

УДК 81'27

DOI: 10.17223/22274200/20/5

З.И. Резанова

ФРАГМЕНТ РАЗМЕТКИ В ЛИНГВИСТИЧЕСКОМ КОРПУСЕ RuTuBiC. ПЕРЕКЛЮЧЕНИЕ КОДА ИЛИ ЛЕКСИЧЕСКОЕ ЗАИМСТВОВАНИЕ?

Представлено решение одной из проблем специальной лингвистической разметки в корпусе RuTuBiC – корпусе русской речи русско-тюркских билингвов, связанной с маркированием отклонений от речевого стандарта на лексическом уровне. Предлагаются решения, основывающиеся на разграничении лексем, полностью освоенных системой русского литературного языка, и единиц, имеющих статус заимствований из других подсистем национального языка и других языков. В последнем случае выделяются языковые и контекстные признаки, противопоставляющие лексическое заимствование и переключение кодов.

Ключевые слова: корпус текстов, лексическое заимствование, переключение кодов, шорско-русский билингвизм, тюркско-русский билингвизм, аннотирование, отклонение от речевого стандарта.

Введение. Постановка проблемы

В статье представлено обоснование одного из типов лингвистического аннотирования в корпусе устной речи тюркско-русских билингвов (RuTuBiC) – разметки отклонений от речевого стандарта (ОРС) в русской речи билингвов, возникающих как нарушение норм словоупотребления в стандартной русской литературной речи под влиянием иносистемных единиц и их отношений. При этом рассматриваются случаи проникновения норм словоупотреблений других форм русского языка и других языков, прежде всего материнских языков билингвов. Рассмотрение межъязыковых и внутриязыковых межсистемных влияний в качестве вариантов единых интерферентных процессов опирается на существующую в литературе традицию, с одной стороны [1–3], с другой – на выявленное авторами проекта взаимодействие данных процессов в синхронной и диахронной динамике языков. Так, включенная в литературную речь единица из лексической системы диалекта может появиться в ней под иноязычным влиянием (ср., например, наблюдения в [4. С. 206–207]). В статье разграничиваются лексиче-

ские заимствования разных типов, являющиеся результатом межъязыковой интерференции и явления переключения кодов, характеризуются типичные позиции использования лексических заимствований и переключения кодов – коммуникативные условия реализации данного типа включений в спонтанную речь респондентов.

Отметим значительную степень актуальности данной проблематики в исследовании билингвальных речевых практик. При том, что сущностные различия между данными типами включений в речи билингвов никем не оспариваются, исследователями отмечается наличие пограничных, переходных явлений, однозначная интерпретация которых практически невозможна [5]. Это случаи типа *Всё, мы с ним на лошадь садимся, он меня: «Всё, **қызым**, поехали, дочь моя, поехали»* (М.А., в., 2)¹, относительно которых в литературе представлены различные решения: одни исследователи интерпретируют один тот же тип словоупотреблений как вид заимствований, определяя их как «заимствования для данного случая», например S. Poplack [7], другие – как вид переключения кодов, называя их переключением кодов инсертационного типа (P. Muysken) [8] или «встроенным словом» (C. MyersScotton) [9].

Данный теоретический разнобой, находящий терминологическое отражение при выполнении эмпирически направленных исследований, в частности при решении задачи не выборочной характеристики материала, но его сплошного маркирования (в том числе при создании корпусов текстов) обуславливает необходимость выработки критериев разграничения явлений в данной переходной зоне. В статье характеризуются переходные типы от лексических заимствований к полностью освоенным литературным языком лексемам, которые в корпусе не маркируются специальными тегами, с одной стороны,

¹ Здесь и далее фрагменты речи приводятся из базы данных корпуса текстов тюркско-русских билингвов Южной Сибири, хранящейся в Лаборатории лингвистической антропологии Томского государственного университета. Вследствие обязательств о конфиденциальности информации имена респондентов не приводятся. В корпусе информация хранится под кодовыми обозначениями, пользователю корпуса будут доступны только социолингвистически и психолингвистически релевантные характеристики информанта, личная идентификация недоступна. В статье контекст сопровождается инициальным идентификатором; фиксируются возраст на время записи, отнесенность к группе (социолингвистическая характеристика групп 1, 2, 3, выделяемых при описании корпуса, представлена в [6], образование (н. – начальное, с. – среднее, в. – высшее, нв. – неполное высшее).

