

## ФИЛОЛОГИЯ

УДК 811.511.151

**Кашкин Е.В.**

*Институт русского языка им. В.В. Виноградова РАН  
(г. Москва)*

### **ПОЛЕВЫЕ ИССЛЕДОВАНИЯ ВОСТОЧНЫХ ГОРНОМАРИЙСКИХ ГОВОРОВ: КОРПУС ТЕКСТОВ**

*Работа поддержана грантом РФФИ № 19-012-00627.*

В статье описан корпус текстов, подготовленный коллективом лингвистических экспедиций МГУ им. М.В. Ломоносова и в основном включающий расшифровки устных текстов. Обсуждается его состав и принципы разработки. Приводятся примеры использования корпуса для решения исследовательских задач.

**Ключевые слова:** горномарийский язык, корпус, методика полевых исследований, документация.

#### **1. Введение**

Статья продолжает цикл публикаций о результатах лингвистических экспедиций МГУ имени М.В. Ломоносова по изучению горномарийского языка [см. Кашкин 2018а, 2019]. В этих работах обсуждаются материалы проекта, в рамках которого с 2016 г. исследуются говоры с. Кузнецово и нескольких окрестных деревень (Апшак-Пеляк, Кожланангер, Кукшилиды, Никишкино, Паулкино, Тюманово, Яшпатрово). Опубликованные работы и новости проекта доступны на сайте <http://hillmari-exp.tilda.ws>

В этой статье речь пойдет о корпусе текстов, разработанном в ходе экспедиций. В целом под корпусом понимается

база текстов, в которой имеется разметка (морфологическая, семантическая или какая-либо иная), см. справку на сайте Национального корпуса русского языка (далее НКРЯ <http://ruscorpora.ru/corpora-intro.html>). Корпус используется в наших исследованиях наряду с материалами, полученными методом анкетирования. В разделе 2 мы охарактеризуем объем и состав корпуса. Раздел 3 содержит примеры его использования в исследовательских целях. В разделе 4 подводятся итоги.

## **2. Объем и состав корпуса**

По состоянию на апрель 2019 г. корпус включает 63 522 словоупотребления (т.е. всего слов в текстах с учетом всех повторений каждого слова; в англоязычной литературе этому соответствует термин *tokens*). В дальнейшем планируется его пополнение. Он размещен в свободном доступе на сайте <http://hillmari-exp.tilda.ws/corpus>. Для пользователей доступна оригинальная база, которая открывается в свободно распространяемой программе SIL FieldWorks. Дополнительно все тексты выложены в файлах .pdf и .txt. Горномарийские словоформы записаны в фонологической транскрипции латиницей, в основном соответствующей системе из [Саваткова, 2002]. Для каждой словоформы приведено поморфемное глоссирование в системе, разработанной в нашем проекте. Каждое предложение сопровождается литературным переводом. Звуковые файлы на момент написания статьи не выложены на сайте проекта, но могут предоставляться по запросу. Запись текстов кириллицей на данный момент не произведена, но может составлять одну из задач дальнейшей работы.

В основном корпус включает тексты, записанные и расшифрованные в экспедициях при участии всех членов коллектива. Дополнительно добавлено несколько текстов на тех же говорах из [Акцорин, 1991; Саватеева, 2005]. При каждом тексте приведены метаданные, где указан его источник, а

также перечислены работавшие с ним люди (рассказчик; участники экспедиции, занимавшиеся разбором; носители языка, участвовавшие в разборе). Корпус включает тексты различных жанров: бытовые рассказы, рассказы об истории населенных пунктов и различных объектов (например, Кузнецовской школы), сказки, описания визуальных стимулов (рассказы по картинкам, пересказы мультфильмов и пр.). Имеется серия текстов с описаниями игр (за идею и помощь в их обработке мы благодарны М.А. Ключевой).

Обработка каждого текста для корпуса устроена следующим образом. Сначала устный рассказ записывается на диктофон. После этого участник экспедиции вместе с носителем языка готовит первую версию расшифровки и перевода. Далее расшифровка сверяется с аудиозаписью, появившиеся вопросы уточняются у носителя языка.

