

# МУЛЬТИМЕДИЙНЫЙ КОРПУС РУССКИХ ГОВОРОВ УДМУРТИИ: ЭЛЕКТРОННЫЙ ПОДКОРПУС УСТНОЙ РЕЧИ

Multimedia Corpus of Russian Dialects of Udmurtia:  
Electronic Subcorpus of Spoken Speech

*Регина Александровна Верняева*  
*anikina.regina@gmail.com*

*Ижевский государственный технический университет имени М.Т. Калашникова*  
*(Ижевск, Россия)*

*Екатерина Анатольевна Жданова*  
*zhdanovaea@gmail.com*

*Ижевский государственный технический университет имени М.Т. Калашникова*  
*(Ижевск, Россия)*

Regina A. Vernyaeva  
anikina.regina@gmail.com

Kalashnikov Izhevsk State Technical University (Izhevsk, Russia)

Ekaterina A. Zhdanova  
zhdanovaea@gmail.com

Kalashnikov Izhevsk State Technical University (Izhevsk, Russia)

ISSN: 1698-322X ISSN INTERNET: 2340-8146

Fecha de recepción: 07.05.2023

Fecha de evaluación: 27.11.2023

*Cuadernos de Rusística Española n° 19 (2023), 67 - 79*

## РЕЗЮМЕ

В статье представлено описание электронного подкорпуса устной речи, который является неотъемлемой частью мультимедийного Корпуса русских говоров Удмуртии (<http://dialect.manuscripts.ru>). Подкорпус позволяет изучать диалектную лексику не только методами корпусной лингвистики, но и методами лингвистической географии и электронной лексикографии. Важной составляющей подкорпуса является наличие в нем аудиозаписей устной диалектной речи, которые позволяют выявлять и анализировать фонетические и грамматические особенности русских говоров республики. В связи с тем, что Удмуртия является мультILINGВАЛЬНЫМ регионом, изучение аудиозаписей помогает выявить факты влияния на русские говоры других языков, распространенных в республике. Речь диалектоносителей, представленная в виде аудиозаписей, а также результаты разметки этой речи являются новым языковым материалом, уточняющим сведения о лексико-семантических особенностях русских говоров Удмуртии. Лингвогеографический модуль дает возможность произвести выборку всех ответов на один из вопросов программы, на которой основана разметка корпуса, результатом становится отображение полученных слов на электронной карте Удмуртии. В статье представлены результаты разметки аудиозаписей, приведены примеры карт и словарных статей, составленных при помощи корпусного менеджера.

*Ключевые слова:* лингвистический корпус, аудиоразметка, русские говоры Удмуртии, лингвистическая география.

**ABSTRACT**

The article presents a description of the electronic subcorpus of oral speech, which is an integral part of the multimedia Corpus of Russian Dialects of Udmurtia (<http://dialect.manuscripts.ru>). The subcorpus allows you to study dialect vocabulary not only by the methods of corpus linguistics, but also by the methods of linguistic geography and electronic lexicography. An important component of the subcorpus is the presence in it of audio recordings of oral dialect speech, which makes it possible to identify and analyze the phonetic and grammatical features of various dialects of the republic. Due to the fact that Udmurtia is a multilingual region, the study of audio recordings helps to identify the interaction of languages spoken in the republic. The speech of dialect speakers, presented in the form of audio recordings, as well as the results of marking this speech, are new language material that clarifies information about the lexical and semantic features of the Russian dialects of Udmurtia. The linguistic and geographical database module makes it possible to select all the answers to one of the questions of the program on which the corpus marking is based, the result is the display of this word on the electronic map of Udmurtia. The article presents the results of marking up audio recordings, examples of maps and dictionary entries compiled using the corpus manager.

*Keywords:* linguistic corpus, audio markup, Russian dialects of Udmurtia, linguistic geography.

