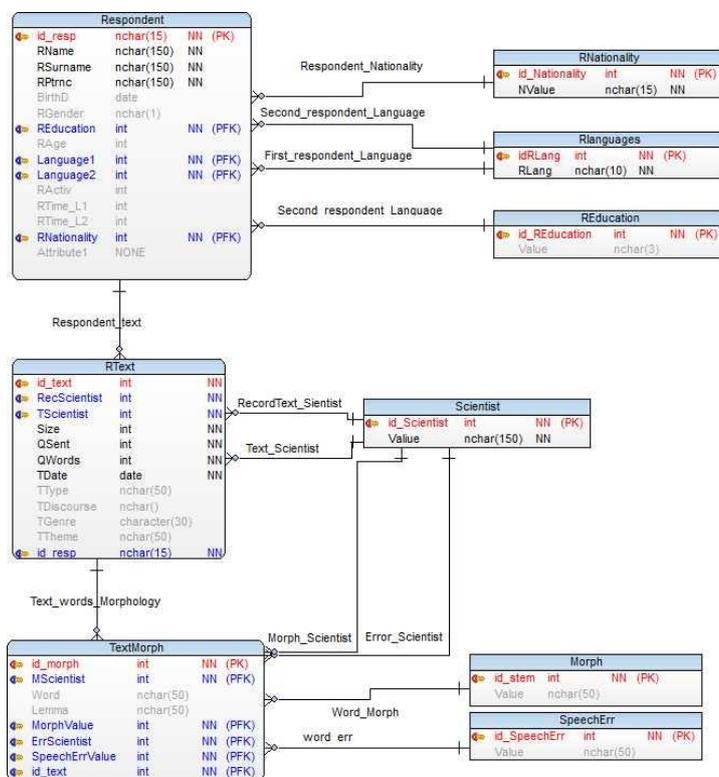


Рисунок 1 - Итоговая схема базы данных

В тестовом режиме была проверена работоспособность схемы на локальном сервере ТГУ: метаразметки на основе СУБД PostgreSQL (Рисунок 2).

Id	File Name	Size(Mb)	Date	Respondent	Age	Education	Ids
Sh0003_	Sh0003_001.mp3	78	2017-07-20	[REDACTED]	1970-03-11	V	45
Sh0003_	Sh0003_002.mp3	165	2017-07-18	[REDACTED]	1971-03-11	V	46
Sh0003_	Sh0003_003.mp3	8	2017-07-20	[REDACTED]	1972-03-11	V	47
Sh0013_	Sh0013_001.mp3	127	2017-08-19	[REDACTED]	1974-12-20	Student	135
Sh0013_	Sh0013_001.mp3	128	2017-08-17	[REDACTED]	1974-12-21	Student	136
T0043_	T0043_001.mp3	90	2017-07-24	[REDACTED]	1976-08-06	N	57
Sh0009_	Sh0009_001.mp3	91	2017-07-22	[REDACTED]	1977-08-06	N	58
Sh0009_	Sh0009_002.mp3	10	2017-07-24	[REDACTED]	1978-08-06	N	59
Sh0009_	Sh0009_003.mp3	4	2017-07-25	[REDACTED]	1979-08-06	N	60

Рисунок 2 - Графический интерфейс корпуса

Разметка текстов звучащей речи осуществляется через программу «ELAN» (Рисунок 3), а морфологическая основывается на консольной программе «Mystem»

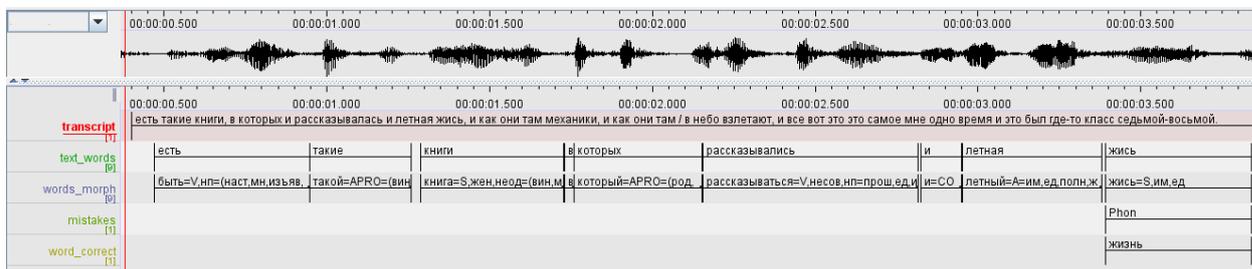


Рисунок 3 - Пример аннотирования фрагментов текстов билингвов (русско-татарский билингвизм).

2.3 Особенности разметки корпуса RuTuVic

Как уже указывалось в п. 2.2 вся разметка осуществляется при помощи определенного программного обеспечения. В частности, морфологическая разметка основывается на алгоритмах консольной программы компании «Яндекс»: «Mystem», морфологическая информация ограничивается возможностью работы этой программы. Программа поддерживает составление пользовательских словарей, в которых указываются те лексические единицы, которые некорректно отображает программа.

Следующий тип лингвистической разметки – разметка отклонений от речевого стандарта. Нормированная и обработанная письменная литературная форма русского языка является в данном случае стандартом, на её основе были определены принципы грамматического анализа языка в «Грамматическом словаре русского языка» А.А. Зализняка. На системной основе определённых принципов и правил (о них писалось ранее) существует корпусная система морфологической автоматической разметки Mystem. Корпус представляет собой Записи устной речи составили основу данного корпуса, поэтому некоторая часть отклонений письменной речи от стандарта связана с устным,

неподготовленным характером общения коммуникантов. Определение отклонений, речевых ошибок, по мнению создателей корпуса RuTuVic, способно расширить «...информативные возможности корпуса за счет аннотирования маркеров разговорности коммуникации» [36, С. 128]

Основное направление общности в системе разметки – ориентация на поуровневое представление отклонений от речевого стандарта.

В настоящее время исследователями принята модель одного или двух уровней классификации подвида отклонения, небольшим количеством примеров характеризуется и третий уровень конкретизации, который объяснён с точки зрения важности необходимости объяснения и разъяснения интерферентных особенностей:

первым обозначенным исследователями уровнем осуществляется маркирование уровня языковой системы, к примеру: Phon – фонетика, Lex – лексика, Morph – морфология, Synt – синтаксис, , Disk – дискурс и т.п;

Благодаря второму уровню происходит маркирование конкретного языкового явления исследуемого уровня, к примеру: PhonAcc – фонетика, ударение; Morph – морфология, флексия; SyntGov – синтаксис, управление и т.п.;

третий уровень позволяет определить варианты отклонений в речи билингва при помощи единиц и форм конкретного языкового яруса: SyntAgrGen – отклонения от речевого стандарта в согласовании по роду.

Заключение о внутрисистемном источнике отклонений от стандарта литературного языка определяется и выносится с использованием данных, зафиксированных в работах, которые посвящены описанию рассматриваемых подсистем русского языка [48] [49] [16].

2.4. Особенности метаразметки корпуса RuTuVic

Важным дополнением разметки корпуса устных текстов являются языковые и социолингвистические анкеты, разработанные в Институте

языкознания РАН, а они, в свою очередь, опираются на анкеты, созданные О. А. Казакевич и применяются для исследования языков малых народов России.

В социолингвистической анкете определён 41 вопрос с подробной детализацией информации о респонденте: время и место его рождения, проживания, обучение и образование, профессиональная деятельность.

Сюда же включены и сведения о родственниках по различному признаку родства, о характере приобретения и использования респондентом языков.

Языковая анкета билингва, разработанная на основе анкеты языкового опыта и уровня владения языком Marian V., K. Blumenfeld H., Kaushanskaya M., включает 14 блоков параметризации характера и типа владения билингвом взаимодействующими языками: языки ранжируются по мере активности их использования, порядку усвоения, количеству времени пользования языков во время интервьюирования, по предпочтению выбора языков при чтении и при коммуникации с другим человеком; также характеризуется история пользования языком – время изучения или вхождения в язык, время пребывания в среде языка, самооценка информантом уровня владения языком и факторов, стимулирующих изучение каждого из языков, которым владеет информант, определение предпочитаемых тем и сфер коммуникации для говорения на каждом из языков.

Предложенная корпусом метаразметка структурируется относительно владения респондентом русским языком и, соответственно, определяет его статус: является ли русский язык по самооценке информанта родным или неродным; используется ли в период записи текста в разных формах коммуникации (активный vs. пассивный), порядок усвоения языка (первый vs. второй); сфера преимущественного использования (письменная vs. устная; бытовая vs. официальная vs. эстетическая vs. другие. По тем же параметрам оцениваются другие языки речевых практик билингва (полилингва).

Информация о тексте включает также данные о времени и месте записи текста, его размере в словах. Далее следуют лингвистически релевантные параметры текста.

Первый, наиболее общий параметр квалификации текста, – его принадлежность к типу дискурса, противопоставляется в соответствии со сложившимися в мировой и российской лингвистике параметрами: институциональные и личностные. Запись текста, в целом, может характеризоваться переключением дискурсивного регистра. Так, фрагменты записи интервью, фиксирующие представление информанта интервьюеру, студенту или преподавателю университета, и соответствующие жесткому регламенту анкеты, интерпретируются как институциональный. Обозначается тэгом [InstD]. Однако интервью может перетекать в дружескую беседу, воспоминания. Такие тексты квалифицируются в корпусе как относящиеся к личностным дискурсам. Маркируется тэгом [PersD].

Глава 2. Отклонения от речевого стандарта в речи шорско-русских билингов как проявление внутри и межязыковой интерференции

2.1. Языковая интерференция и ее типы.

2.1.1 Интерференция и проблемы контактирования.

У. Вайнрайх, подчеркивая, что языковое сообщество никогда не бывает однородным, и то, что говорящий всегда приспособливает свою речь к слушающему, отмечал, что это приспособление имеет значительные отличия в случаях, когда мы находимся в одной языковой системе и в разных. Во втором случае говорящий постоянно переключается с одной языковой системы на другую, что может быть охарактеризовано как идеальная двуязычная ситуация. Двуязычием он называет «практику попеременного пользования двумя языками». Из явления двуязычия У. Вайнрайх выводит определение интерференции: «Случаи отклонения от норм любого из языков, которые происходят в речи двуязычных в результате того, что они знают больше языков, чем один, т. е. вследствие языкового контакта». Интерференция являет собой результат перестройки моделей одного языка под влиянием другой, при этом необходимо осознавать значимость такого фактора, как различие языковых систем вступающих в контакт языков, хотя механизм интерференции будет одним и тем же как в близкородственных языках, так и типологически различных языках. Стремясь вычленить и определить явления интерференции, необходимо учитывать следующее [12, с. 22-23]:

1) список возможных интерферентных явлений лингвист получает при сравнении подсистем вступающих в контакт языков;

2) есть специфические пути анализа: например, при изучении интерференции в сфере лексики необходимо изучить несоответствие словаря одного языка потребностям культурного окружения, в котором происходит языковой контакт;

3) необходимо учитывать экстралингвистические факторы языкового контактирования.

2.1.2 Отношения индивида к языкам контактирования как фактор интерференции

Изучение социально-культурного контекста и психологического субстрата представляется чрезвычайно важным для понимания природы интерференции.

К числу экстралингвистических факторов, связанных с отношением билингва к языкам контактирования, которые важны при социолингвистическом анализе, относятся:

1) способность находить общие элементы вербальной коммуникации, способность тех, кто общается, определять различия между двумя-тремя языками, которыми владеет говорящий, и также возможность определять совместные средства;

2) относительная степень компетенции говорящего в степени владения каждым из известных ему языков;

3) необходимость в специальной подготовке в использовании каждого из языков в зависимости от тем коммуникации;

4) необходимость определять методы и средства овладения языками;

5) важно отношение говорящего к языкам (идиосинкретическое или стереотипное).

2.1.3 Межгрупповое взаимодействие как фактор интерференции

Однако влияние интерференции на изменение языковой системы будет более сильным, если наблюдается межгрупповое двуязычие. В этом случае важно учитывать:

1) количество участников двуязычной системы, социальное и культурное единство или различие, формирование микрогрупп, которые

используют тот или иной язык в качестве родного; демография коммуникантов; политические и общественные связи между говорящими в микрогруппах;

2) главенствование в некоторых микрогруппах билингов определенного типа.