После экспедиции расшифровка и перевод вычитываются опытными участниками проекта. В программе SIL FieldWorks делается автоматический разбор. Вопросы, возникшие на этом этапе работы, уточняются у носителя языка в следующей экспедиции или дистанционно.

Некоторых комментариев требует объем корпуса. Сейчас он составляет 63 522 словоупотребления. Современные корпуса «крупных» языков имеют несопоставимо больший объем (ср., например, основной корпус НКРЯ, насчитывающий на момент подготовки статьи 210 млн. словоупотреблений). Однако такие корпуса включают в основном письменные тексты, которые подвергаются в первую очередь автоматической разметке без снятия омонимии (так, словоформа *лечу* в большинстве примеров из НКРЯ, не проходивших ручную проверку разметки, соотносится и с лексемой *лететь*, и с лексемой *лечить*; словоформа *дочери* получает вариативную характеристику Р. п. ед. ч., Д. п. ед. ч., П. п. ед. ч., И. п. мн. ч. – снятие омонимии во всех этих случаях возлагается на самого пользователя). Наш корпус включает

устные тексты, которые были вручную расшифрованы, размечены частично автоматически, после чего вся разметка была проверена и уточнена вручную. Такая процедура неизбежно снижает возможное количество словоупотреблений в корпусе, что можно видеть и при сравнении с другими корпусами устных текстов на языках народов России, ср., например, корпус бесермянского диалекта удмуртского языка (<http://beserman.ru/>), который по состоянию на апрель 2019 г. насчитывает «65 тыс. словоупотреблений и постоянно увеличивается», или корпуса некоторых сибирских языков на сайте ИЭА РАН (<http://corpora.iea.ras.ru/corpora/statistics.php>), насчитывающие от 10 тыс. до 171 тыс. словоупотреблений. В то же время частотность различных грамматических показателей в нашем корпусе не столь мала, ср., например, 2665 вхождений посессивного показателя 3 л. ед. ч. *-жы* (с учетом алломорфов), 686 вхождений частицы *=ок*, 227 вхождений каузативного показателя *-ыкт* (с учетом алломорфов).

Следует особо оговорить, что данный корпус не является готовым образовательным ресурсом. Все тексты записаны от людей, свободно владеющих горномарийским языком. В то же время наша запись отражает особенности речи, которые могут не соответствовать литературной орфографии. В текстах встречаются диалектные особенности, индивидуальные особенности речи, вставки русских слов при наличии горномарийского аналога, а также естественные для устной речи на любом языке сбои, оговорки и проч. Поэтому использование корпуса в образовательных целях предполагает предварительную работу учителя с нужным ему текстом. Для научных задач при работе с такими данными всегда учитывается частотность явления и производится его проверка разными методами (анкетированием многих носителей, сопоставлением с работами предшественников и др.).

### 3. Корпус и научные исследования

#### 3.1. Иллюстрация 1: сложные глагольные комплексы

Рассмотрим два примера применения нашего корпуса для исследовательских задач. Первый из них касается употребления предикатов *шб́ндәш* ‘сажать’ и *шуаш* ‘бросать’ в сложных глагольных комплексах, образованных формой деепричастия смыслового глагола и грамматикализированным глаголом (ср. *поген шб́нденәм* ‘я собрал’, букв. ‘собрав посадил’), см. [Кашкин, 2018b, 2018c] и приводимые в этих работах ссылки.

В качестве примеров толкований таких конструкций в литературе рассмотрим работы [Современный, 1961] (в этой книге в большей степени рассматриваются луговые марийские варианты) и [Саваткова, 2008]. Глагол *шб́ндәш* трактуется в [Современный, 1961: 204–205] как образующий «составные глаголы от некоторых переходных глаголов» и придающий им «видовое значение однократной завершенности действия». Согласно [Саваткова, 2008: 372], этот глагол «употребляется в составных глаголах для выражения действия, доведенного до предела». О глаголе *шуаш* в [Современный, 1961: 206] сказано, что он «образует составные глаголы от некоторых переходных глаголов и придает им видовое значение однократности, стремительности совершения действия». Согласно [Саваткова, 2008: 353], он «образует составные глаголы со значением завершенности действия». Эти толкования, безусловно, помогают понимать общий смысл таких конструкций в текстах, однако не содержат информации достаточного уровня глубины о том, в каких в точности случаях используются предикаты *шб́ндәш* и *шуаш* в сложных глагольных комплексах. Кроме того, из процитированных работ не вполне ясны различия, проявляемые этими предикатами в конструкциях рассматриваемого типа (например, не очевидно, чем отличаются ярлыки

«действие, доведенное до предела» и «завершенность действия»).