**КОРПУСНАЯ ЛИНГВИСТИКА И КОРПУСЫ УСТНОЙ РЕЧИ В СЕТИ ИНТЕРНЕТ**

Одним из перспективных направлений развития прикладной лингвистики в настоящее время, несомненно, является корпусная лингвистика, которая позволяет исследовать факты речи и языка на основе анализа большого объема текстового материала.

Известно, что текст как объект базы данных может быть представлен не только в письменной форме, но и в устной.

В сети Интернет существует большое количество текстовых корпусов (национальных, параллельных, исторических и др.), однако далеко не каждый такой корпус способен не только представить материал в виде письменного текста (набор транскрипции текста, фотокопия текста), но и продемонстрировать его звучащую форму. В последнем случае следует говорить о корпусах устной речи, под которыми понимается коллекция аутентичных устных текстов, представленных в виде аудио- или видеозаписей. Большинство из них содержит аудиофайл записанной монологической или диалогической речи и транскрипцию записи.

Среди наиболее значимых лингвистических корпусов русского языка нужно назвать следующие:

1. *Общие корпуса современного русского языка:*

Национальный корпус русского языка (<https://ruscorpora.ru/>), Корпус русского литературного языка (<https://narusco.ru/>), Открытый корпус русского языка (<http://opencorpora.org/>), Генеральный интернет-корпус русского языка (<http://www.webcorpora.ru/>), Araneum Russicum ([http://ucts.uniba.sk/aranea\\_about/\\_russicum.html](http://ucts.uniba.sk/aranea_about/_russicum.html)) и др.

2. *Устные корпуса русского языка:*

«Рассказы о сновидениях» и другие корпуса звучащей речи (<http://spokencorpora.ru/>), Корпуса русской речи билингов-носителей других языков: башкирского (<http://lingconlab.ru/BashkirRus/#!/>), бесермянского

(<http://lingconlab.ru/BesermanRus/#!/>), дагестанских (<http://parasolcorpus.org/dagrus/#!/>), карельского (<http://lingconlab.ru/karelrus/#!/>), цыганского (<http://lingconlab.ru/romarus/#!/>), чувашского (<http://lingconlab.ru/ChuvashRus/#!/>), якутского ([http://lingconlab.ru/cs\\_yakut/](http://lingconlab.ru/cs_yakut/)) и др.

3. *Параллельные корпусы:*

Параллельный корпус переводов «Слова о полку Игореве» (<http://www.nevmenandr.net/slovo/>), Параллельный корпус русских и французских поэтических текстов первой трети XIX в. (<http://www.nevmenandr.net/fr/>), InterCorp (<https://www.korpus.cz/intercorp>), ParaSol ([https://parasol.vmguest.uni-jena.de/parallel\\_crystal/#corpus?tab=basic&cat=all&sketches=0&lang=&lang2=&query=&showOld=0](https://parasol.vmguest.uni-jena.de/parallel_crystal/#corpus?tab=basic&cat=all&sketches=0&lang=&lang2=&query=&showOld=0)) и др.

4. *Исторические корпусы:*

Манускрипт: Славянское письменное наследие (<http://manuscripts.ru/>), СКАТ – Санкт-Петербургский корпус агиографических текстов (<http://project.phil.spbu.ru/scat/page.php?page=project>), Древнерусские берестяные грамоты (<http://gramoty.ru/birchbark/>), Регенбургский диахронический корпус русского языка (<https://www.slawistik.hu-berlin.de/de/member/meyerrol/subjekte/rrudi>) и др.

5. *Корпусы диалектного языка:*

Говор бассейна Устьи. Корпус севернорусской диалектной речи (<http://www.parasolcorpus.org/Pushkino/>), Электронная библиотека русских народных говоров (<https://dialekt.corpus.tatar/>), Говор северной деревни (<https://borbushino.vogu35.ru/>), The Oslo Corpus of Pskov Dialects (<http://tekstlab.uio.no/pskov/>), Томский диалектный корпус (<http://losl.tsu.ru/corpus/demo>) и др.