с другой – переход от лексического заимствования (тег маркирования Lex) к переключению кодов (CodeSw). Наша позиция при решении проблемы соотносится с существующими в лингвистике теоретическими подходами и с общими установками авторского коллектива при создании корпуса RuTuBiC.

Отметим также, что актуальность проблемы использования лексических заимствований и переключения кодов в речи билингвов подтверждается и широтой круга обсуждаемых в настоящее время частных проблем: типы лексических заимствований и переключения кодов, типы включений в грамматические структуры принимающего языка, социальные и коммуникативные условия, способствующие их активации, влияния типологических характеристик вступающих во взаимодействие в языковом сознании билингва языковых пар и т.д. [10, 11]. При этом в сферу внимания при обсуждении данных аспектов билингвальных речевых практик привлекался и материал языков малочисленных народов с другими типами морфологических структур второго контактного языка. Однако, как правило, в этих работах внимание привлекается к вариантам включения русскоязычных фрагментов в устную речь на родном языке (см., напр.: [12]). Отметим, что в настоящее время активизируется «переворачивание» языковой пары при решении данной проблемы, исследуется включение фрагментов кода родного языка в речи на втором, функционально доминирующем русском языке.

В исследованиях билингвизма накоплен также значительный материал, подтверждающий влияние на процессы заимствования и переключения кода типа билингвизма (ранний vs поздний, естественный vs искусственный и др.). Вместе с тем John M. Lipski в статье 2014 г. отмечает, что в настоящее время «существует мало доступной информации о каком-либо типе языкового смешения среди тех, кто не использует язык семейного наследия часто (и даже, возможно, не говорит на нем много лет) и кто не контактирует с носителями родного языка» (перевод наш. – *З.Р.*) [13. С. 25]. В создаваемом корпусе представлен анализ речевых практик билингвов данного типа, что, по нашему времени, будет не только способствовать включению актуального эмпирического материала в исследовательскую парадигму, но и позволит сделать доказательные выводы о характере влияния данного фактора на появление заимствований из материнских языков и переключений кодов в речи билингвов.

В литературе отмечается также значимость исследований социальных коммуникативных практик, стимулирующих появление исследуемых включений [13]. Мы проследим в первую очередь влияние коммуникативных практик на появление лексических заимствований и переключений кода, охарактеризовав ограничения материала корпуса определенными типами речи, речевых жанров и тематики общения.

Объект и материал исследования: лингвистический корпус устной русской речи тюркско-русских билингвов RuTuBiC

Структура корпуса и теоретические основания принимаемых решений авторами корпуса RuTuBiC были представлены в [14], однако считаем необходимым кратко охарактеризовать те общие особенности корпуса, которые связаны с обсуждением поставленной в статье проблемы.

Создаваемый корпус принадлежит к типу так называемых корпусов носителей языка [15], т.е. фиксирует речь на втором, осваиваемом русском языке. Корпус включает три подкорпуса, представляющих материалы русской речи носителей шорско-русского, татарско-русского и хакасского-русского билингвизма.

Особенность речи носителей языка, фиксируемой в корпусе, определяется следующими основными факторами: во-первых, своеобразием проявляющегося в ней типа билингвизма – ранний естественный билингвизм с безусловным функциональным доминированием русского языка. Материнские языки с функциональной точки зрения характеризуются как языки семейного наследия (херитажные языки) (о термине см.: [16]), уровень владения которыми вариативен. Отметим также значимую характеристику языковых пар, вступающих во взаимодействие – это типологически различные языки: флективный русский язык и агглютинативные тюркские, имеющие значительные различия как в составе грамматических классов слов, так и в системных связях грамматических категорий и в составе внутрикатегориальных оппозиций. Во-вторых, в речи наших информантов проявляются характерные черты разных форм регионального варианта русского языка – литературной, просторечной, диалектной.

Корпус является бимодальным – включает синхронизированные записи устного общения интервьюеров с информантами и их транскрипты. В шорско-русском подкорпусе нашими информантами были

носители шорского языка как родного от 32 до 81 года, имеющие образование от начального до высшего, в том числе филологическое со специализацией в области шорского языка и литературы, жители г. Шерегеш и Таштагол, пос. Большая Суета. Записи проводились в экспедициях 2017–2019 гг. Следует отметить ограничения материала, определяемые условиями его сбора – это записи бесед с информантами, следствием чего является доминирование (но не исключительная представленность) диалогического типа речи и его жанров: интервью, разговора, беседы, которые могут перетекать в монологические жанры рассказа, истории.