С помощью корпуса мы можем получить частотные списки смысловых глаголов, которые используются в форме деепричастия в сложных глагольных комплексах с *шб́ндәш* и *шуаш*. На основании этих списков можно строить обобщения о классах смысловых глаголов.

Глаголы, с которыми в нашем корпусе наиболее часто сочетаются в указанных конструкциях предикаты *шб́ндәш* и *шуаш*, представлены в Таблицах 1 и 2. Имеется, кроме того, общее различие в частотности этих конструкций: 340 вхождений конструкции с *шб́ндәш* и 44 вхождения конструкции с *шуаш*.

Таблица 1. Сочетаемость глагола *шб́ндәш* в сложных глагольных комплексах

Глагол	Перевод	Частотность
<i>опташ</i>	класть, накладывать	29
<i>кычаш</i>	держать, ловить	20
<i>б́штәш</i>	делать	20
<i>пырташ</i>	вносить	16
<i>чиәш</i>	надевать	16
<i>виктәш</i>	протягивать	15
<i>питб́рәш</i>	закрывать, запирать	14
<i>ялилташ / ялышлташ</i>	завязывать	12
<i>шолташ</i>	варить	12
<i>йәмдб́ләш</i>	готовить	10
<i>качкаш</i>	есть	10
<i>тб́рләш</i>	вышивать	9
<i>сирәш</i>	писать	8
<i>чистаяш / чистәйәш</i>	чистить	5

<i>нәләш</i>	брать	5
<i>леведәш</i>	накрывать	4
<i>шәрәш</i>	расстилать, разворачивать	4
<i>рисуяш</i>	рисовать	4
И др. глаголы по 1–3 вхождения.		

Таблица 2. Сочетаемость глагола *шуаш* в сложных глагольных комплексах

Глагол	Перевод	Частотность
<i>кыдашаш</i>	снимать	5
<i>анжааш</i>	смотреть	3
<i>шиәш</i>	бить	3
<i>пёчкәш / пёчкәдәш</i>	резать	3
<i>мондаш</i>	забывать	3
<i>салаш</i>	косить	3
<i>шәләтәш</i>	разваливать, разбрасывать	2
<i>лыкташ</i>	извлекать, доставать	2
<i>пушташ</i>	убивать	1
<i>пачкалташ</i>	стряхивать	1
<i>кёрәш</i>	рвать	1
<i>качкаш</i>	есть	1
<i>панежәш</i>	опаливать	1
<i>пыдырташ</i>	ломать	1
<i>потопаш</i>	затапливать	1
<i>йылаташ</i>	жечь	1
<i>итёрәйәш</i>	очищать	1
И др. глаголы по 1 вхождению.		

Как видно из Таблиц 1 и 2, глаголы *шб́ндáш* и *шуаш* сочетаются в сложных глагольных комплексах с разными классами смысловых глаголов. Глагол *шб́ндáш* в первую очередь используется с глаголами наступления долговременного состояния объекта, создания нового объекта, накопления ресурса. В свою очередь глагол *шуаш* сочетается чаще всего с глаголами разрушения, уничтожения, удаления объекта.

### 3.2. Иллюстрация 2: частица =ок

Следующая иллюстрация касается частицы =ок, см. ее предыдущие описания в [Saarinen, 1986; Сибатрова, 1987; Эрцикова, 2016; Кашкин и др., 2017]. Согласно [Saarinen, 1986], эта частица имеет тенденцию маркировать обстоятельственные группы, однако подробных количественных данных, по которым можно было бы проверить это утверждение, в цитируемой статье не приводится. В нашем корпусе частица =ок встретилась 686 раз. Данные о ее сочетаемости с различными падежными формами существительных (всего таких контекстов нашлось 151) представлены в Таблице 3.