Объектом исследования в рамках данной работы является представление звучащей диалектной речи в цифровой форме.

В связи с тем, что большинство существующих в настоящее время диалектных корпусов ключевой целью ставят загрузку письменной формы диалектных текстов, то актуальным становится отображение не только транскрипции письменного текста, но и представление звуковой формы речи носителей того или иного говора, что, несомненно, является важным источником для дальнейшего исследования особенностей диалектов.

## **ОЦИФРОВКА, РАЗМЕТКА И ПРЕОБРАЗОВАНИЕ АУДИОЗАПИСЕЙ В ЛГИС «ДИАЛЕКТ»**

Мультимедийная часть корпуса русских говоров Удмуртии представляет собой одну из составляющих лингвогеографической системы (ЛГИС) «Диалект» – многофункционального электронного лингвистического ресурса, целью создания которого являлось разностороннее изучение русских говоров Удмуртии. Если изначально данный ресурс был ориентирован на картографирование в масштабе региона слов, собранных для лексического атласа русских народных говоров на территории Удмуртии и, как всякий региональный атлас, предполагал детальное представление местных языковых особенностей, выявление существенных для данной

территории диалектных различий, то в ходе работы над ЛГИС «Диалект» появилось представление о необходимости пополнения системы языковым материалом иного формата, а также доработки с целью расширения ее возможностей.

В результате к собственно лингвогеографическому был добавлен лексикографический модуль, позволяющий осуществлять поиск слов в базе данных ЛГИС «Диалект». Далее был создан корпус русских говоров Удмуртии, состоящий из письменных текстов, собранных в ходе диалектологических экспедиций в 70–80-х гг. XX в. Основной целью создания корпуса русских говоров Удмуртии было получение дополнительных сведений о лексическом составе местного диалектного языка, нахождение контекстов для подтверждения и выявления лексико-семантических и грамматических особенностей диалектных слов.

Кроме того, работа с текстами корпуса позволила обнаружить ряд фонетических, морфологических и синтаксических диалектных черт, которые не могли быть зафиксированы при сборе материала по программе лексического атласа русских народных говоров (см. Жданова: 2022). В то же время при работе с диалектным материалом, зафиксированным в письменной форме, приходится учитывать, что мы имеем дело с интерпретацией собирателя, которая является более или менее точной в зависимости от его филологической подготовки, целей и условий записи.

В связи с этим было принято решение о создании мультимедийного подкорпуса русских говоров Удмуртии, который составили записи с дисков и кассет, сделанные в ходе сбора диалектного материала в 2000-е годы. Часть записей подкорпуса представляет опрос диалектоносителей по различным разделам программы лексического атласа русских народных говоров. Такие записи дают возможность убедиться в правильности фиксации диалектных слов, а также услышать их произношение, которое не отражалось при письменной фиксации. В ряде случаев диалектные лексемы употребляются в контексте, который дает необходимую иллюстрацию их значения.

Большую ценность представляют записи монологов диалектоносителей или их диалогов с односельчанами, в которых говорящий ведет себя более непосредственно. В таких случаях речь информанта демонстрирует большее количество фонетических и грамматических особенностей, которые могут стать предметом отдельного исследования.

Ранее в работе (Баранов, Верняева, Жданова: 2020) был представлен принцип аудиоразметки, которым мы пользовались при фрагментировании записей.

В настоящей статье необходимо подробнее рассказать о:

- 1) значении оцифровки аудиозаписей и их загрузке в базу данных,
- 2) предпринятых попытках конвертирования звуковых файлов в письменную форму для последующей синхронизации текста и идентичного ему по содержанию аудиофайла.

1. Диалектный аудиоматериал, несомненно, является важным источником при анализе фонетических особенностей языка: звучащая речь позволяет уточнить произношение той или иной единицы говора, позицию ударения в слове и т. д.