3) привычное отношение коммуникантов к разным языкам, которыми он владеет (его выделение); родной или заимствованный статус языка;

4) уважение и понимание культуры всех языковых сообществ, которые знакомы билингу;

5) отношение к билингвизму как к существующему в мировой цивилизации явлению;

6) принятие или неприятие такой особенности, как билингвальное смешение языков и к нарушениям речи на каждом из них;

7) взаимовлияние говорящих в билингвальной группе, в двух разных речевых сообществах, по отношению к которым она оказывается в роли «маргинального сегмента» (У. Вайнрайх), утверждая свою собственную систему ценностей [12, с.25-26].

2.1.4 Типы интерференции по отнесенности к уровням языковой системы

В типологии интерференции основное место занимает противопоставление по уровням языка, а именно - фонетическая, грамматическая, семантическая и стилистическая интеграция элементов одного языка в систему другого.

Рассмотрим последовательно данные виды интерференции.

2.1.4.1 Фонетическая интерференция.

Как отмечает У. Вайнрайх, звуки, которые произносятся билингвом, «находятся как бы в ничейной полосе между двумя системами фонем». Наиболее значимые явления интерференции [12, с.45-46]:

1) недодифференциация фонем - смешение двух звуков принимающей языковой системы, соответствия которым в первичном языке не различаются (в качестве примера приводится звук [й] в *яблоко* или *перед гласным в безударной позиции: молодая, молодой*);

2) сверхдифференциация фонем - в данном случае на звуки вторичной системы накладываются фонологические различия первичной системы там, где их не должно быть. Примером может явиться палатализация гласных в тюркских языках;

3) реинтерпретация различий - различение фонемы в принимающем языке по тем признакам, которые для нее являются сопутствующими;

4) субституция звуков - в обоих языках фонемы определяются одинаково, но в их нормальном произношении существуют различия, примером чего может служить различие специфических гласных и согласных звуков в тюркских языках в отличие от русского.

Следует отметить еще одно важное замечание, сделанное У. Вайнрайхом со ссылкой на мнение Р. Якобсона, - «язык «принимает иноязычные структурные элементы только в тех случаях, когда они соответствуют его собственным тенденциям развития» [12, с.55].

2.1.4.2 Грамматическая интерференция

По мнению многих исследователей, вопрос о грамматической интерференции является одним из наиболее спорных проблем общего языкознания. Точки зрения варьируются от отрицания самой возможности

грамматической интерференции (А. Мейе) до возможности ее проникновения в самые формализованные элементы (Г. Шухардт).

Осложняется решение проблемы 1) несогласованностью терминологии, применяемой при описании разных языков; 2) различиями в определении границ между языковыми единицами и уровнями языка (например, морфологии и синтаксиса, лексики и морфологии и под.); 3) обязательностью грамматических категорий.

Типы грамматической интерференции:

1) употребление аффиксов родного языка в бытовой или литературной речи во втором неродном языке билингва;

2) употребление частного грамматического явления родного языка в речи на языке неродном или непонимание языковых единиц, свойственных неродному языку, поскольку нет таких категорий в родном языке (примером этого явления может быть категория рода, а именно, её отсутствие в тюркских языках, приводит к тому, что билингв не различает род в русском языке);

3) расширение или сужение значений морфемы неродного языка в соответствии с грамматическим строем и категориями родного языка в результате отождествления морфемы неродного языка с морфемой родного (*в качестве примера приведем однозначность и многозначность флексий тюркского и русского языков, соответственно*).

Интерференция, которая отмечается в области грамматики.

1) в синтаксисе отмечается перенесение норм порядка слов родного языка; моделей согласования и подчинения. В качестве примера приведем согласование в роде, числе, падеже и управление падежными формами в тюркском и русском языках – тюркоязычные билингвы нарушают нормы согласования в роде прошедшего времени глагола русского языка с подлежащим и поэтому, в соотношении с моделью родного языка, говорят по-русски неверно: сын не ходила магазин, вместо – не ходил в магазин.

Значительную степень распространенности данного типа интерференции можно объяснить тем, что, не будучи сегментами высказывания,

грамматические отношения с точки зрения наивных носителей языка менее заметны.

Распространение такого типа интерференции как копирование функций эквивалентных морфем видится У Вайнрайхом в следующем: «Если двуязычный индивид отождествляет морфему или грамматическую категорию языка А с морфемой или категорией языка Б, то он может использовать форму Б для выполнения тех грамматических функций, которые он берет из системы А. К установлению междуязычной эквивалентности морфем или категорий двуязычный приходит на основе либо их формального сходства, либо сходства их предыдущих функций» [12,с.73-74]. Примером такой интерференции может быть, на наш взгляд, неправильное употребление тюркскими билингвами предлогов родительного и винительного падежа.

2.1.4.3 Лексическая интерференция

При воздействии лексического состава одного языка на лексикон другой системы могут использоваться разные способы: 1) могут заимствоваться целостные лексемы в единстве значения и формы (*например, пицца*); 2) лексемы могут использоваться во втором языке с новыми функциями/значениями (*например, появление в якутском языке у слова *tahut* влиянием русского языка кроме исконного «уровень воды» обобщающего значения - «уровни» — уровень воды, развития, мастерства*); 3) слабая интерференция в лексической системе языка возникает в том случае, если выражение знаковой единицы меняется в сравнении с родственным словом в контактирующем языке без изменений в плане содержания.

Среди составных лексических единиц, состоящих более чем из одного аффикса, определяют три типа интерференции: все их элементы, единицы можно перенести в аналитической форме; все единицы воспроизвести, используя семантическое расширение; и всё же одни элементы могут переноситься, а другие только воспроизводятся.

Воспроизведение при помощи своих аналогичных слов применительно к сложным словам, словосочетаниям и даже к большим единицам, например, пословицам – называется калькированием. Выделяют собственно кальки, воспроизводящие точно неродную языковую модель, компонент за компонентом; кальки - толкования, при их появлении сложный элемент неродной языковой структуры определяет только общее положение для воспроизведения; например, немецкие: Vater-land «отечество» по образцу латинского patr-ia; кальки - образования.

Интерферентные явления наблюдаются не только в сфере нарицательных существительных, но и собственных. Это, например, такие названия, как итальянское *Monfalcone* — словенское *Trzic* в Италии; *Кондома* – шорское (длинная), *Кундома* – южносамодийское (извилистая). Не связанные между собой одно и то же лицо или одна и та же местность в очень редких случаях имеют похожее наименование в языке разных народов.

В качестве причин лексического заимствования выделяют системно-языковые и социальные (языковой опыт билингва).

К системно-языковым причинам относят: 1) наличие лексических лакун, необходимость в обозначении новых предметов, лиц, местностей и понятий; 2) наличие вредной омонимии, стремление к ее преодолению; 3) утрата аффективными словами родного языка их исконной выразительности.

Причины интерференции, возникшие в результате приобретённого языкового опыта билингва: 1) сопоставление со вторым неродным языком может вызывать чувства, что некоторые семантические поля родного языка не чётко различимы; 2) возможно появление символических ассоциаций, возникших в результате контакта языка-источника и определенными социальными значимостями: положительными или отрицательными; 3) речевая небрежность.

2.2. Межъязыковая и внутриязыковая интерференция

В типологии интерференции принято также противопоставление по источникам интерферентных явлений в речевых практиках носителей языка или целых групп.

Теория интерференции складывалась при обращении к материалам межсистемного межъязыкового взаимодействия. Именно этот тип рассматривается в большинстве работ данного направления. Однако существуют явления межсистемного взаимодействия в пределах одного этнического языка. В данном случае исследователи исходят из того, что в современном мире большинство этнических языков представляют собой макросистемы относительно автономных подсистем, в другой терминологии – форм языка, своеобразие которых определяется социальными факторами. Наряду с функционально доминирующей подсистемой литературного языка выделяются диалектная, просторечная подсистемы, профессиональные подъязыки, жаргоны и аргот, в разной степени противопоставленные на различных языковых уровнях – фонетическом, морфологическом, синтаксическом, дискурсивном (текстовом). Носитель одной из форм национального языка может, находясь в диалектном или просторечном окружении, испытывать их влияние, что является причиной проникновения данных явлений в речевые практики.

2.2.1 Особенности языковых систем – источников внутриязыковой и межъязыковой интерференции в речи шорско-русских билингвов

В работе исследуются интерферентные проявления в русской устной речи шорско-русских билингвов. Вслед за рядом исследователей мы полагаем, что источником отклонений от речевого стандарта, в качестве которого в корпусе тюркско-русских билингвов приняты отклонения от норм русской литературной речи могут быть как интерферентные явления (межъязыковые и

внутриязыковые), так и нормы русской разговорной речи. В работе мы разграничиваем эти явления, вследствие чего считаем необходимым далее охарактеризовать основные особенности языковых подсистем, в среде которых существуют носители билингвизма, речевые особенности которых изучаются.

2.2.1.1 Особенности шорского языка – источник интерферентных явлений в речи шорско-русских билингвов

Фонетика. В данном разделе мы охарактеризуем только те особенности фонетики, которые связаны с явлениями грамматической и лексической интференции. Наиболее яркое отличие – явление сингармонизма. В шорском языке определяется восемь гласных звуков: *a, o, y, э, ы, и, ö, ү*. Шорские гласные классифицируются как гласные заднего ряда (твердые): *a, y, o, ы; и* гласные переднего ряда (мягкие): *э, и, ö, ү*. В русском языке твёрдость и мягкость гласных не определяется. В тюркских языках эта характеристика является важной, так как связана с общим законом сингармонизма: все звуки в словах либо твёрдые, либо мягкие: *ул (сын, мальчик) – үй (дом)*. Кроме вышесказанного гармония гласных заключается в том, что в одном и том же слове могут сочетаться или только гласные заднего ряда, или только гласные переднего ряда. При этом к основам с гласным заднего ряда прибавляются аффиксы только заднего ряда, а к основам с гласным переднего ряда прибавляются аффиксы только переднего ряда. Например: *қазан «котел», азық «запас», аяс (айас) «ясный», тебир «железо», кебеге «печь», тегелек «колесо», келескин «ящерица», күртүк «тетерев», түргек «клубок»*. Нёбная гармония - гармония гласных по принципу переднеязычности и заднеязычности. Губная гармония заключается в том, что после губного гласного (*o, y или ö, ү*) в следующем слове употребляются также только губные *o, y или ö, ү*. Например: *торчуқ «соловей», қомус «музыкальный инструмент», қусқун «ворон»; пөрү «волк», көрүк «бурундук»*. Однако губная гармония

применяется непоследовательно, особенно у шорцев бассейна р. Кондома. Эта непоследовательность проявления губной гармонии выразилась в допущении двух вариантов написания и произношения многих слов в литературном языке (*Кондома - тюркское, а Кундома – южносамалийское, полдум — полдым «я был», тўлгү — тўлге «лисица»*). В шорском языке при употреблении удвоенных гласных наблюдается явление стяжения звуков, например, *ан – аан* (подобное явление наблюдаем в результате интерференции «прииск» произносится как «прыск» и подобное произношение ещё связано с узуальностью в тюркских языках звуков [и, ы]) [18, с.19, 29]

Долгота гласных в шорском языке имеет важное значение, являясь признаком фонем, и выполняет смыслоразличительную функцию, например, *кол – «рука», коол – «пустой», «дупло»*.

В соответствии с гармонией гласных, как видом прогрессивной ассимиляции, в шорском языке существует и ассимиляция согласных: после глухих согласных, как правило, следуют глухие, после звонких согласных и гласных следуют звонкие, после носовых — носовые (в особенных случаях для носовых), например: *шошқа «свинья», асқыр (наряду сайгыр) «жеребец»; қарға «ворона», тилбис «кулик»*.

Нужно отметить, что в шорском языке не было губно-зубного звука [в], его заменял билабиальный [w] в словах иноязычного происхождения, это отражается в таком явлении интерференции, как опущение или изменение при построении словосочетаний предлога «в» в устной речи шорцев, например, *на подготовительный (в подготовительный класс; ну, эт... уже... семидесятом (в семидесятом))*.

Ударение в тюркских языках является дистинктивным, но может быть только конечным в слове, например: *бала, балапан, дантерлер, кыздар* и т. д.

Ударение в русском языке является демаркационным; в слове оно свободно, кроме того, может быть подвижным или неподвижным, например, *опт, оптовый, крокодил, торты* и др. Отсюда возникает отмечаемое

стремление шорцев поставить ударение в русских словах с фиксацией на последнем слоге: *селезЕнь, КондоМА*.