В большинстве записей начало общения представлено формализованными и полужформализованными интервью¹, что обуславливает значительную степень тематической общности записей: языковой и социальный (связанный с речевыми практиками) опыт, семья, культура, социальное окружение, природа, события / случаи в жизни. В качестве своеобразной макротемы, к которой возвращаются коммуниканты при переключении тем, – шорская культура, шорский язык.

Корпус RuTuBiC является лингвистически размеченным корпусом. Метаразметка корпуса включает данные о тексте: о типе речи (диалог–монолог–полилог), дискурса (институциональный–персональный), жанра (интервью, разговор, беседа, рассказ, история), теме текста.

Лингвистическое аннотирование включает морфологическую разметку на основе MyStem и разметку отклонений от речевого стандарта (ОРС), что составляет основную особенность создаваемого корпуса². Первоначально введение такого типа разметки связывалось с постановкой исследовательской задачи выявления интерферентных влияний материнских языков билингвов на речевые практики на русском языке. Однако, как показал первичный анализ материала, сдвиги в речи могут быть обусловлены различными факторами, разноплановыми влияниями языковой среды говорящих. Приняв в качестве «условной нормы» письменную речь, соответствующую нормам литературного языка, авторы фиксируют в корпусе отклонения, вызванные спонтанным характером устной неподготовленной коммуникации: самоперебивание, смысловую неполноту высказывания, слова-паразиты и т.п.;

¹ Интервью проводилось на основе Социолингвистической анкеты О.А. Казакевич [17] и анкеты языкового опыта билингва [18].

² Основной объем разметки ОРС наряду с автором статьи проведен И.Г. Темниковой, Е.Б. Степаненко, А.С. Буб, Н. Жильцовой.

ОРС, обусловленные влиянием диалектов, просторечия, жаргона / сленга. То есть наряду с проявлениями межъязыковой интерференции фиксируются проявления внутриязыковой межсистемной интерференции – элементы просторечия, диалектов, жаргонов, арг, индивидуально-авторские образования. ОРС фиксируются на всех уровнях языка – в фонетике, морфологии, синтаксисе, лексике, дискурсе¹.

Маркеры отклонений от речевого стандарта на лексическом уровне русского языка в RuTuBic

Далее охарактеризуем решения, принимаемые при маркировании ОРС лексического типа. Общим основанием является отсутствие лексемы в системе русского литературного языка; заключение об этом выносится на основе данных толковых словарей: отсутствие фиксации в словарях или включение лексемы в словарь с пометами функциональной ограниченности, принадлежности к другим системам (рег., диал., обл., прост., жарг., сленг и т.д.) и фиксация в соответствующих словарях. Тег Lex используется в основном варианте и с конкретизациями: LexId, LexSem, LexSemAgr.

Тег LexId маркирует идиоматические выражения, не зафиксированные в словарях литературной речи, варианты формы общерусских единиц: *Сноха мне говорит: Ты мне посади, мне, **от моей руки не растут** (Н.Н., с., 2), ...потому что, ну, это такое **храмное блюдо** (Г.Ф., нв., 1), Он с испугу уже не знал, что делать, с испугу закричал, ну, **благим криком** (Г.П., с., 3).*

Тег LexSem маркирует использование общерусского слова в ином значении: ***Кропить** – это значит, это... как бы... или молоком там брызгаешь, ну, как бы, из ложечки... Ну, арачкой уже пореже, ну, а так вот, в основном, молоком (М.А., в., 2); Ну, мы думаем, ладно, **подкатим** (обратимся с просьбой) к Владимиру Николаевичу молоком (М.А., в, 2); Недавно у вас.... ударились вы правой стороной головы? ...Ну, или лбом ударились, или вот этой стороной, но все равно осталось.... небольшой **натёс** (травма) (И.А., нв., 2).*

Тег LexSemAgr маркирует нарушение норм лексико-семантического согласования: *Мы, во-первых, должны были и **сценки** на своём языке **говорить**, и песни свои (З.И., с., 2); Они оттуда и на вертолётё...,*

¹ Обоснование такого подхода, связанное с расширением возможности в исследовательских практиках, см.: [4].

раньше там в аэро... эропорт был, это, в Спасске, и на это... на **вертолётах приезжали** (М.Н., н., 3); Я Вас на **картиночку** сниму... (М.А., в., 2).