Таблица 3. Сочетаемость частицы =ок с падежными формами существительных

Падежная форма	Количество сочетаний
Номинатив	39
Генитив	18
Аккузатив	12
Датив	1
Инессив	56
Иллатив	7
Латив	16
Комитатив	–
Каритив	2



Как видно из Таблицы 3, частица =ок чаще сочетается с формами локативных падежей (инессива, иллатива и латива) – таких контекстов всего обнаружено 79. Именно для этих падежных форм типичны обстоятельственные контексты. Более того, и в других случаях =ок часто оформляет обстоятельства: например, из 39 сочетаний =ок с номинативом 7 представлены формой *ирок=ок* (где *ирок* – ‘утро’). Частотны и сочетания этой частицы с наречиями, см., например, *тенге=ок* (65 вхождений), *пйтәри=ок* (22 вхождения), где *тенге* – ‘так’, *пйтәри* – ‘сначала’. Тем самым, мы получили данные о частотности =ок, с помощью которых можно проверить и уточнить выводы предшественников. Далее эти данные могут служить основой для различных теоретических интерпретаций [см. Кашкин и др., 2017].

### 3.3. Некоторые ограничения

Использование корпуса, описываемого в этой статье, как исследовательского инструмента имеет ряд ограничений. Во-первых, некоторые из единиц, которые могут интересовать исследователя, встречаются в текстах не очень часто, в силу чего в корпусе может не найтись данных по каким-либо их релевантным свойствам. Например, в нашем проекте подробно исследуется показатель приближенности *нәрйи* (ср. *лу эдем нәрйи* ‘примерно 10 человек’), [см. подробнее Сидорова, 2018]. В этом исследовании необходимо получить ответы на множество вопросов о порядке элементов в конструкции с *нәрйи* и о морфологическом маркировании разных элементов. В частности, нужно учитывать факторы арифметического значения числительного, падежа существительного в конструкции, числового маркирования существительного и др. Однако в текущей версии корпуса показатель *нәрйи* встретился всего 19 раз. Такого количества вхождений недостаточно для корректного количественного анализа названных факторов, поэтому основным методом в данном случае должно быть анкетирование. В то же время

при анкетировании можно проверить все модели устройства конструкций с *нәрйи*, которые встретились в корпусе, и в этом смысле корпусные данные тоже ценны.

Во-вторых, иногда происходят сдвиги количественных показателей, обусловленные спецификой вошедших в корпус текстов. Ярким примером служит глагол *кйишкйиш*, см. о различных его свойствах [Мордашова, 2017; Кашкин, 2018с]. В [Саваткова, 2008: 127] этому глаголу приписаны значения ‘бросать, кидать, подбрасывать, забрасывать’, ‘лить, выливать, поливать’, ‘метать (стог)’. Дополнительно отмечены некоторые фразеологизмы.

В корпусе насчитывается 51 вхождение глагола *кйишкйиш*, в 26 из которых он выражает значение ‘метать (икру)’. Использование глагола *кйишкйиш* в таком значении особенно не противоречит лексикографическим данным, однако считать его основным для данного глагола было бы спорно. Объяснение такого смещения количественных данных довольно простое. От одного из носителей языка был записан длинный рассказ о жизни рыб. Помимо прочего, он рассказал, когда разные рыбы мечут икру. Все 26 вхождений глагола *кйишкйиш*, в которых он передает такое значение, содержатся в одном этом тексте. Так, приводимый далее фрагмент посвящен рыбе, которую говорящий обозначил как *шим туп лавал* ‘лещ с черной спинкой’. В нем содержится 3 вхождения глагола *кйишкйиш* в обсуждаемом значении: *Мөртньём кйишкйимби годымжы тупвлйштби веле кайын шбйнзйт. Уже амыжвлй лошты ыжаргаш тбйнгйлеш, амыжвлй веле мыльгат. Мөртньём кйишкйит. Лавал кытланок кйишкйй мөртньём* ‘Когда мечут икру, спинки только у них торчат. Уже в камышах зеленеть начинает, камыши только трясутся. Икру мечут. Одновременно с лещом мечет икру’.