В ряде случаев ударение обуславливает семантику той или иной грамматической формы: под ударением -та / -да является аффиксом местного падежа, а в неударном положении — частицей: *анда* «у зверя», *қушта* «у птицы», *аң-да көп қуш-такөп* «и зверей и птиц много» [17, с. 49].

Морфемика шорского языка

Отметим агглютинативный характер шорского языка, как и всех тюркских языков: аффиксы последовательно нанизываются на слово, образуя либо формы слова, либо новое слово, например, «*паиш*» - *голова, вершина, начало*, «*паиштык*» - *староста, руководитель*, «*паиштыктар*» - *старосты, руководители*. Аффиксы определяют конкретный падеж или число слова в сочетании с другими словами, являясь однозначными формами, но выбор первого звука аффикса связан с законом прогрессивной ассимиляции и с его вариантом – сингармонизмом, и может изменяться. Фонетические варианты аффикса множественного числа есть результат ассимиляции согласных и гармонии гласных. Фонетический облик аффикса зависит от характера основы, от ее гласных и конечного согласного или гласного. Множественное число существительных образуется присоединением аффикса -лар / -лер после конечного гласного или звонкого согласного, кроме носового, -тар / тер- после глухого согласного или -нар / нер после конечного носового (м, н, ң) [46, с.21].

Имена существительные в шорском языке бывают одушевлёнными и неодушевлёнными, собственными и нарицательными. В грамматике существительного определяется категория числа: единственное и множественное, категория падежа (7 падежей: основной или именительный, винительный, напрвительно-дательный, местно-временной (предложный), исходный, совместно-орудный (творительный). Особенности образования падежных форм шорского языка связаны прежде всего с категорией неопределенности. Выделяется неоформленный безаффиксальный падеж существительного [46,с.61].

Существительное в родительном и винительном падежах, когда речь идет о предмете вообще, т. е. когда он не выделяется из числа ему подобных (ср. существительное с неопределенным артиклем в английском и немецком языках), ставится в нулевой (безаффиксальной) форме. В грамматике такую падежную форму называют неопределенным (или неоформленным) падежом. Неопределенный падеж существительного выражает категорию неопределенности, противопоставляемую категории определенности, которую выражают оформленные притяжательно-родительный и винительный падежи. Этим объясняются ошибки шорских билингвов в употреблении падежных форм русского языка, чаще всего в падежном управлении: *со Спасском (в Спасском) работали* [18, с.40,41, 52,53,59].

Категория рода, как и во всех тюркских языках, в шорском не определяется. Родовые признаки в словах отличаются только по полу: *ана – мама, аба- отец, кыз- девочка, дочь, ул- сын, мальчик.*

Имена прилагательные шорского языка морфологически являются одновременно и существительными: *таш –«каменный», «камень», агаш- «дерево», «деревянный», кара – «чернота», «чёрный».* Если имя в позиции атрибута, определения в предложении – это прилагательное с признаками качества, если субъект (подлежащее) или употребляется изолированно – существительное. Имя прилагательное в шорском языке не изменяется ни по числам, ни по падежам (*улуг кол-«большое озеро», ср. улуг коллар – «большие озёра»*). Предшествует в предложении имени существительному. При существительном может быть несколько определений (прилагательных). Употребляясь с глаголом, прилагательное в предложении переходит в наречие [18, с. 70-71].

Особенности глагольных форм шорского языка. Глагол в шорском языке обозначает действие или состояние предмета. Есть в языке неопределённая форма глагола, которая называется **супин**: к основе слова присоединяются аффиксы –рча/-рче. Определяются словоизменительные категории: видовременные (10 форм) и наклонения (изъявительное, желательное и

повелительное), изменяются глаголы по лицам и числам с использованием как простых так и сложных аффиксов (-чалар, -чазаар, -алчан и др.). Глагол в шорском языке имеет сложную систему временных форм 1. Прошедшее время (недавнепрошедшее). Значение данного времени — действие, происходившее незадолго до настоящего времени, действие законченное, а также сообщение о достоверном факте. 2. Прошедшее время на -ган (давнопрошедшее): значение данного времени — действие, совершившееся в прошлом, но не приуроченное к определённом моменту. 3. Прошедшее-будущее время (обычное, постоянное, многократное): данное время определяет обычность, повторяемость действия, важность совершения; действие имело место в прошлом, оно было обычным, повторялось много раз; время выражает, кроме того, действие, которое должно произойти в будущем. 4. Прошедшее несовершенное время указывает на то, что действие ещё не совершилось, но ожидается. 5. Прошедшее неожиданное время раскрывается действием, совершённым в прошлом, и которое способно вызвать удивление. 6. Настоящее-прошедшее несовершенное время глагола обозначает значение данного действие, которое совершалось в прошлом (реже в настоящем), без указания на его завершение, причём действие характеризуется как длительное. 7. Настоящее неожиданное время на -чаттыр: действие, совершающееся до момента речи или в момент речи; действие, о котором говорящий узнаёт во время его совершения или после; действие, свидетелем которого говорящий не был, которому он несколько удивлён, которое для него неожиданно. 8. Настоящее сложное длительное время. 9. Настоящее сложное продолженное время заключается в действии, протекающем в настоящем времени, но не законченном, но оно в ближайшее время должно завершиться; а также действие, которое вот-вот должно начаться. 10. Будущее вероятное время глагола обозначает действие или состояние, которое, возможно, наступит или совершится в будущем [48, с.240-252].

От глагола в шорском языке образуются причастия и деепричастия, представляя собой простое предложение, входящее в состав сложного,

например: *алтын шарчын тозин эбире тактап, алтын шарчин кезе кепеп аг-ой турча* – «бело-серый конь стоит, вокруг золотой коновязи, утаптывая, верхушку золотой коновязи отгрызая».

Служебные слова и послелогои определяют тюркский характер шорского языка. Служебные слова, управляемые падежами и соответствующие русским предлогам, называются послелогоми. Послелог, в отличие от предлога, стоит после имени. Послелогои играют в предложении служебную роль, уточняя значение падежей. Некоторые послелогои (па, ча) очень близки к падежным аффиксам и при некоторых условиях могли бы рассматриваться таковыми. Эти послелогои могут писаться слитно с именем.

Наречие характеризуется неразложимой основой. Сравнительно небольшая группа наречий для современной стадии языка представляет пока неразложимые основы (неразложимой может быть и часть сложносоставного наречия): *Қарчы «назад», пере «сюда», қарчы-перчи «взад-вперёд», ам «сейчас», ылтам «скоро», элееде «много», «многовато», арай, арий «едва», «чуть», айбын «поздно», шагам «только что», кӧӧче «медленно».*

Синтаксические особенности шорского языка. В шорском языке выделяются подлежащее и сказуемое. Из второстепенных членов предложения – определение. Подлежащее в шорском языке занимает место впереди сказуемого. . Прямой порядок слов в грамматике тюркских языков имеет особенность: сказуемое всегда стоит в конце предложения, перед ним ставится и подлежащее, и второстепенные члены предложения. Если подлежащее при себе имеет определение, то последнее стоит перед ним. Подлежащее простого предложения в шорском языке — всегда в основном падеже и отвечает на вопросы: *кем? «кто?», ноо? «что?» ноо небе? «что?».* Функцию подлежащего могут выполнять все имена (существительное, прилагательное, местоимение, числительное), в том числе и глагольные имена (причастия), с определением или без определения, в единственном или множественном числе, с аффиксом принадлежности или без него. Сказуемое в шорском языке — это главный член двусоставного предложения,

выражаемый чаще всего формами глагола, а также причастиями и деепричастиями (в придаточном предложении, отсюда отмечаем ошибки в речи шорских билингвов: *Первый рабочий свой, начинающую работу (начала работать) как восемнадцать лет мне было (когда мне было восемнадцать лет)*), существительными, прилагательными и другими частями речи. Подлежащее со сказуемым в предложении всегда согласованы в лице и в числе. Причем согласование в лице в шорском языке всегда существует, но оно не всегда приемлемо в сочетаемости по числу [18, с.230].

Согласования в роде в шорском языке нет, так как в грамматике шорского языка отсутствует род. Например: *Колхозчы мазарга парды. «Колхозник поехал на базар».* *Колхозчылар мазарга пардылар. «Колхозники поехали на базар».* Имя, выступающее в функции предикатива, может сопровождаться послелогом. При этом само имя может выступать в притяжательно-родительном падеже. Например: *Турна уйазы узун састың ортада. «Гнездо журавля посредине длинного болота».* *Кебе чар тӧзинде. «Лодка у крутого берега».* Определение бывает: предметное, принадлежностное, личное притяжательное и составное [18, с. 216].

2.2.1.2 Фонетические и морфологические особенности русских говоров Сибири – источник интерферентных явлений в речи шорско-русских билингвов

Своеобразие говоров Южной Сибири объясняется поздним поселением людей в азиатской части России и определяется разными условиями: смешанный, разнодиалектный состав населения, оторванность от материнских говоров, взаимодействие с языками населявших Сибирь других народов и национальностей (финно-угоров, самодийцев, тюрков), новые географические условия и условия быта, условия уклада и традиций, живших на этой территории аборигенов, а также сибирского населения, предки которого жили в Сибири с начала её освоения и до середины 18 века. Сначала поселились в Сибири выходцы из северных краёв: новгородских, пермских, вятских,

олонецких, вологодских, архангельских. Связано это было с колонизацией северных народов. Позже появились поселенцы из среднерусской части России. «Выходцы из этих местностей до сих пор сохраняют черты говоров средневеликорусских и южновеликорусских» (А. М. Селищев) [42, с 9]. Русские говоры Среднего Приобья по своим характеристикам тоже относятся к «старожильческим говорам Сибири» [31, с.19].

Фонетическая система говоров. Отличительной фонетической чертой среднеобских говоров, в сравнении с другими говорами России, является «аканье». Только в отдельных случаях отмечалось «оканье». «Еканье» является диалектной особенностью, присущей среднеобским говорам. В меньшей степени среднеобским говорам в безударной позиции свойственно «иканье». В речи шорцев мы тоже отмечаем эту особенность («еканье») как диалектную. В прикетских говорах описано наличие губно-губного звука «в» как диалектная особенность. Это же свойственно было ранее, в качестве речевой нормы, и шорскому языку, где ранее не было ни буквы, ни звука «в». Его заменял губно-губной звук. Особенностью приобских говоров было также употребление, как взрывного, так и фрикативного звука «г». В конце и середине слов и «х»: *в охтябре не мох (не мог), берех (берег)*. Вокализм среднеобских говоров представляет собой некое единство, а вот расхождения в произношении чаще наблюдаются в системе консонантизма. Гласные в своей основе в сильной позиции не меняются, слышатся отчётливо, но нужно отметить, что в некоторых случаях фонема «э» под ударением произносится как «и»: *«сивер», «ис» («есть»)*, а вместо фонемы «а» - «э»: *«стучэли», «кричэл», вместо «и» - «э»: «тваех», «адне», «батеначки», вместо «а» - «о»: «пасодют», «заплотишь», «намотывать»*. Шипящие «ж, ш, щ» в части среднеобских говоров являются мягкими, соответственно им и произношение гласных: *(ж'онка, ж'ена, ш'ес')*. Также отмечается ассимиляция звуков «з», «с», «ж'», «ш'»: *«уежж'аю» (уезжаю)*.

Протеза звука «в» является одной из особенностей среднеобских говоров: *«вострый», «воно», «вумны»*. В связи с переходом спيرانтов губно-губных

звуков к губно-зубным связана замена «ф» на заднеязычные или губные :«вахли», «ахтобус». Также в говорах отмечается лексикализованное изменение «вн» - «мн»: «мнучик», «нарамне». Фонема «л» может иметь варианты в среднеобских говорах «ф», «п»: «тофстоногость», «карап'» (корабль). Диалектные различия звуков «р, р'» обнаруживаются перед мягкими согласными: «грып», «товарышшы», «рыцен» (в речи шорца: ешишо). Это явление лексикализовано, варианты выступают как лексико-фонетические варианты слов. В целом все среднеобские говоры имеют общую систему консонантизма и вокализма, которая, так или иначе, прослеживается и в речи шорских билингвов.