Маркирование единицы тегом Lex обозначает иносистемный характер лексической единицы. Проиллюстрируем положение конкретными примерами. *Вообще аристократка, то есть если у неё там в **стайке** (хлеву) будет как-то грязно...* (Г.Ф., нв., 1). Лексема *стайка* в значении «постройка для домашнего скота, хлев» представлена непоследовательно в словарях литературного языка, если отмечается, как в словаре Т.Ф. Ефремовой, то с пометой «местн.» [19], отмечена в «Словаре русских народных говоров» [20]. Фразеологизм *ни фига*, представленный в контексте *Хотя **ни фига** это не помогло, вот...* (Г.Ф., нв., 1), маркируется в словарях в разных значениях с пометами «разг.-сниж.», «прост.». Лексема *капец* (*Ну, потому что, если забодает, и всё, и **капец** мне пришёл* (Г.Ф., нв., 1)) отмечена в «Словаре русского арго» в значении «конец, провал» [21]. Лексема *оказался* в следующем контексте (*Вот он сам говорил то, что вот... ока... ока... **оказался**, короче* (Э.Н., в., 1)) не отмечена в словарях, интерпретируется как окказионализм и маркируется тегом.

Переходные явления в сфере отклонений от речевого стандарта

Анализ единиц, заимствованных из материнских языков респондентов, показывает, что они могут быть рассмотрены как находящиеся на некоей шкале переходов, на которой отдельные единицы могут рассматриваться как ядерные, прототипические для данного типа выделяемых единиц, другие проявляют свойства переходности: *Не крепкая совсем, **арачка**, ну, градусов десять-двенадцать* (М.А., в., 2); *Да, вот у нас есть **толкан**. Толкан – это уже маленько забытое, это то есть перетертое – рожь, да, или ячмень...* (Э.Н., в., 1); *Это я районный **наштык**, а в каждом посёлке есь ещё свои **наштыки** избранные* (М.А., в., 2). *И они вот, как раз вот, как говорят наши, тоже, что... надо держать **шамана**, когда он **камлает*** (М.А., в., 2). Для пометы Lex такие зоны переходов открываются «слева» – в состав полностью освоенной лексики, и «справа» – в сферу переключения кодов. Так, условно на левой границе шкалы находятся лексемы типа *шаман, наштык, кудай*. При этом лексема *шаман*, заимствованная из тунгусо-манчжурских языков, интерпретируется как полностью освоенная в русском языке: имеет полную морфологию

синтаксическую парадигму, свободные лексико-семантические связи, переносные значения, функционирует в разных формах и дискурсах русского языка, освоена деривационной системой (*шаманить, шаманство*). Высокий уровень освоенности отражается в отсутствии ограничительных помет в толковых словарях: с пометой «этнолог.» представлена только в словаре под ред. Д.Н. Ушакова, представляющем лексикон первой трети прошлого века [22]. В корпусе данное слово дается без тега Lex.

Слова *паитык* и *кудай* (*А я всё время говорю: «Религия – она единая. Бог, вообще – Кудай, он у нас Кудаем называется, у кого-то Кришина, у кого-то Христос, у кого-то там Будда, у кого Аллах, у кого-то Магомед, но Бог, он, все равно един, и религию я всегда сравниваю* (С.Г., в., 2)) маркируются тегом Lex в корпусе. Обе лексемы не отмечены в словарях литературного языка, характеризуются фонетической и морфологической освоенностью, имеют свободные лексико-семантические связи, а слово *паитык* и деривационной – отмечено производное слово *паитычество*: *Много у нас собирались, ну, сейчас вот... хочу в этом году уйти с паитычества* (М.А., в., 2). *Паитык* при этом в нашем материале представлен как широко употребительный.

Таким образом, *паитык* и *кудай* интерпретируются как единицы, характеризующиеся ограниченностью употребления, но освоенные в системе русского языка, что отличает их от единиц типа *шаман*, с одной стороны, с другой – от иносистемных единиц, появляющихся в речи в механизме переключения кода.

Переключение кодов на границе отклонений от речевого стандарта русского языка в билингвальной речи

При представлении вариантов переключения кодов также выделяем ядерные, прототипические единицы, однозначно определяемые как переключение кода, и пограничные. Так, однозначно определяются как случаи переключения кодов включения целых фразовых контекстов, например: *...и тут бабушка, короче, мимо проходит, о..., и со мной говорит: «Э... эзен-ок, улуг кижжи!», – ну, это означает «здравствуй, большой человек, ну, великий человек». Я: «Эзен-ок, куртуйақ!», – «Здравствуй, бабушка». Он такой посмотрел, посмотрел, и grit: «Сен уцнапчаң турк язык, ты?!». Я говорю: «Пистиң – это... шор кижжи тилин» (И.А., нв., 2); *А вот внук, он у меня русский, он меня понимает. Я говорю: «Костя, кель пеере!», – «Иди сюда»**

(Р.С., в., 3); *И вот она не могла по-русски сказать: «Вот, пойдём домой, поменяемся, там это, игрушкой какой-нибудь».* А, а она по-шорски говорила: *«Пульге' нара'мас».* Это, *пурна'с*, то есть, домой пойдём, поменяем (Г.Н., в., 2)¹.