В базу данных ЛГИС «Диалект» в настоящее время загружена 41 аудиозапись.

Первоначально оцифрованные записи были конвертированы в формат WAV в связи с необходимостью предварительной и промежуточной обработки, что в дальнейшем помогло избежать потерь качества записи, которые могли произойти при многократном перекодировании. Затем с целью уменьшения размера файла, представляемого пользователю, было принято решение конвертировать оцифрованные записи в формат OGG.

Инструменты, принцип оцифровки и загрузка в базу данных были описаны ранее в работах (Верняева, Жданова: 2018; Баранов, Верняева, Жданова: 2020). Технически функции разметки реализованы на стороне браузера, используется JavaScript, HTML 5 и библиотека для работы с аудио WaveSurfer. Взаимодействие с базой данных происходит через серверную часть, реализованную на Java с использованием фреймворка (библиотеки) Spring Boot и Thymeleaf.

Работу по загрузке аудиоматериала и дальнейшей работе с ним можно представить в виде схемы: оцифровка -> загрузка в БД -> разметка в БД -> автоматическая загрузка в словарь системы / автоматическое отображение на карте региона.

После загрузки в базу данных оцифрованные аудиозаписи автоматически публикуются на странице «Тексты» ЛГИС «Диалект». Зарегистрированный пользователь имеет возможность прослушать данные записи и исследовать языковые особенности текста.

Размеченные аудиозаписи позволяют осуществлять полный лингвистический анализ, в частности выявлять фонетические особенности говоров, формировать запрос по словоформам и автоматически получать закрепление той или иной диалектной единицы на карте.

2. Сегодня в сети Интернет существует большое количество онлайн-инструментов, позволяющих конвертировать файл звукового формата в письменный текст.

Помимо онлайн-ресурсов существуют локальные программы, помогающие решить интересующую нас задачу. Однако большая часть такого программного обеспечения является платной, свободно распространяемые же инструменты как правило справляются с поставленной задачей с большими погрешностями.

Мы проанализировали и протестировали возможности нескольких популярных программ, предназначенных для преобразования звучащей речи в письменный текст и рекомендуемых специалистами, на одном и том же диалектном тексте.

В целом результаты работы инструментов не отличаются высоким качеством.

1) Программа *RealSpeaker* (<https://realspeaker.net/>) является платной. Перевод аудио в текст можно осуществить бесплатно только в том случае, если длительность аудиозаписи не превышает полутора минут.

Нами данный продукт протестирован. Однако использовать его в качестве инструмента не представляется возможным, так как ряд записей превышают длительность, равную полутора минутам. Кроме того, необходимо отметить, что описываемая программа недостаточно качественно транскрибирует диалектный текст.

2) *Сервис Transcribe* (<https://transcribe.wreally.com/>) позволяет расшифровать файл либо надиктовать текст. Программа является платной, но существующая демоверсия позволяет пользоваться сервисом в течение 7 дней, получать автоматическую транскрипцию 30-минутной записи. Тестирование программы

на материале диалектной аудиозаписи показало очень низкое качество транскрипции, в частности нераспознавание части слов, замена слов.

- 3) *Konverter App* (<https://converter.app/ru/mp3-text/>) работает только с форматом MP3, что является неудобным в нашем случае, так как аудиозаписи в ЛГИС «Диалект» загружены в формате \*ogg.
- 4) *AnyConv* (<https://anyconv.com/ru/konverter-audio-v-txt/>) не справился с распознаванием диалектного текста. В результатах выдал ошибку распознавания.
- 5) *Сервис Gglot* (<https://gglot.com/ru/mp3-to-text/>) так же, как *Konverter App*, работает только с форматом MP3.
- 6) *360Converter* (<https://www.360converter.com/>) удачнее всех остальных преобразовал звучащую речь в текст, автоматически разделил текст на фрагменты с возможностью проставления временного интервала соответствующего фрагмента записи. Однако текст содержит достаточно большое количество ошибок, для дальнейшей загрузки такого текста в базу данных необходима проверка преобразованного текста специалистом.