Отличия в морфологической системе связаны также с тем, что существовала своя система и к ней присоединились другие говоры, завезённые из других регионов. И это явление оставило след в морфологической системе среднеобских говоров. Наиболее яркие противопоставления в употреблении падежных форм и падежной категории: *ножов – ножей, рублёв - рублей, яблок – яблоков, крестьяне- крестьяня, татары, снеги, на Кете- на Кети, на Томе – на Томи и т.д.* Различались по диалектам и местоимения: «мене» или «мине» - меня, «тебе» или «тобя»(тебя), «собя» (себя), «оне» (они), указательных - «тое», «той» (той), «такем» местом. Как и у прилагательных у местоимений-прилагательных наблюдаем стяжение форм: *така, тако место.* Необычно и диалектное употребление указательных местоимений это, эта с начальным е: *ето, ета, етих и т.д.* Притяжательные местоимения имеют в среднеобских диалектах по екающей норме окончания -ёго,-ёй, -ёму: *своёго, своёму, своей мойей*, а также в середине букву –н-: *ихный, ихной.* В Р. и В. П. употребляются спорадически сокращённые формы: *мово, твово, свово.* Из вопросительно-относительных местоимений в говорах отмечаются следующие: *кто, чё, чё-то, какой, который; отрицательные – ничё, некава, нечаво, неча.* Местоимения чё, ничё часто употребляются в речи шорцами. Диалектные отличия были и в образовании глаголов и в образовании

деепричастий: «идтить», «сести», «бояся», «не дошед». Иногда отмечаются случаи стяжения окончаний имён прилагательных «задни ноги».

У глаголов в форме 3-го лица наблюдается частое употребление флексии –ют: *топят, смотрют*. Необычно употребление глаголов с основой на –г-, –к-: *текёт, ляжете, испекём*. В 1-ом лице: вместо *молочу* – *молотю, захвачу* – *захватю, шучу* – *шутю, топлю* – *топлю* (п//пл), *кормлю* – *кормю* (мл//м)

В области категории рода тоже наблюдаются лексикализованные диалектные явления под влиянием «акающей нормы произношения» такие как переход существительных среднего рода в женский: *тоскует сердца, лукошка просы, солнышка выглянула*. У одушевлённых существительных ср. р. ед. и мн. числа - единичные формы совпадения И.п. и В.п.: «*и соболи и лисицы вытопило водою*», «*зайцы добывали*». В Р. п. мужского и среднего рода наблюдается флексия –у (-ю): *с уезду, против...списку, край берегу, до Томску*. Это было свойственно именам с широким значением. Необычным было и склонение существительных с суффиксом _-ишк-: *сынишка, сынишко, с сынишком, сынишку моему*. Флексия –е осталась в основном у существительных в Д. и П. п. с основой на «і» и «а»: *доче, по Обе, на памяти, по грязи, на Кете холодна вода*. У прилагательных И. и В.п. м. р. наблюдаются окончания –ой (-ей): *кафтан лисей, бедной беззаступной*. Интересно употребление формы глагола 2 и 3-го лица, где происходит стяжение глагольных форм: *убывает (убывает), не хватат, чё стукашь*. Возвратный постфикс –ся и –сь сохраняется традиционно во всех формах: *открылась, принялась, принялся*. Иногда, в редких случаях, может быть –си: *боишьси*.

Диалектную особенность получили причастия: *наделат, настлат, смолон, расколон*, даден или безударный суффикс в говорах – ударный: *распущёны, разрушоны, стрежённы*. В употреблении некоторых деепричастий изменения связаны с фонетическими диалектными признаками, описанными ранее: *убрамши, стирамши* (в//м).

К существительным мужского рода относят в диалекте слова: *берлог, мыши, облак, полотенец, яблок и др.*; к словам женского рода: *кедра, овоща,*

организма, проса, тигра, свитра. Слово «путь» в ряде территорий зафиксировано м.р, у остальных – ж.р.. Нужно отметить, что между тремя формами «кедр – кедрa - кедрo» в среднеобских диалектах сложилось следующее: «кедр» - под влиянием литературного языка, «кедрa» - распространено повсеместно в среднеобских говорах, а «кедрo» (ср. р.) – только в принарымских и прикетских говорах. *Вариант «кедрo» мы наблюдаем в речи шорского билингва как диалектный вариант.* Причины такой диалектной специфики объясняются преобразованием неудобно произносимых согласных звуков на конце иноязычных слов (*метра, литра, кедрa*) или сохранением говорами древних родовых форм (*мыш, облак*). Вариантность диалектных слов сохраняется по причине сложения вторичных языков из материнских. Категория числа в среднеобских говорах организована также как и в литературном языке. Нарушения есть у существительных с собирательным значением: *падат шишка, комар всюду, ударят по кедре.*

В словообразовании существительных и прилагательных в среднеобских говорах отмечаются различия и количественные, и качественные в сравнении с аналогичной системой литературного языка. Качественная характеристика проявляется в наличии или отсутствии определённого словообразовательного явления, а степень распространения словообразующей единицы является количественным показателем. В говорах меньше слов, образованных приставочным способом, и вообще не замечены иноязычные префиксы с книжной окрашенностью, немного в речи жителей Среднего Приобья наблюдается и сложных прилагательных. Некоторые аффиксы в говорах имеют иное значение, не соответствующее значению литературного языка, например, *присталость, хворость, испривеливый, загустенный, затоций.* Кроме того, отмечается иная структура морфемы, в сравнении с литературным языком (*«стекловатый», «меловатый»*). Отличия есть и в продуктивности словообразовательных элементов: в говорах отмечается большая продуктивность суффикса –анн: *бревянный, икрянна (рыба)*. В отличие от литературного языка в среднеобских говорах у существительных не всегда

сохраняется родовая характеристика мотивирующего слова: *житьё - житишка, покрывало-покрывалочка*, также нет строгой словообразовательной зависимости мужского рода от женского: *чистун-чистотка, говорун-говоруха, кашлюн-кашлюнья*. Здесь же отичаем опредмечивание слова. Десемантизация приставок, количественные и качественные изменения аффиксов наблюдаются и в словообразовании глаголов в среднеобских говорах, например, *обсоветовать, распотеть, прочерпнуть воды и др.* Что мы наблюдаем и сегодня в речи билингвов.

Лексическая система, как показывают исследования учёных, в среднеобских говорах не имеет такого постоянства, как фонетическая, словообразовательная и морфологическая. Это объясняется сложностью и своеобразием лексической системы каждого приобского региона, диффузностью лексической семантики, разной лингвистической сегментацией соотносительных семантических полей. Например, *болотина-балахтинка - балахтина – болотце – болотичко; фартук-запон – передник; таёжное озеро: сорник – поньжа – чвор – мурлинка*. Лексическая система русских говоров связана с лексикой и местных аборигенов. Что подтверждают примеры: *наст - чир – чарым - чурым – канасок*. Но вместе с тем видно, что и план выражения слова и план содержания на всех уровнях языка среднеобских говоров взаимосвязаны, имеют общее объяснение и характеристику.

К синтаксическим особенностям говоров Средней Оби относят изменения и сближение в сравнении с литературным языком синтаксических возможностей конкретных и абстрактных существительных, то есть и в синтаксисе наблюдается тенденция к опредмечиванию объекта, его конкретизации в процессе номинации, например: *Мы под этим председателем двенадцать лет жили. Коровы за пастухом ходят. Они в пастухе ходят*. Отмечается в говорах и употребление синонимичных приставок и предлогов: *Он вчера отпросился от работы. Они занимаются за пчёлами*. Многие подобные явления объясняются законом действия синтаксической аналогии. В предложении характерной чертой говора является

отсутствие чётко противопоставленной временной категории глагола: *Сенокос без мяса не жили. Волки эти года отвалились. Нынешнее лето тепло было.* Говорам свойственно употребление союза А в начале предложения: *А ты дрова с ём пойдёшь резать? А автобус видно нынче не придёт.* Зафиксированно преимущественное употребление неопределённо-личных предложений в описании собственной деятельности: *Здесь хлебопашесвом занимались. Хлеб сеяли.* Сложные предложения в диалекте чаще всего передаются без союза: *Год пройдёт – их стригут. Рематизма – мучит – муравьиным ядом, скипидаром лечат. Он один остался, вот его и ухлопали.* Союз так употребляется как дак, дык: *Я уже стара была, дак с меня много не справлялось. Придут, дык все перестынут.* Или отмечаются не точные употребления союзов в зависимости от их значения: *Мне было страшно, потому как чёрный лежит. Не пуцу её с тобой, оттого ты не надёжный парень.* Часто форма предложения не соответствует содержанию. Отмечается также в диалектах частое отсутствие относительных и указательных слов в предложениях. В построении диалектных предложений наблюдается по большей части аналитизм. Синтаксис среднеобских говоров схож с синтаксисом других говоров, имеет общие черты, но есть, как представлено выше, свои семантические и синтаксические особенности. Многие синтаксические конструкции соответствуют литературной норме.

2.2.1.3 Фонетические и морфологические особенности разговорной речи как источник отклонений от речевого стандарта

Как отмечалось ранее в корпусе, материалы которого послужили источником анализа, отмечаются отклонения от речевого стандарта, обусловленные особенностями устной спонтанной разговорной речи, что мы отграничиваем от явлений интерференции.

Анализ устной речи билингов показал, что суперсегментные средства разговорной речи характеризуются рядом особенностей, нередко резко контрастирующих друг с другом. С одной стороны, можно отметить яркость, богатство, разнообразие просодических характеристик устной речи. Характерны и ярко выражены темп речи, пауза, тембр, модуляции тона. Эти признаки интонации позволяют расширить модификации основных типов интонирования. В единстве с лексическими и грамматическими компонентами высказывания, они расширяют его возможные семантические конструкции. В то же время в разговорной речи отмечается "нечёткость", невыраженность суперсегментных единиц. Одной из причин невыраженности суперсегментных показателей является убыстренный темп речи, характеризующий разговорную речь. Он может ослабить или проявиться вообще отсутствием некоторых суперсегментных единиц. Достаточно часто отличия интонирования речи шорцев можно объяснить индивидуальными особенностями речи носителей. Кто-то выражает свои мысли ярко, экспрессивно, используя разные средства языковых уровней: фонетического, лексического (использование образных слов, высказываний, принадлежащих разным речевым дискурсам), словообразовательного (создание незуальных слов) и т.д. Другой респондент говорит более стандартно, речь интонационно не окрашена. Однако и яркость просодических характеристик бывает неоднородной и у одного и того же респондента в зависимости от темы разговора, возрастных различий, настроения, профессии, образования, интересов и др. Эти же особенности наблюдаются в устной речи билинга на всех языковых уровнях, подуровнях и микроуровнях, что свидетельствует о психолингвистических и социолингвистических особенностях говорящих. О проявлении расчлененности разговорной речи в плане выражения может свидетельствовать большее разнообразие суперсегментных средств по сравнению с нормированным литературным языком. Об этом свидетельствуют случаи яркости, контрастности просодики. Явления же

ослабления, "нейтрализации" суперсегментных показателей реализуют тенденцию к нерасчлененности, синкретизму.

В устной разговорной речи при безударности одного из слов такта наблюдается примыкание безударного слова к ударному: энклитики и проклитики, примкнувшие к слову, образуют одно целое образование — фонетическое слово. Оно строится как "монолитное" слово, т.е. с одним ударением, местоположением всех гласных и согласных в необходимых позициях — в состоявшихся градациях тембра, длительности и силы. Только в этих случаях такт представляет собой одновременно и фонетическое слово, например: «*токаздесь*», «*Ониоттуда// инавертолёте. (из речи шорцев)*». В разговорной речи под влиянием различных факторов слова могут изменять ударение (могут быть под ударением, слабоударными или безударными). Они называются динамически неустойчивыми.

Существует два фактора, которые определяют динамическую неустойчивость слов в разговорной речи: 1) ритмический фактор, 2) фактор информативной нагрузки слова.

Слово становится в речи ударным или безударным в соответствии с ритмическим рисунком. Например, в начале фразы: «*Ну, шорцы// и... шорцы, и всё...*». Динамическая неустойчивость союза и (его ударность в первом случае и безударность во втором) объясняется различной ритмической позицией данного союза.

В конце фразы оптимальной для русской ритмики является ударность одного из трех последних слогов фразы. Ударность слов, завершающих фразу, подстраивается под эту ритмическую закономерность, к примеру: ... *у меня младший брат женился, девушку привёл оттуда, со стороны Кабырзы — мы, знаете, как над ней хохотали дома? (из речи билингва)*

В середине фразы наблюдаем динамическую неустойчивость слов, расположенных в середине фразы (т.е. неначальных и неконечных слов фразы). От ритмики одно и то же слово может быть либо ударным или

слабоударным, либо безударным. Следующая фраза записи устной речи тюркоязычного билингва:

Эээ... как же... там, где...я, вишь, к-... был, Кри.. Кристальный, по-моему, я был там три месяца. Кристалл — это... щас Екатеринбург называется, раньше Свердловск.