О переключении кодов могут свидетельствовать метатекстовые показатели. В качестве таковых выступают маркеры цитирования чужой речи в более широком контексте рассуждений респондентов о характере использовании родного языка: *У нас есть мальчишка такой интересный, Андрей Хипсон, он: «Эзеноқтар!», – а я... я и говорю: «Вот... во множественном числе ты можешь говорить мне «Эзеноқтар!», – ну, если мы с тобой так общаемся, наедине, ты можешь сказать: «Эзеноқ, Екатерина Николаевна, здравствуйте», – ну, там, ну... (Е.Н., в., 2); Они пришли спросить, продайте молоко, сметаны или там ещё... Сметана – я не знала... Но вот молоко-то я знала точно, что молоко, **сум** – это молоко. Я говорю: бабушка... она говорит всё мне: «Нет». Она одно слово знала, что нет. Я говорю, бабушка, они просят молоко. Она: «Ну, молоко есть – **бар, бар**» (Р.С., в., 3).*

В следующем контексте респондент рассуждает о различии диалектов шорского языка и приводит пример коммуникативного сбоя в общении носителей разных диалектов: *Ну, тахтэ оне grit: «Давай, grit, **такте**, – grit, – это... *разбирать*». Кого ты собралась, – грю, – с... *пол разбирать, что ли?*». «Ну почему, – grit, – *стелить я буду матрас*». «Так это же, – *грю*, – не... не так». «Ну, мы, – grit, – так говорим». Ну, то есь, **такте** (М.Н., н., 3). Следующий контекст представляет рассуждение респондента о сходстве и различии языков в тюркской группе: *А там **қанче час** вот – на шорском, на киргизском – **канча саат**... как бы..., ну, *схоже там* (Е.Н., в., 2).**

Единичные лексические включения как вариант переключения кодов могут маркироваться в тексте переводом единицы родного языка респондента на русский язык в широких контекстах бесед о своеобразии шорской культуры: *Я *рассольник* люблю, «*ыргэ*» по-шорски (Н.Н., с., 2).* Однако в данном случае в качестве основного показателя переключения кодов мы используем фонетический критерий, противопоставляя случаи, когда слово может быть представлено в системе русской фонетики, в механизме субституции (*И *собиралися*, и это, все старики, шорский кайко'мус называется (Г.П., с., 3)), и варианты*

¹ Транскрипции и переводы шорских фрагментов подготовлены Д.М. Токмашевым при подготовке доклада на конференции [12].

коммуникации, когда наш собеседник, рассуждая о шорской культуре, намеренно переходит на шорский язык, называя явление, артефакт культуры словом родного языка, как, например, в следующем фрагменте: *Раньше делали нитки из... из конопли: кендирдең... Могли, эта... холст... холст ткать... ткать. Ткань делать, эта... Ну, шорцы знали это, видимо, ну, шорцы знали торгу – шёлк, ну, ... материал, э... там украшения, например, всякие. Ну, различные украшения чынча э... наkostenные украшения – чынча... ...ызырга – эта... как эта... серьги... Эта... чўстўк – эта... ...потом чўстўк* (С.Г., в., 2).

Заключение

Таким образом, при маркировании ОРС на лексическом уровне в русской речи тюркско-русских билингвов единиц лексико-фразеологического фонда располагаются на шкале переходов, где крайняя «левая позиция» – заимствованная, но полностью освоенная единица (не маркируется тегом), крайняя «правая позиция» – переключение кодов, о чем свидетельствуют как метатекстовые показатели, так и собственно языковые, прежде всего фонетические (маркируются тегом CodeSw), в центре – включаемые в речь респондентов лексические единицы нелитературных форм русского языка и освоенные заимствования из родных языков носителей тюркско-русского билингвизма.