#### 4. Заключение

В статье было описано устройство корпуса текстов на восточных говорах горномарийского языка, разработанного

в ходе лингвистических экспедиций МГУ и включающего в основном расшифровки устных рассказов. Приведены примеры использования корпуса для решения исследовательских задач. В текущей версии корпус представляет статистические данные для описания частотных языковых единиц и хотя бы образцы естественных примеров для менее частотных единиц. В дальнейшем корпус планируется расширять. Несомненно, актуальна и задача разработки других корпусов для горномарийского языка, в т.ч. больших корпусов письменных текстов с автоматизированной разметкой.

## Литература

*Акцорин В.А.* Марийский фольклор. Мифы, легенды, предания. – Йошкар-Ола: Мар. кн. изд-во, 1991. – 288 с.

*Кашкин Е.В., Сидорова М.А., Мордашова Д.Д., Гарейшина А.Р., Учитель И.К.* Всё =ок: о семантике одной горномарийской энклитики // Доклад на конференции «Малые языки в большой лингвистике» (Москва, 2–3 ноября 2017 г.).

*Кашкин Е.В.* Полевое изучение восточных говоров горномарийского языка: лингвистическая экспедиция МГУ // XVII Игнатьевские чтения. Материалы докладов и выступлений на ежегодной Региональной научно-практической конференции с международным участием «Горные мари в поликультурном пространстве», посвященной 90-летию со дня рождения доктора филологических наук, почетного профессора МарГУ Л.П. Васиковой (1927–2012). (Пайгузово, 28 мая 2017 г.). – Йошкар-Ола, 2018а. – С. 3–12.

*Кашкин Е.В.* О семантике сложных глагольных комплексов в горномарийском языке: конструкции со значением достижения предела // Проблемы языка. Сборник научных статей по материалам Шестой конференции-школы «Проблемы языка: взгляд молодых ученых». – М.: ИЯз РАН, Канцлер, 2018b. – С. 110–129.

*Кашкин Е.В.* К типологии семантического развития глаголов со значением ‘бросать’: данные горномарийского языка // Урало-алтайские исследования. – № 4. – 2018с. – С. 112–129.

*Кашкин Е.В.* Исследование горномарийской лексики в свете современной типологии: опыт лингвистической экспедиции МГУ // XVIII Игнатьевские чтения. Материалы докладов и выступлений на ежегодной Региональной научно-практической конференции «Роль просветителей в истории, культуре и образовании горных мари», посвященной 175-летию со дня рождения педагога-священника Г.Я. Яковлева (1843–1912). (Козьмодемьянск, 10 апреля 2018 г.). – Йошкар-Ола, 2019. – С. 10–19.

*Мордашова Д.Д.* Особенности лексико-семантического поля бросания в горномарийском языке (на материале говора с. Кузнецово) // Доклад на Чтениях, посвященных 85-летию со дня рождения А.И. Кузнецовой (Москва, 3–4 марта 2017 г.).

Национальный корпус русского языка (НКРЯ) // URL: <http://ruscorpora.ru/> (дата обращения: 17.04.2019).

*Саватеева Г.А.* Лексические особенности правобережных говоров горномарийского языка: дисс ... канд. филол. наук. – Йошкар-Ола: МарГУ, 2005. – 321 с.

*Саваткова А.А.* Горное наречие марийского языка. – Szombathely: Berzsenyi Dániel Főiskola, 2002. – 292 с.

*Саваткова А.А.* Словарь горномарийского языка. – Йошкар-Ола: Мар. кн. изд-во, 2008. – 404 с.

*Сибатрова С.С.* Частицы горного наречия марийского языка // Материалы и исследования по марийской диалектологии. – Йошкар-Ола: МарНИИЯЛИ, 1987. – С. 113–134.

*Сидорова М.А.* Аппроксимативные количественные конструкции в горномарийском языке // Томский журнал лингвистических и антропологических исследований. – 2018. – № 2 (20). – С. 51–63.

Современный марийский язык. Морфология. – Йошкар-Ола: Мар. кн. изд-во, 1961. – 324 с.

*Эрцикова Г. А.* Сравнительный анализ частиц-прилеп в марийском и мордовских языках // Ежегодник финно-угорских исследований. – 2016. – Т. 10. № 1. – С. 9–16.

*Saarinen S.* Über einige tscheremissische Enklitika // Ural-Altäische Jahrbücher, Band 6, 1986. – P. 37–46.