Интересным является наблюдение за типическими функциями частиц в разговорной речи. Частицы необходимы для организации важных смысловых и стилистических сообщений, выполняют коммуникативную контактную функцию, являются показателями границ текста (частицы, начинающие речь, вводящие новую тему, завершающие речь или тему), служат заполнителями ритмических пауз, используются в качестве показателей актуального членения, как слова-заместители при паузах колебания или обдумывания. И во всех этих и некоторых других функциях они выступают как аналитические средства выражения тех или иных отношений.

Наиболее распространены в речи шорских билингвов по нашим наблюдениям частицы — ну, вот, а, и, да, так, же, вроде, только, прямо, просто, уже, ли, даже и др.

Широкая употребительность частиц объясняется такими ее экстраязыковыми особенностями, как непринужденность и неподготовленность общения, эмоциональность.

В устной речи респондентов соединение слов иногда сопровождается нарушением правил синтаксической, семантической и лексической сочетаемости, которые являются нормой в литературном языке.

Ослабление и снятие периферийных компонентов значения в разговорной речи происходит, как указывалось Ю.Д. Апресяном, в условиях нейтрализации семантических различий. Разговорная речь нивелирует тонкие нюансы значений слов, и именно поэтому возможно такое словоупотребление, какого нет в литературном языке, например [11, с.206]:

Ну, вон в Совете у нас десять человек. Мы вот, на..., мы и говорим, что подходит время выборов, там по уставу через три года, но к этому времени,

допустим, вот щас, вот тоже в конце ноября или в начале декабря будет. Три года уже как пройдёт. Я же, примерно, девять лет уже паптыком работаю. Ну, у нас членов тридцать семь сегодня, на сегодняшнее... тока члены, которые уплачивают. А так в каждом посёлке практически... есь же паптыки свои избранные.

Фонетический состав сегментных единиц, в частности заимствованных слов, меняется в речи билингвов по аналогии с фонетическими процессами, происходящими в просторечии и диалектах. Распространены следующие типы нарушений литературной нормы:

1. Употребление вставного [j] при стечении гласных (*принципы \i \аль-ный, ти\i\анист, комцсси [i]онка, ки\i)оск*). Например: Сигарет бы купить// Да вон ки [i] оск напротив//

2. Метатеза (*леригия, лериквия, леревантный*): *Это же моя лериквия.*

3. Смягчение согласного перед [э] [*резю\м 'э*], *ка[ф'э], хризан [т'э] мы, Пойдет кто ай нет?; Я его видел// — Когда? - Вчерась!//; А. Куда карточку класть-то? Б. Сюды; А. Я тебе позвоню//ан [т'э] нна*)

4. Смягчение з в заимствованных словах на -изм. Например: *Что же ты все работаешь/ работаешь/ головку не жалеешь// Это же прям мазохи[з'м] какой-то//; Ну у тебя прям что ни слово/ афори [з'м]; по пути шамани[з'м]а.*

2.3 Отклонения от речевого стандарта в речи шорско-русских билингвов в аспекте интерферентных влияний

Далее мы приведем типичные примеры отклонений от речевого стандарта в речи шорско-русских билингвов и охарактеризуем по уровням языковой системы и опишем их в аспекте интерферентных влияний.

2.3.1 Отклонения от речевого стандарта в фонетике

На фонетическом уровне нами были выделены такие типы отклонений от речевого стандарта как:

1. Диалектное, свойственное среднеобским говорам «еканье\эканье», где во всех предударных слогах не происходит позиционной мены звука [э], например: *Че [э]псараков Николай Тимофеич. (Ч. М., 81) Спасенский сельский совет, поселок Усть-Се[э]ле[э]зЕнь. Да, колхоз «Че[э]пай» (Чапай) там было (С.О., 79)*

2. Диалектное-просторечное региональное «чёканье»: *«Ну, иголка, - грю, - эт (это) иголка - грю, котор... которой можно шить, там чё-нить, (что-нибудь) а это, - грю, - совсем другое». Или ета..., вот я чё (что) начал раз..., грить... (Ч. М., 81) Так чтоб помочь где-то, на праздники там, чё-то (что-то) кому-то..., посредникам это выписываем там. Ну, и восемнадцать лет он собирал, и потом чё-то... (что-то) трагически погиб, и вот, книга осталась неизданная. (И.М., 60).* Кроме того, [ЧО] и [ШО] в речи респондента могут варьироваться: *На шорском, шо (что) ль, говорить если?.. Ну... ну, так-так конечно, не... не.. сказать, шоб (чтобы) ну, до... этот...до десятого класса там эт... (Ч. М., 81) Так чтоб (штоб)помочь где-то, на праздники там, чё-то (что-то) кому-то..., посредникам это выписываем там. (И.М., 60).*

3. Отклонения, обозначающиеся как собственно диалектные; они встречаются нечасто, в единичных случаях:

1) j протетическое перед [Э]: *Или ета (это)..., вот я чё [Phon] (что) начал раз..., грить... (Ч.М., 81)*

2) наличие твердых долгих [ш:] и [ж:]: *Они отдохнули, щас они меня будут пытать ишишо. (Ч. М., 81).* Это явление обозначает или самостоятельные фонемы, или выделяется в качестве лексико-фонетического (Русские говоры Среднего Приобья) [31, с. 121].

3) Многие диалектные и просторечные фонетические особенности отмечаются лишь в отдельных словах респондентов, которые можно, видимо,

представить в виде списка: *теперича, энто, кажный, слободно, ланно, маво, сваво, поддрит, пуцай, его, хто, жисть, обчество, щас, например:*

Есь (есть), Еси б, (если бы), де-то (где-то), Дак (так) нас около пяти тыщи (тысяч) Но вот из них примерно... тыща (тысяча) девятьсот... тыще (тысяча). А многие там вот, когда это самое..., остались практически шорцы, а русские, которые были пидисят (пятьдесят). Может быть, я у ней (у неё) никогда не спрашивала. Ну вот, стоко (столько) мы с ней намучались... . (И.М., 60); Бригадиром и мастером, начальником работал. Всяко-разно. Я один раз выезжал на три месяца токо (только) — и всё. А так больше никуды (никуда) не ездил. (только) здесь (Ч. М., 81).

4) к диалектному явлению и отклонению от литературной нормы в речи билингвов (Русские говоры Среднего Приобья, ч I) [30] можно отнести и произношение мягкого звука [ж:], например: *приежж'Ие (приезжие), приежж'Яем. (приезжаем), уежж'Яла (уезжала)*(И.М. 60) [30, с. 124]

5) диалектно-просторечным считается смягчение з в заимствованных словах на -изм. Например: *Но, по крайней мере, так получилось, мне кажется, что..., в основном, у нас в семье, ну, в ш..., по шамани[з']му пошли дети.*(И. М., 60.)

4. Интересно употребление шорскими билингвами губно-зубного звука [В]: в более раннем периоде развития языков и шорского, и диалектного русского этот звук произносился как билабиальный [w]; в шорском алфавите отсутствовала буква В (Н.П.Дыренкова «Грамматика шорского языка» [18, с.14], Э. Ф. Чиспияков «Шорская грамматика» [46, с. 8], Русские говоры Среднего Приобья, ч. 1), поэтому отсутствие или замена в речи шорцев приставки (обычно опускается) или предлога В (опускается или заменяется) не случайное явление: *Прошлый год (в прошлом году)... зимой... сsss Новокузнецка, (из Новокузнецка) э..., приезжали две женщины, это, тоже меня вызывали, нашли. (Ч. М., 81). Спасенский сельский совет, поселок (в посёлке). Мы на всех (во всех) районных мероприятиях, мы уже на международных, на всех (во всех) участвуем. Я ещё на подготовительный (в*

подготовительный) там немножко успела походить, один год, и всё. (И. М., 60) [30, с. 96,97]

5. В речи шорских билингвов можно заметить замены [к]и[x]: *Мне нравится икный (их язык.) (С.О., 79); Лекше (легче) вылетает! В охтябре (в октябре) как уехал, ноябрь, декабрь — в январе тока приехал. (Ч.), а также замену[д и т]: Ну, вишь (видишь), эта... , уже давно дак (так) эта... не... (Ч.М., 81) Дак (так) нас около пяти тыщ всего здесь (И. М., 60).*

Эти замены можно классифицировать и как внутриязыковую, так и как межъязыковую интерференцию. Слабая противопоставленность в тюркских [К] и [X], [Г] и [X] привела к заменам [К]/[X]. Звук [Г] в тюркских языках не только взрывной, но он может произноситься как спирант. Подобное фрикативное произношение звука актуально и для русских говоров Сибири, но как южнорусский его вариант [27, с. 107]. Поэтому эту черту можно было бы отнести и на счет русского диалектного влияния, хотя эта особенность в среднеобских говорах отмечается редко, она достаточно узка и направлена на замену к \ х [Образование севернорусского наречия..., с. 247]. Учёные предполагают, «что «слабая противопоставленность» заднеязычного взрывного и спиранта была, наоборот, индуцирована в среднеобских говорах тюркским субстратом» [37, с. 208]

6. В области вокализма можно отметить сильную, вплоть до выпадения, редукцию узких гласных Ы и И в первом слоге (это встречается не часто), также подобное произношение можно объяснить и наличием в шорском языке долгих гласных: *прыск, (прииск) (С.О., 79); ср. шорский язык: кол – «рука», коол – «пустой».*

7. К отклонениям от речевого стандарта в области фонетики можно отнести и постановку ударения: с одной стороны, это можно объяснить тюркской особенностью функционирования ударения (всегда на последний слог в слове), а с другой – подвижностью и свободностью ударения русского языка. Отсюда в речи шорских билингвов наблюдаются отклонения от речевого стандарта типичные и для носителей русского языка ошибки и

относящиеся к тюркскому произношению, например: Эээ... Спасенский сельский совет, поселок Усть-Селезень (СЕлезень). С языком учиться надо дитЯм (дЕтям) молодым. Ой, я на рУднике (на рудникЕ) стрелочницей работала, работала-а, ааа, когда уехала, там на заводе работала и на шахте работала. Все, мне звОнят (звонЯт) (С.О., 79) Уже дошло до того, что мы Путину началИ (нАчали) писать, ответ оттуда приходит. Многие же приежжяют (приезжают) из северОв,(сЕверов) сюда домой возвращаются. Ну, к кондОмскому (кОндомскому) диалекту я отношусь. Образовался в связи с дОбычей (добЫчей) золота. (И.М.,60)

8. В системе вокализма тюркских языков есть особенность: гласные различаются, в отличие от русского языка, степенью палатализации, и это явление, как и основной закон тюркских языков, сингармонизм во многом объясняет следующее отклонение от стандарта, например: СамойловЫч (Самойлович), мамына (мамина) Я кЫргЫзский-то (киргизский) сдавала же, в институте же своём. (С.О., 79); Ну, старший – Чепсараков... Владислав Михайловыч (Михайлович). А... года два, навэрно, (наверно)... тому назад приезжали.«Ну, вроде как матрас надо вытаскавать, (вытаскивать) стелить надо». (Ч. М., 81)

9. Устная речь русско-шорских билингвов характеризуется и общепринятыми правилами русской разговорной речи, так называемой неканонической фонетикой (редукцией гласных): А многие там вот, когда это самое..., остались практически шорцы, а русские, которые были пидисят (пятьдесят). (И.М., 60)); Э, это... и вот, он мне читает: «Мож, - грит, (говорит) - так? Можт, эдак? Можт, вот так слово?». Я грю (говорю): «Нет, нет. Нет, – грю, (говорю) – не подходит». Короче, у нас вот так вот. Разногласие, как грится. (говорится). Ну, и это... вот, я грю, (говорю) три слова только и не смог. (Ч.М., 81). б) изменением порядка слов, парцеллирующей: Вот, три слова, вишь, (видишь) это... нечасто разговариваешь, не общаешься часто – забываются. (Ч.М.,81.) Мы вот, на..., мы и говорим, что подходит время выборов, там по уставу через три года, но к этому времени, допустим,

вот щас, вот тоже в конце ноября или в начале декабря будет. Три года уже как пройдёт. (И. М., 60).; в) слова-средства ритмической организации речи: Ну, там вот в этих... есть, а вот допустим, Мундыбаиш, который оттуда едешь уже, в Казе, в Темире — уже там практически мало совсем.(И. М., 60); Э, это... и вот, он мне читает: «Мож, (может) - grit, (говорит) - так? Можт, (может) эдак? Можт, (может) вот так слово?». Я грю (говорю): «Нет, нет. Нет, – грю, (говорю) – не подходит». Короче, у нас вот так вот. Разногласие, как gritся. (говорится). Ну, и это... вот, я грю, (говорю) три слова только и не смог. (Ч.М., 81).