Отметим незначительное количество и лексических заимствований, и случаев переключения кодов в речи респондентов корпуса RuTuBiC, что во многом определяется типом фиксируемого билингвизма, при котором второй язык (русский) является полностью освоенным и функционально доминирующим, а родной перемещен в сферу языка семейного наследия. Типичность метатекстового маркирования как лексических заимствований, так и переключения кодов определяется дискурсивными и жанровыми ограничениями материалов корпуса, направленностью интервью и свободных бесед на обсуждение разных аспектов родных для респондентов языков и культур.

Литература

1. Немзер У. Проблемы и перспективы контрастивной лингвистики // Новое в зарубежной лингвистике. М., 1989. Т. 25. С. 128–144.
2. Чойбонова Б.М. Влияние межязыковой и внутриязыковой интерференции на формирование ошибок в речи обучаемых // Вестник Бурятского университета. 2009. № 11. С. 157–162.

3. Слама-Казаку Т. Методика психолингвистического исследования при контрастивном анализе. Корпус языковых факторов, полученных в процессе обучения языку: корпус ошибок и их иерархия // Психолингвистика : сб. ст. / под ред. А.М. Шахнаровича. М., 1984. С. 283–306.

4. Резанова З.И., Дыбо А.В. Отклонения от речевого стандарта в региональном варианте устной русской речи: внутриязыковое и межъязыковое взаимодействие // Русин. 2020. № 62. С. 144–158. DOI: 10.17223/18572685/62

5. Галямина Ю.Е. Структурные типы переключения кодов в речи носителей исчезающих языков (на материале кетского и эвенкийского языков) // Вестник Томского государственного педагогического университета. 2012. № 1 (116). С. 96–100.

6. Резанова З.И., Темникова И.Г., Некрасова Е.Д. Динамика социолингвистических процессов в Южной Сибири в зеркале билингвизма (русско-шорское и русско-татарское языковое взаимодействие) // Вестник Томского государственного университета. 2018. № 436. С. 56–68. DOI: 10.17223/15617793/436/7

7. Poplack S. Sometimes I'll start a sentence in Spanish y termino en español // Linguistics. 1980. № 18. P. 581–618.

8. Muysken P. Bilingual speech: a typology of code-mixing. Cambridge : Cambridge University Press, 2000. xvi+306 p.

9. Myers-Scotton C. Social Motivations for Code-Switching: Evidence from Africa. New York : Oxford University Press, 1993. xii, 177 p.

10. Gardner-Chloros P. Code-Switching and Language Change // The Handbook of Language Contact / ed. by R. Hickey. P. 188–208.

11. Балакина Ю.В., Соснин А.В. Теоретические основы переключения кодов и функционирования заимствований с позиций контактной лингвистики // Вестник Воронежского государственного университета. Сер. Лингвистика и межкультурная коммуникация. 2015. № 2. С. 5–11.

12. Языки России в контакте с русским языком II. 11–13 февраля 2021 г. : тез. конф. / Ин-т русского языка им. В.В. Виноградова РАН, Ин-т языкознания РАН; М.-Э.А. Винклер, Е.В. Кашкин, О.А. Казакевич, А.Б. Летучий, Н.М. Стойнова, О.В. Ханина, И.А. Хомченкова. М., 2021. 92 с.

13. Lipski J.M. Spanish-English code-switching among low-fluency bilinguals: Towards an expanded typology // Sociolinguistic Studies. 2014. Vol. 8, № 1. P. 23–55. DOI: 10.1558/sols.v8i1.23

14. Резанова З.И. Подкорпус устной речи русско-тюркских билингвов Южной Сибири: типологически релевантные признаки // Вопросы лексикографии. 2017. № 11. С. 105–118.

15. Коптев М. Введение в корпусную лингвистику. Прага : Animedia Company, 2014. 194 с.

16. Азимов Э.Г., Щукин А.Н. Проблемы лексикографического описания лингводидактической терминологии // Вопросы лексикографии. 2018. № 14. С. 5–23. DOI: 10.17223/22274200/14/1

17. Казакевич О.А. Документация исчезающих языков Сибири (на материале двух поселков Красноярского края) // Вестник Российского гуманитарного научного фонда. 2006. № 3 (44). С. 221–231.

18. Marian V., Blumenfeld H.K., Kaushanskaya M. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals // *Journal of Speech, Language, and Hearing Research*. 2007. Vol. 50, № 4. P. 940–967.

19. Ефремова Т.Ф. Новый толково-словообразовательный словарь русского языка. М. : Рус. яз., 2000. Т. 2. 1084 с.