Разговорная речь шорцев связана с особенностями спонтанной речи, но можно отметить в речи шорцев и их индивидуальность, раскрывающуюся посредством коммуникации с интервьюером и групповым общением в билингвальной среде. Используя языковые средства ритмической организации речи (например, один использует указательные местоимения ето / это, у другой респондентки), отмечаем частицу ну, у третьей — вот.

Таким образом, можно сделать вывод, что в устной речи билингвов отмечается интерфернтное влияние закона слогового сингармонизма; а также ассимилятивные явления и связанное (не свободное) ударение; активна и внутренняя интерференция: влияние говоров и диалектных просторечий среднеобских говоров (а они в свою очередь, как писалось выше, сложились из разных, существовавших в России говоров: южнорусских, севернорусских и среднерусских).

2.3.2 Отклонения от речевого стандарта в грамматике

1. Стяжение формы флексий имен прилагательных и прилагательных-местоимений: *Ну, русские, понятно (понятное) дело. Да мне кака (какая) разница, ёлки-палки? Кто встретится — с тем и...всяко-разно (по-разному) поговоришь. И чё тока ни это такое.... Вот-вот эти, которы (которые) из Франции приезжали. (Ч.М., 81).*

2. Вариантные возвратные формы глагола с СЬ/СЯ: *Они на русском училися (учились), по-русски и по-русски. Даже у меня вот сестренка родилися (сестрёнки родились), я старалася(старалась) ними по-русски, чтобы они знали даже вот у меня дочка училася (училась) на-а... . . . высшее образование хотела получить. Старалася (старалась) только по-русски говорить. (С. О., 79). В школу пойти, и вот тут как раз школа-то и закрылася. (закрылась). Поэтому как-то общалися. (общались) (И. М.,60) Ну, вот здесь (здесь) маленько понятно!(Ч.М.,81)*

3. Следующие варианты диалектных особенностей в области морфологии, ряд из которых является особенностью и городского просторечия, проявляются в речи возрастных респондентов:

1) предлог меняется на С в родительном падеже: *Они потихоньку потом ушли с ансамбля (из ансамбля), потому что работа большая. А откуда везут — с России (из России) ж это всё. Сначала наш шаман встаёт, счас духовных лиц с молодёжи (из молодёжи) очень... много. И с Карбалык, (из Карбалыка) и там с мелких, (из мелких) с этого, (из этого) со всей округи (из всей округи) ходили, из Базаса, тоже оттуда, от..., с Ворошиловки, (из Ворошиловки) со всех (из всех) ходили в Сайзак. И всё, всё размещается в газете, на телевидении, и... м-мы определяем, сколько делегатов с каждой (из каждой) территории будет, в зависимости от численности шорцев, проживающих в том или ином посёлке. (И.М.,60). Он оттуда же, с деревни (из деревни). Прошлый год (в прошлом году)... зимой... сссс Новокузнецка, (из Новокузнецка) э..., приезжали две женщины, это, тоже меня вызывали, нашли. Ну, э... видать, наверно, знают, что я, наерна... последний этот... мокиган осталсА ... и я пок... вот..., с Новосибирска (из Новосибирска) вот были, с Новокузнецка... (из Новокузнецка), эээ... ой, с н... ну да, с Новокузнецка, (из Новокузнецка) с Новосибирска, (из Новосибирска), с Франции (из Франции) были, ну, и ещё какие-то были вощем...(Ч.М., 81).*

2) вариантыные падежные формы местоимений (проявляются не часто в речи билингвов) они / оне, меня / мене: *Ну, тахтэ, оне (она) грит : «Давай грит тахтэ - грит, - это... разбирать».* (Ч.М., 81).

4. В тюркских языках и в шорском, в частности, отсутствует категория грамматического рода и согласование определения с главным словом. Соответственно, естественными следствиями шорского влияния на русскую речь билингвов являются ошибки в согласовании: *А они уже, оказывается, там, и в Кабырзе были, в трёх или в четырёх уже посёлках - и везде по-разному эта слова написано (эти слова написаны) Или эта (это)...., вот я чё (что) начал раз..., грить... (Ч.М., 81); Потом женщины - от старшему к младшему (от старшей к младшей)(И.М., 60) Да, колхоз «Чепай» (Чанай) там было (был).(С.О., 79).*

Имеет место и разграничение слов по категории рода (мужской, средний, женский и общий). Хотя выделяются существительные с диалектными родовыми признаками: мужской род - берлог, клиник, крылец, мыш, облак, полотенец; женский род — овоща, организма, свитра, тигра и др.; средний род: коноплѣ, фамилиѣ и др.. Но есть и слова, которые употребляются в сибирских говорах во всех трёх родах, к примеру: кедр (м. р.) – кедр (ж.р.) – кедро (ср.р.). В речи билингва встречаем данное слово в среднем роде с ударением на последний слог. Это можно охарактеризовать как лексико-морфологическое явление в речи, связанное с диалектными особенностями жителей Среднего Приобья [27, с. 144-147]: *Они вошли недавно сюда, но вырубил практически вс..., всё кедро. Ну, а потом, если кедро-то вырубил, не кедро, а даже вокруг даже кедр там всё вырубил, то кедр сам упадѣт. Потому что он же... ему, у нег корневая система-то не так глубоко уходит, как у других, и он за счёт других деревьев тоже поддерживают себя.* (И. М., 60)

Можно предположить, что отклонения в речи шорских билингвов связаны как с отсутствием в грамматике тюркских языков категоризации по роду, так и с общими языковыми явлениями, связанными с формами существования рода в речи жителей Среднего Приобья.

5. Категория числа в тюркских языках представлена единственным и множественным числом, но употребление форм времени в контексте и взаимодействии с другими языковыми параметрами отлично от согласования слов данной категории в русском языке. Так единственное число не обозначается в тюркских языках и не всегда является обязательным при согласовании слов по числу в тюркской речи. Если количественный показатель количества не является в высказывании чем-то важным, то вместо него может быть использована форма множественного числа. *А они уже, оказывается, там, и в Кабырзе были, в трёх или в четырёх уже посёлках — и везде по-разному эта слова написано. Только это... эээ... русские эти литературы (русскую литературу) читаешь. (Ч.М., 81) Она же сходится слова (они слова сходятся). После войны это... нам очень тяжело было, дети войны же я (я дитя войны/ мы дети войны) Я целый год училась... и во вторых... осталась (на второй) (осталась). Даже у меня вот сестренка родились (сестрёнки родились) (С.О., 79). А потом уже, когда он шаманил, допустим, щас же шаман — это немножко... другие (это другое). Она как это..., преждевременные роды. В начале-то у нас писала Надежда Алексеевна, Нина Алексеевна. (И.М., 60)*

Хотя подобные отклонения от речевого стандарта можно отнести не только межъязыковым, но и к диалектным, потому что в среднеобских говорах принято использование единственного числа вместо множественного и можно также отметить специфичность формы множественного числа, которыми обозначаются наименования по родству в среднеобских говорах.

6. Падежная категория существительного определяется и в русском, и тюркских языках, но отличается формальными показателями и грамматическим значением. Падеж в тюркских языках и число не всегда бывают выражены грамматически. Так как в шорском языке существуют послелого как языковая форма, то большинство слов употребляется с послелогом в именительном падеже, например: *Я родилась в Таштагольский район (в Таштагольском районе). Эээ... Спасенский сельский совет, поселок*

Усть-Селезень (С.О, 79); . По русский язык, только по русский язык (на русском языке, говорили по-русски) и все. (С.О., 79); б) в позиции приименного определения, в том числе в количественной конструкции: В Усть-Селезенский сельский совет. (В Усть-Селезенском сельском совете). А школа - русский язык, конечно (в школе). Они не знают. По русский язык, только по русский язык (на русском языке, говорили по-русски) и все. На ээ... алтайском? Алтайский кто со мной разговаривает - да. С языком (языку) учиться надо дитЯм молодым. (С.О.,79) в) прямое и косвенное дополнения: Ну,она у меня это..., ДЦП была, (она была больна ДЦП) поэтому щас вот ещё хорошо ходит. Старшему — семь, младший (младшему) уже, наверное, год и восемь... (И.М., 60)

7. Употребление послелогов и предлогов. Отклонения от речевого стандарта.

В тюркских языках предлоги заменяются послелогамии. Этим же можно объяснить и диалектное опущение или замену в синтаксических конструкциях билингвами предлога со значением «места, времени или направления объекта или на объект», например: *Когда тайком надо от ребяташек, чё-нибудь (о чём-нибудь) поговорить, тоже по-шорски разговаривали, чтобы они не понимали Я целый год училась... и во вторых... осталась (на второй)(осталась). . По русский язык, только по русский язык(на русском языке, и всё.(С.О.,79). Спасенский сельский совет, поселок (в посёлке). Нет, у нас золотарный, эээ... алтайский какой-то прыск, поселок Спасск, со Спасском (в Спасском) работали. Тридцать девятом году (в тридцать девятом году.) Прошлого года (в прошлом году) другие были. Прошлый год (в прошлом году)... зимой... сссс Новокузнецка, (из Новокузнецка) э..., приезжали две женщины, это, тоже меня вызывали, нашли. Татар, манкурт, кумык, лезгин, удмурт, чуваш, саха... мордва (у татаров, манкуртов,.....). (Ч.М., 81). Когда вот техникум кончила, поехала в деревню. Отдохнуть... мужу. Но, честно говоря, потом выводы (с выводами) этой экологической экспертизы... , хотя деньги тратятся-то... нормальные, я имею в виду..., не знакомят они*

нас. С национальным парком здесь уже... две тыщи (с две тысячи) одиннадцатого года мы всё это тоже.... (И. М, 60).

8. Интересной межъязыковой интерференцией можно считать отклонение от речевого стандарта в русской речи шорскими билингвами употребление видо-временных категорий, так как в шорском языке таковых насчитывается 10, а в русском – 3 формы времени и две категории вида, например: *Уже дошло до того, что мы Путину началИ писать, ответ оттуда приходит (пришёл). Но теперь уже, как бы, немножко подвижки пошли. Три года уже как пройдёт. (И.М., 60).*

В области грамматики также мы наблюдаем межъязыковую и внутриязыковую интерференцию, которая связана а) с агглютинацией: формы слова в тюркских языках образуются наращиванием элементов, а не изменением формы слова, аффикс имеет одно значение, в отличие от флективного русского языка, где есть изменяемая многозначная флексия, каждая категория флективного строя языка участвует, в отличие от аффиксов тюркских языков, в согласовании и управлении словами разных синтаксических конструкций; б) оказывает влияние на отклонения от речевого стандарта отсутствие категории рода в тюркских языках; в) слабовыраженное явление грамматикализации в тюркских языках в сравнении с языками флективного строя, но выражено ярко во всех языках проявление лексикализации.

Некоторые явления, сложившиеся исторически под влиянием разных по своему строю и диалектных влияний языков разных народов, имеют общие проявления и выражения, поэтому бывает сложно определить интерферентные влияния и их закономерности.

1.3.3 Отклонения от речевого стандарта в лексике.

В области лексики отмечаем следующие явления:

1) диалектные межсистемные синонимы: *Она же сходится слова (слова похожи). (С.О., 79); Хотя вот раньше обучали..., а уже мне не досталось.*

(М.А.); Пошла в подземку (в шахту) работать. Ну, мы как-то обращались знакомые (общались) только с шаманами. (С.О., 79); По поступлениям в ВУЗы..., ну, с этими..., с спиртовозом (неблагополучные семьи) тоже работали. (И. М., 60).

2) лексико-фразеологические диалектизмы: система конструкций, выражающая отрицание возможности осуществления события, утвердительные или вопросительные конструкции: «Кого (что) ты собралась, - грю. И вот кто учителем, вот это... выучился — он ток будет на свой язык тянуть. (Ч.М., 81).

3) просторечно-диалектная лексика: в речи шорцев употребляются как приметы речи крестьянской отдельные лексемы, чаще всего слова, имеющие отвлеченную семантику и широкую сферу распространения, типа наречий: вчерась, надьсь, тама, туды, сюды и под), например: Ну, у нас членов тридцать семь сегодня, на сегодняшнее... тока (только) члены, которые уплачивают. Да мне кака (какая) разница, ёлки-палки? Детям-то нельзя было присутствовать шибко. (И. М., 60). Всяко-разно (по-разному) — по-шорски и по-русски. . Ну, будут тада (тогда) разговаривать между... между собой на том языке. Кто встретится — на таком и бушь (будешь) говорить. (Ч.М., 81).

4) жаргонная лексика: Мы вот всё тусуемся. (И. М., 60).

5) наиболее заметными проявлениями интерференции на лексическом уровне являются заимствования, которые довольно часто рефлексированы в данном статусе носителями тюркско-русского билингвизма: Встретимся, для шутки, по-своему «кайды чан», ну вот так вот. (С.О., 79); Ну, **паитык** — это вот для нас как ... **паш** — **ты, тык**, ну как бы голова, да? Не крепкая совсем, **арачка**... И они вот, как раз вот, как говорят наши, тоже, что... надо держать шамана, когда он **камлает**. **Чыл-паже** у нас есть... день этой, м-м... день это весеннего солнцестояния это, да. Ну, как говорят историки, что шорцы назывались...(назывались) по... наименованию самого большого рода **Шора**, поэтому. Говорим «**таг эзи**» - дух горыю. А вот, допустим, как посёлок

Мундыбааш, да? У него есть наименований перевода. Допустим, Мыньдабааш — это «здесь глава», потом Мунбааш — это «тысяча голов», и ещё там есть некоторые. Ну, а потом ещё ленты-чаламы после этого, как покормим огонь, да, уже, благопожелания. (И. М., 60). Ну, то есть, тахтэ — это пол. Эна — как мать, она. Эна — видишь (Ч. М., 81).

Таким образом, в устной разговорной речи шорско-русских билингвов на всех языковых уровнях (фонетическом, лексическом и грамматическом), их подуровнях и микроуровнях отмечается внутриязыковая – общерусских разговорных явлений и русских говоров среднеобского типа – и межъязыковая – влияние исходного шорского языка в его диалектной форме – интерференция.

В результате анализа отклонений от речевого стандарта в устной речи шорско-русских билингвов была составлена таблица частот выявленных отклонений (Таблица 1).

	ИМА	СОЕ	ЧМН
DiscHes	275	97	224
CodeSw	46	1	14
SyntLess	41	44	44
SyntPrep	22	9	17
SyntCon	3	1	0
MorphAff	11	9	28
Lex	13	4	2
SyntagrNum	1	4	0
SyntagrGen	1	1	1
SyntGov	4	10	12
SyntNum	2	0	2
LexsemAgr	0	2	0
LexSem	0	2	0
SyntStruct	0	2	0
Phon	0	149	2

SyntagGov	0	1	0
PhonAcc	0	3	0
MorphStem	0	1	0
MorpHnum	0	0	1
SyntagCase	0	0	1
SyntagRasp	0	0	1

Таблица 1 - Частотная таблица ОРС в устной речи шорско-русских билингвов

Из Таблицы 1 видно, что наиболее частотными отклонения являются: DiscHes, CodeSw, SyntLess, SyntPrep, MorphAff.

Визуализируем в виде гистограммы частоту ОРС в устной речи шорско-русских билингвов (Рисунок 4):

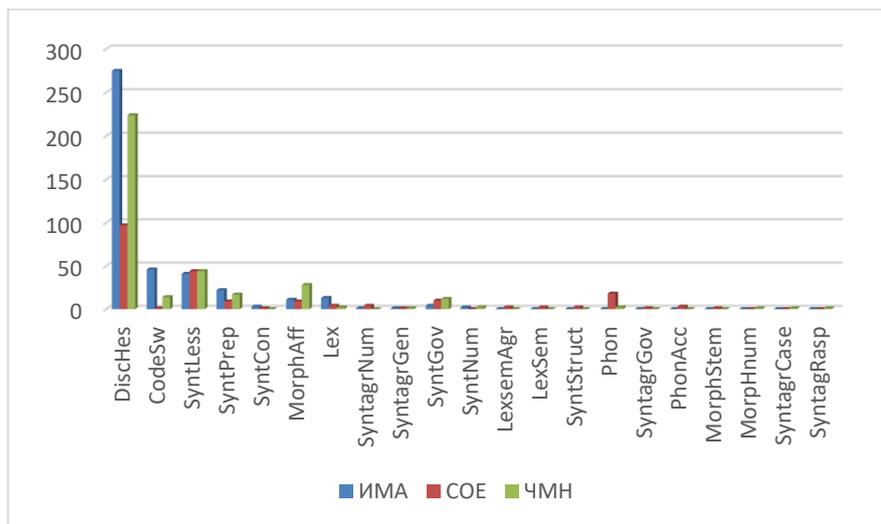


Рисунок 4 - частота ОРС в устной речи шорско-русских билингвов

Заключение

Проведенное магистерское исследование описывает современное состояние корпусной лингвистики, которая начинается с традиционных исследований в национальных корпусах и получает новые вехи развития в узкоспециализированных. Последние представлены сегодня несколькими популярными проектами: «Исследование контактно-обусловленной русской речи малых народов России» и «DagRus». Их актуальность поддерживается постоянным пополнением экспедиционными записями устной речи, что дает возможность описывать текущее состояние и развитие малых языков Российской Федерации и проводить квантитативный (частотный) анализ отклонений от речевого стандарта и выявлять закономерности их использования в речи билингов. Данная работа основывается на методах и идеологии выше обозначенных корпусов и берет за основу своего исследования результаты работы Лаборатории лингвистической антропологии Национального исследовательского Томского университета, где ведется работа по сбору и оцифровке экспедиционных материалов, теоретизация и типологизация тегов отклонения от речевого стандарта в устной речи тюркских билингов в трех языковых уровнях. Данное исследование является анализом устной речи шорско-русских билингов и позволяет сделать следующие выводы:

1. Современные корпусные исследования осуществляются с опорой на междисциплинарные методы исследования, что повышает скорость и надежность изучения отклонений от речевого стандарта в речи билингов.

2. В результате анализа трех текстов шорско-русских билингов (общий объем которых составил более 35000 слов) были обозначены основные типы отклонений от речевого стандарта и составлена их частотная матрица. Что, в свою очередь, позволяет выдвинуть предположение, что подобные характерные отклонения характерны на синтаксическом и фонетическом уровнях. Чтобы подобное заключение было достоверным, важно дальнейшее

пополнение материала и применение в исследованиях количественных методов корпусной лингвистики.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1 Adolphs S. Carter R. Spoken Corpus Linguistics: From Monomodal to Multimodal / Adolphs S. Carter R., London: Routledge, 2013.

2 Bayda K. Kholodilova M. Kozhemjakina A. Romanova E. Remizova T. Storozheva A. Tarasova N. Zorina A. Morozova V. Panova A. Dobrushina N. ChuvashRus Corpus, Moscow: Linguistic Convergence Laboratory, NRU HSE [Электронный ресурс]. URL: <http://www.parasolcorpus.org/chuvashrus/>.

3 Bonacchi S. Karpiński M. Remarks about the use of the term “multimodality” // Journal of Multimodal Communication Studies. 2014. (1). С. 1–7.

4 Chafe W. The pear stories: Cognitive, cultural, and linguistic aspects of narrative production, Norwood. / Chafe W., 1980.

5 Daniel M.A. Dobrushina N.R. A corpus of Russian as L2: the case of Daghestan [Russkij jazyk v Dagestane: problemy jazykovej interferencii] // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”. 2013. № 12 (1). С. 186–211.

6 Gibbon D. Mertins I. Moore R.K. Handbook of multimodal and spoken dialogue systems: Resources, terminology and product evaluation. , 2000. / Gibbon D. Mertins I. Moore R.K., Berlin: Springer, 2000.

7 Khomchenkova I.A. Pleshak P.S. Stoyanova N.M. The Corpus of Contact-Influenced Russian of Northern Siberia and the Russian far East // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2019”. 2019. С. 276–288.

8 Mustajoki A. Protassova E. Vakhtin N. Instrumentarium of linguistics: sociolinguistic approach to nonstandard Russian. / Mustajoki A. Protassova E. Vakhtin N., 2010.

9 Scollon R. Multimodality and the language of politics // Encyclopedia of language and linguistics. 2006. (9). С. 386–387.

10 Т. McEnery С., Gabriellatos English Corpus Linguistics // The Handbook of English Linguistics: McMahon-Blackwell Publishing, 2006. С. 33–72.

11 Апресян Ю.Д. Богуславский И.М. Иомдин Б.Л. Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы // Национальный корпус русского языка: 2003—2005. 2005. С. 193—214.

12 Вайнрайх У. Языковые контакты. Состояние и проблемы исследования / Вайнрайх У., Высшая шко-е изд., Киев:, 1979. 259 с.

13 Виноградов В.А. Языковые контакты в Восточной Африке: суахили и луганда // Социо- и этнокультурные процессы современной Африки. 1992. С. 118–131.

14 Володарская Э.Ф. Заимствование как отражение русско-английских контактов // Вопросы языкознания. 2002. № 4. С. 96–118.

15 Выренкова А.С. Полинская М. С. Рахилина Е.В. Грамматика ошибок и грамматика конструкций: «эритажный» («унаследованный») русский язык // Вопросы языкознания. 2014. (3). С. 3–19.

16 Гаджиева Н.З. Серебренников Б.А. Сравнительно-историческая грамматика тюркских языков. Синтаксис / Гаджиева Н.З. Серебренников Б.А., Наука-е изд., М:, 1986. 286 с.

17 Дмитриев Н.К. Грамматика башкирского языка / Дмитриев Н.К., Ленинград: Академия Наук СССР, 1948. 276 с.

18 Дыренкова Н.П. Грамматика шорского языка / Дыренкова Н.П., Москва:, 1941. 309 с.

19 Жирмунский В.М. Немецкая диалектология / Жирмунский В.М., М:, 1956. 635 с.

20 Захаров В.П. Корпусная лингвистика: учеб.-метод. Пособие / Захаров В.П., СПб.:', 2005. 48 с.

- 21 Казак М.В. Паузы хезитации в спонтанной речи на родном и неродном языках (на материале речи франкофонов): дисс. маг. лингв. 2015.
- 22 Кибрик А.А. Просодические портреты говорящих // Рассказы о сновидениях: корпусное исследование устного русского дискурса. 2009. С. 478–487.
- 23 Кибрик А.А. Русский мультиканальный дискурс. Часть I. Постановка проблемы // Психологический журнал. 2018.
- 24 Копотев В.М. Введение в корпусную лингвистику / Копотев В.М., Прага:, 2014. 195 с.
- 25 Кравцов Н.И. Славянский фольклор: учебное пособие / Кравцов Н.И., М: Изд-во Моск. ун-та, 2009. 352 с.
- 26 Крысин Л.П. Современный русский язык: Активные процессы на рубеже XX-XXI веков / Крысин Л.П., 2008. 712 с.
- 27 Ли Х.-Й. Особенности реализации согласных в русской речи тайваньцев: дисс. ... канд. филол. наук. СПб. 1999.
- 28 Ляшкова И.П. Русский язык для мигрантов: РКЯ или РКИ? (к вопросу о создании термонов) // Международная филологическая конференция. 2009. (21). С. 181–186.
- 29 Николаева Ю.В. Жесты, которые оформляют устный рассказ // Когнитивная наука в Москве: новые исследования. Материалы конференции Е. В. Печенкова, М. В. Фаликман (ред.). 2017. С. 545–549.
- 30 Палагина В.В. Русские говоры среднего Приобья. Часть 1 / Палагина В.В., Томск: Издательство Томского университета, 1984. 203 с.
- 31 Палагина В.В. Русские говоры среднего Приобья. Часть 2 / Палагина В.В., Томск:, 1989. 324 с.
- 32 Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном

освещении. 2008. № 16 (2). С. 7–20.

33 Постовалова В.И. Картина мира в жизнедеятельности человека // Роль человеческого фактора в языке: Язык и картина мира. отв. ред Б.А. Серебренников. 1988. С. 8–69.

34 Пужаева С. Ю. Зевахина Н. А. Джакупова С. С. Контаминация конструкций в речи нестандартных русскоговорящих на материале корпуса русских учебных текстов // В кн.: Труды Международной научной конференции «Корпусная лингвистика-2015». 2015. С. 390–397.

35 Резанова З.И. Подкорпус устной речи русско-тюркских билингвов южной сибери: типологически релевантные признаки // Вопросы лексикографии. 2017. (11). С. 105–118.