20. *Словарь русских народных говоров*. М. : Наука, 2007. Вып. 41. 343 с.

21. Елистратов В.С. Словарь русского арго. ГРАМОТА.РУ. 2002. URL: <https://rus-russian-argo.slovaronline.com> (дата обращения: 01.03.2021).

22. *Толковый словарь русского языка* / под ред. Д.Н. Ушакова. М. : Гос. изд-во иностр. и нац. слов, 1940. Т. 4. 1502 с.

Markup Fragment in the RuTuBiC Linguistic Corpus. Code-Switching or Lexical Borrowing?

Voprosy leksikografii – Russian Journal of Lexicography, 2021, 20, pp. 91–104. DOI: 10.17223/22274200/20/5

Zoya I. Rezanova, Tomsk State University (Tomsk, Russian Federation). E-mail: rezanovazi@mail.ru

Keywords: lexical borrowing, code-switching, Shor-Russian bilingualism, Turkic-Russian bilingualism, text corpus, annotation, deviation from speech standard.

The article presents a solution to one of the problems of special linguistic markup in the RuTuBiC corpus – the Russian Speech Corpus of Russian-Turkic Bilinguals, associated with error annotation at the lexical level. The corpus includes three subcorpuses representing materials of the Russian speech of Shor-Russian, Tatar-Russian and Khakass-Russian bilinguals. The article presents solutions developed on the basis of all subcorpuses; the illustrative contexts are drawn from the Shor-Russian subcorpus, recordings of interviews with 14 respondents, about 20 hours of sound. The recordings were made during expeditions to Shoria in 2017–2019. Bilingualism of the respondents is defined as early natural bilingualism with the dominance of the second Russian language, mother tongues are languages of the family heritage. The theoretical basis of the research was works on linguistic contact at the lexical level. Solutions based on the differentiation of lexemes fully mastered by the system of standard Russian and units with the status of borrowings from other subsystems of the national language and other languages are proposed. In the latter case, linguistic and contextual features are distinguished that oppose lexical borrowing and code-switching. The typical errors singled out at the lexical level are: [LexId] – idiomatic expressions that are not fixed in the standard language (dialectal and vernacular, slang, etc.), they can also be Turkic calques; [LexSem] – general Russian words used in meanings different from those fixed in the normative sources; [LexSemAgr] – violations of the lexical and semantic agreement norms. The units borrowed from the mother tongue of the respondents are located on the scale of transitions from nuclear to borderline. The nuclear units marked with the [Lex] tag are dialectal units, common words, other word usage cases that are outside the standard, as well as borrowings from the Turkic languages that are not included in the dictionaries of standard Russian. On the border “to the left” are borrowings

assimilated to different degrees. On the border “to the right” are non-assimilated borrowings and code-switches. The [CodeSw] marks code-switching, insertion of mother tongue elements into Russian speech. The author considers the inclusion of statements as nuclear cases of code-switching, and single lexical inclusions as transitional cases. Code-switching is evidenced by metatext and linguistic proper, primarily phonetic, indicators. There is an insignificant number of both lexical borrowings and cases of code-switching in the speech of the respondents of the RuTuBiC corpus, which depends on the type of bilingualism. The typicality of metatext marking of borrowings and code-switches is determined by the discursive, genre and thematic limitations of the corpus.