36 Резанова З.И. Веснина Г.Ю. Подкорпус русской речи билингвов лингвистического корпуса «Томский региональный текст»: принципы разметки и метаразметки корпуса // Вопр. лексикографии. 2017. № 9 (1). С. 29–39.

37 Резанова З.И. Дыбо А.В. Языковое взаимодействие в речевых практиках шорско-русских билингвов Южной Сибири // Известия Уральского федерального университета. 2019. № 187 (2). С. 195–212.

38 Резанова З.И. Некрасова Е.Д. Миклашевский А.А. Исследование психолингвистических и когнитивных аспектов языкового контактирования в проекте «Языковое и этнокультурное разнообразие Южной Сибири в синхронии и диахронии: взаимодействие языков и культур» // Русин. 2018. № 52 (2). С. 107–117.

39 Резанова З.И. Темникова И.Г. Артеменко Е.Д. Степаненко А.А. Дацюк В.В. Дыбо А.В. Бимодальный корпус устной Речи русско-тюркских Билингвов (rutubic) // Компьютерная лингвистика и интеллектуальные технологии По материалам ежегодной международной конференции «Диалог» (2019). 2019. № 18. С. 200–210.

40 Резанова З.И. Темникова И.Г. Некрасова Е.Д. Динамика социолингвистических процессов в Южной Сибири в зеркале билингвизма (шорско-русское и татарско-русское языковое взаимодействие) // Вестн. Том. гос. ун-та. 2018. (436). С. 56–68.

41 Рыков В.В. Лекций по корпусной лингвистике [Электронный ресурс]. URL: <http://rykov-cl.narod.ru/c.html> (дата обращения: 20.05.2020)

42 Селищев В.А. Диалектологический очерк Сибири / Селищев В.А., Иркутск:, 1921. 294 с.

43 Ушаков Д. Соколов Н. Краткий очерк возникновения Московской диалектологической комиссии и ее деятельности за первое десятилетие (1904-1914 гг.). 1914 г. // Д.Н. Ушаков. Русский язык. 1995. С. 26.

44 Федорова О. В. О русской жестикуляции с лингвистической точки зрения (к выходу монографии Е. А. Гришиной) // Вопросы языкознания. 2018. № 5. С. 114–123.

45 Фрей А. Грамматика ошибок. (Систематическое описание речевых отклонений от литературной нормы французского языка) / Фрей А., 2006. 304 с.

46 Чиспияков Э.Ф. Учебник шорского языка / Чиспияков Э.Ф., Кемеровско-е изд., Кемерово:, 1992. 318 с.

47 Шаров С.А. Представительный корпус русского языка в контексте мирового опыта // Научно–техническая информация. 2005. № 2 (6). С. 12–16.

48 Тенишев Э.Р. Сравнительно-историческая грамматика тюркских языков. Лексика / Тенишев Э.Р., Дыбо А.В., Москва: Наука, 2001. 908 с.

49 Тенишев Э.Р. Сравнительно-историческая грамматика тюркских языков. Региональные реконструкции / Тенишев Э.Р., Дыбо А.В., Москва: Наука, 2002. 767 с.

50 Национальный корпус русского языка [Электронный ресурс]. URL: <http://www.ruscorpora.ru/new> (дата обращения: 18.05.2020)

51 The Brown Corpus [Электронный ресурс]. URL: https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html (дата обращения: 10.05.2020)

52 Lancaster-Oslo/Bergen Corpus (LOB) [Электронный ресурс]. URL: https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/private/LOB/lob.html (дата обращения: 10.05.2020)

53 Синтаксически размеченный корпус русского языка (СинТагРус) [Электронный ресурс]. URL: <http://www.ruscorpora.ru/new/instruction-syntax.html> (дата обращения: 28.05.2020)

54 The Corpus of Professional Spoken American English (CPSA) [Электронный ресурс]. URL: <http://www.athel.com/cpsa.html> (дата обращения: 28.05.2020)

55 The Michigan Corpus of Academic Spoken English (MICASE) [Электронный ресурс]. URL: <https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple> (дата обращения: 28.05.2020)

56 The Epistolary Corpus of Victorian Women Writers 'Letters [Электронный ресурс]. URL: <https://letrs.sitehost.iu.edu/vwwp/index.html> (дата обращения: 03.05.2020)

57 The Shakespeare Corpus [Электронный ресурс]. URL: <https://www.opensourceshakespeare.org/> (дата обращения: 28.05.2020)

58 С Корпус словаря языка Достоевского [Электронный ресурс]. URL: <http://slovari.ru/default.aspx?s=0&p=307> (дата обращения: 28.05.2020)

59 Корпус русских учебных текстов (КРУТ) [Электронный ресурс]. URL: <http://web-corpora.net/CoRST> (дата обращения: 28.05.2020)

60 Русский мультимедийный дискурс [Электронный ресурс]. URL: <https://multidiscourse.ru/main/> (дата обращения: 28.05.2020)

61 Русский учебный корпус (Russian learning corpus) [Электронный

ресурс]. URL: <http://web-corpora.net/RLC> (дата обращения: 28.05.2020)

62 The corpus of contact-influenced russian of northern siberia and the russian far east URL: http://web-corpora.net/tsakorpus_russian_nonst/index.html (дата обращения: 28.05.2020)

63 DagRus (Дагестанский корпус) URL: <http://www.parasolcorpus.org/dagrus/#!/> (дата обращения: 28.05.2020)

Приложение А
(справочное)

Пример разметки отклонений от речевого стандарта
В устной речи шорско-русских билингвов

transcript_resp1 Вот одна, просто ёлочная иголочка — как, как её перевести? Они тоже говорят: «Ну, [DiscHes] иголка, мол». «Ну, иголка, - грю, [Phon] (говорю) – эт[Phon] (это) иголка – грю, [Phon] (говорю) – котор... которой можно шить, там чё-нить, [Phon] (что-нибудь) а это, - грю, [Phon] (говорю) - совсем другое». Или ета [Phon] (это)..., [DiscHes] вот я чё [Phon] (что) начал раз..., грить... [Phon] (говорить) Вот я када-то[Phon] (когда-то) та... вот эти не мо... [DiscHes] три слова, не мог перевести — он берёт этот...ну, [DiscHes] тоже у него бумага така [MorphAff] (такая) толстая — этот...[DiscHes] [SyntLess] А они уже, оказывается, там, и в Кабырзе были, в трёх или в четырёх уже посёлках — и везде по-разному эта слова написано. [MorphNum] (эти слова написаны) Одно и то же слово, например, ну, [DiscHes] например : стол, да — а по деревням[SyntGov] (в деревнях) они - по-разному...

transcript_resp1 ... записаны. То есть, у нас нету, как грится, [Phon] (говорится) одинаково слова. [SyntGov] (одинаковых слов)

transcript_resp1 Э, это... и вот, [DiscHes] он мне читает: «Мож, [Phon] (может) - грит, [Phon] (говорит) - так? Можт, [Phon] (может) эдак? Можт, [Phon] (может) вот так слово?». [SyntLess] Я грю [Phon] (говорю): «Нет, нет. Нет, – грю, [Phon] (говорю) – не подходит». Короче, у нас вот так вот. Разногласие, как грится. [Phon] (говорится). Ну, и это... вот, [DiscHes] я грю, [Phon] (говорю) три слова только и не смог... .

transcript_int А вот они как Вам... когда из Франции приезжали — они, как с Вами получается, на русском с Вами общались или как?

transcript_int Когда из Франции приезжали — на русском Вы общались с ними?

transcript_resp1 Ну, [DiscHes] конечно.

transcript_int То есть, они как бы говорят, да, на говорили на русском.

transcript_resp1 Ну, они, видать, наверно, этот, э, сама..., [DiscHes] наверно, чё-то [Phon] (что-то) знает (знают) [SyntAgrGen] маленьчко. Я чё-то [Phon] (что-то) не это...поин..., [DiscHes] не поинтересовался.

transcript_int А они, получается, передачу какую-то снимали, или, ну то есть как, материал какой-то?

transcript_resp1 Возможно, не знаю, чё [Phon] (что) они, для чего, тоже чё-то [Phon] (что-то) ездили, всё спрашивали. Прошлый год [SyntGov: SyntPrep] (в прошлом году)... зимой... ссс [DiscHes] Новокузнецка, [SyntPrep] (из Новокузнецка) э..., приезжали две женщины, это, [DiscHes] тоже меня вызывали, нашли. Приходил к ним, тожа [Phon] (тоже) разговаривали. Переводил им.

Приложение Б
(справочное)

Типовые примеры отклонений от речевого стандарта

Тип ОРС	Вариант ОРС	Контекст	Кол-во
Phon	Phon	Sh0021_001 у нас золотарный, эээ... алтайский какой-то <i>прыск</i> , (прииск) поселок Спасск <i>СамойловЫч</i> (Самойлович)	2
		Sh007_002 transcript_resp1 ... они тож (тоже) шорцы, но у них язык другой. Вот. В сторону Кабырзы, за Таштаголом там — у них тоже друг они шорцы тож. (тоже) В октябре[Phon] (в октябре) как уехал, ноябрь, декабрь — в январе тока приехал. Ну, вишь [Phon] (видишь), эта...[DiscHes] хе-хе, уже давно дак [MorphAff] (так) эта... не...[DiscHes] [SyntLess] Ну, [DiscHes] с шисят [Phon] (с шестьдесят) восьмого года здесь Эээ... как же... там, где...я, вишь, [Phon] (видишь) к-... был, Кри.. Кристальный, [DiscHes] по-моему, я был там три месяца. Кристалл — это... щас [Phon](сейчас) Екатеринбург называется А щас[Phon] (сейчас) Екатеринбург. resp1 Вот. Я грру [Phon] (говорю): три месяца там где-то был. так-так конечно, не... не.. сказать, шоб [Phon] (чтобы)	149

Продолжение Приложения Б

(справочное)

Типовые примеры отклонений от речевого стандарта

Тип ОРС	Вариант ОРС	Контекст	Кол-во
Morph	MorphAfF	<p>Sh0021_001</p> <p>Я <i>родилася</i> (родилась) в Таштагольский район</p> <p><i>Училася</i> (училась), так сказать.</p> <p>Не знала русского языка. Я целый год училась... и во вторых... <i>осталася</i> (осталась).</p>	9
		<p>Sh007_002</p> <p>А всяко-разно [MorphAff] (по-разному) — где по-русски, где по-шорски.</p> <p>Ну, тахтэ,[CodeSw] оне (она) [MorphAff] грит[Phon] (говорит): «Давай грит</p> <p>Вот-вот эти, которы [MorphAff] (которые) из Франции приезжали.</p> <p>он берёт этот...ну, [DiscHes] тоже у него бумага така [MorphAff] (такая) толстая — этот... в какую-нить [Phon] (в какую-нибудь) деревню приедешь — там уже...[DiscHes]</p> <p>которы[MorphAff] (которые), уже, можно сказать, уже не знают русского языка.</p> <p>Всяко-разно[MorphAff] (по-разному) — по-шорски и по-русски.</p> <p>Один из из внуков... один... ну, [DiscHes] тому токо [MorphAff] (только) четыре годика.</p> <p>А молодые.. я те[MorphAff] (тебе) говорю, не это... не [DiscHes]знают оне...[MorphAff] (они).</p>	28

[\(index.php\)](#)**Поиск заимствований в научных текстах[®]**[\(index.php/ru/\)](#) [\(index.php/en/\)](#)

Введите текст:

...или загрузите файл:

Файл не выбран...

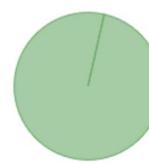
[Выбрать файл...](#)Укажите год публикации: **Выберите коллекции** Все Рефераты Авторефераты Иностранные конференции PubMed Википедия Российские конференции Иностранные журналы Российские журналы Энциклопедии Англоязычная википедия[Анализировать](#)[Проверить по расширенному списку коллекций системы Руконтекст \(<http://text.rucont.ru/like>\)](#)Обработан файл:
VR.pdf.

Год публикации: 2020.

Оценка оригинальности документа - 100.0%**Процент условно корректных заимствований - 0.0%****Процент некорректных заимствований - 0.0%**[Просмотр заимствований в документе](#)

Время выполнения: 27 с.

Заимствования отсутствуют

[Общезвестные фрагменты](#)[Значимые оригинальные фрагменты](#)[Библиографические ссылки](#)[Искать в Интернете](#)

100.00%

[Дополнительно](#)