References

1. Nemser, W. (1989) Problems and prospects in contrastive linguistics. In: Neroznak, V.P. & Gak, V.G. (eds) *Novoe v zarubezhnoy lingvistike* [New in foreign linguistics]. Vol. 25. Moscow: Progress. pp. 128–144. (In Russian).
2. Choybonova, B.M. (2009) Vliyaniye mezh'yazykovoy i vnutriyazykovoy interferentsii na formirovaniye oshibok v rechi obuchaemykh [The influence of interlingual and intralingual interference on the formation of errors in student speech]. *Vestnik Buryatskogo universiteta – BSU Bulletin*. 11. pp. 157–162.
3. Slama-Kazaku, T. (1984) Metodika psikholingvisticheskogo issledovaniya pri kontrastivnom analize. Korpus yazykovykh faktorov, poluchennykh v protsesse obucheniya yazyku: korpus oshibok i ikh ierarkhiya [Methods of psycholinguistic research in contrastive analysis. The corpus of linguistic factors obtained in teaching a language: the corpus of errors and their hierarchy]. In: Shakhnarovich, A.M. (ed.) *Psikholingvistika: sb. st.* [Psycholinguistics: collection of articles]. Moscow: Progress. S. 283–306.
4. Rezanova, Z.I. & Dybo, A.V. (2020) Deviations from the speech standard in the regional version of oral Russian speech: intralingual and interlingual interactions. *Rusin*. 62. pp. 144–158. (In Russian). DOI: 10.17223/18572685/62
5. Galyamina, Yu.E. (2012) Structural typology of code-switching in the texts in endangered languages (On the Ket and Evenk data). *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta – Tomsk State Pedagogical University Bulletin*. 1 (116). pp. 96–100.
6. Rezanova, Z.I., Temnikova, I.G. & Nekrasova, E.D. (2018) Dynamics of sociolinguistic processes in Southern Siberia mirrored in bilingualism (Russian-Shor and Russian-Tatar language interaction). *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 436. pp. 56–68. (In Russian). DOI: 10.17223/15617793/436/7
7. Poplack, S. (1980) Sometimes I'll start a sentence in Spanish y termino en español. *Linguistics*. 18. pp. 581–618.
8. Muysken, P. (2000) *Bilingual speech: a typology of code-mixing*. Cambridge: Cambridge University Press.
9. Myers-Scotton, C. (1993) *Social Motivations for Code-Switching: Evidence from Africa*. New York: Oxford University Press.

10. Gardner-Chloros, P. (2010) Code-Switching and Language Change. In: Hickey, R. (ed.) *The Handbook of Language Contact*. John Wiley & Sons. pp. 188–208.
11. Balakina, Yu.V. & Sosnin, A.V. (2015) Conceptual framework for code-switching and lexical borrowings from the perspective of contact linguistics. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Ser. Lingvistika i mezhkul'turnaya kommunikatsiya – Proceedings of Voronezh State University. Series: Linguistics and Intercultural Communication*. 2. pp. 5–11. (In Russian).
12. Vinkler, M.-E.A. et al. (eds) (2021) *Yazyki Rossii v kontakte s russkim yazykom* [Indigenous Languages of Russia in Contact with Russian]. II Conference. 11–13 February 2021. Book of Abstracts. Moscow: V. V. Vinogradov Russian Language Institute, RAS & Institute of Linguistics, RAS.
13. Lipski, J.M. (2014) Spanish-English code-switching among low-fluency bilinguals: Towards an expanded typology. *Sociolinguistic Studies*. 8 (1). pp. 23–55. DOI: 10.1558/sols.v8i1.23
14. Rezanova, Z.I. (2017) Subcorpus of oral speech of Russian-Turkic bilinguals of Southern Siberia: typologically relevant signs. *Voprosy leksikografii – Russian Journal of Lexicography*. 11. pp. 105–118. (In Russian). DOI: 10.17223/22274200/11/7
15. Kopotev, M. (2014) *Vvedenie v korpusnuyu lingvistiku* [Introduction to corpus linguistics]. Prague: Animedia Company.
16. Azimov, E.G. & Shchukin, A.N. (2018) Problems of Lexicographic Description of Language Teaching Terminology. *Voprosy leksikografii – Russian Journal of Lexicography*. 14. pp. 5–23. (In Russian). DOI: 10.17223/22274200/14/1
17. Kazakevich, O.A. (2006) Documents of Obsolescent Languages of Siberia (Based on the Materials from Two Settlements in Krasnoyarsk Territory). *Vestnik Rossiyskogo gumanitarnogo nauchnogo fonda*. 3 (44). pp. 221–231. (In Russian).
18. Marian, V., Blumenfeld, H.K. & Kaushanskaya, M. (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*. 50 (4). pp. 940–967.
19. Efremova, T.F. (2000) *Novyy tolkovo-slovoobrazovatel'nyy slovar' russkogo yazyka* [New explanatory-derivational dictionary of the Russian language]. Vol. 2. Moscow: Russkii yazyk.
20. Sorokoletov, F.P. (ed.) (2007) *Slovar' russkikh narodnykh govorov* [Dictionary of Russian folk dialects]. Vol. 41. Moscow: Nauka.
21. Elistratov, V.S. (2002) *Slovar' russkogo argo* [Dictionary of Russian argo]. GRAMOTA.RU. [Online] Available from: <https://rus-russian-argo.slovaronline.com> (Accessed: 01.03.2021).
22. Ushakov, D.N. (ed.) (1940) *Tolkovyy slovar' russkogo yazyka* [Explanatory dictionary of the Russian language]. Vol. 4. Moscow: Gos. izd-vo inostr. i nats. slov.