Таким образом, из шести программных продуктов с задачей справился лишь один. Однако использовать его в качестве инструмента для преобразования звучащего диалектного текста в письменный не представляется возможным в связи с достаточно большим количеством ошибок, что недопустимо при автоматической работе со специфическим текстом.

## РОЛЬ МУЛЬТИМЕДИЙНОГО КОРПУСА В ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Несмотря на отсутствие у аудиозаписей письменного аналога, на текущий момент мультимедийный подкорпус русских говоров Удмуртии дает широкие возможности для изучения диалектных особенностей различных языковых уровней.

При работе с текстами мультимедийного подкорпуса отмечены следующие фонетические черты русских говоров Удмуртии: повсеместное оканье и еканье, отдельные случаи ёканья (*платьё, полё, своёй* (с. Черёмушки Можгинского района)), упрощение групп согласных на конце слов (*есь* (есть), *шесть* (шесть)), отвердение губных на конце слов (*церкоф*), произношение долгого твердого [ш] на месте [ш'] (*благовешение, женишина, ташиым, подходяшые*), утрата взрывного элемента аффрикаты *ц* (*больниса, младенес* (с. Новая Казмаска Завьяловского района)), прогрессивная ассимиляция заднеязычного *к* по мягкости (*мамонька, тятенька* (с. Новая Казмаска Завьяловского района)), утрата интервокального *ј* со стяжением гласных в ряде случаев (*отвечашь* (отвечаешь), *ругаца* (ругается), *называм* (называем), *подвеснаа* (подвесная), *каку-нибудь* (какую-нибудь)).

Был отмечен также ряд грамматических особенностей:

- диалектные формы местоимений: *мене* (дательный падеж местоимения первого лица единственного числа), *имя́* (дательный падеж местоимения третьего лица множественного числа);

- совпадение форм творительного падежа множественного числа имен существительных и местоимений с формами дательного падежа: *за плечам, с носкам, с нам*;
- своеобразное морфемное оформление некоторых слов: *дожидать* (литературное *дожидаться*), *казать* (литературное *показывать*), *стараюся* (литературное *стараюсь*), *хороняют* (хоронят);
- частицы *дак, ак* (препозитивная), *то* (постпозитивная).

Особое внимание при работе с текстами мультимедийного подкорпуса было уделено диалектным особенностям, которые можно связать с влиянием на русские говоры Удмуртии финно-угорских языков: удмуртского и коми на севере республики. Ярким примером влияния удмуртского языка является произношение палатальных («шепелявых») [сʰ] и [зʰ] на месте соответствующих русских палатализованных согласных, однако в корпусе аудиозаписей подобное явление отмечено только в речи одной из жительниц д. Кабачигурт Игринского района, родным языком которой является удмуртский: наименование населенного пункта Селигурт звучит как [сʰелʰигурт]. Такое произношение в данном случае можно объяснить межъязыковой интерференцией в речи билингва.

С влиянием финно-угорских языков и удмуртского языка в частности связывают неразличение аффрикат во многих севернорусских говорах. Примеры этого явления (чоканье или цоканье) можно найти в письменных записях корпуса. Но анализ аудиозаписей не подтверждает наличие в современных русских говорах Удмуртии данного фонетического явления. В аудиоматериалах отмечены отдельные примеры мены аффрикат: *тысячка, тысечкой* (тысяцкий) (с. Карсовой Балезинского района), *окрошница* (окрошка) (с. Малиново Завьяловского района), *пекнича* (пятница) (д. Кабачигурт Игринского района, в речи женщины-удмуртки), *цело* (чело, передняя часть печи) на юго-востоке Удмуртии (с. Светлое Воткинского района). Однако других случаев неразличения аффрикат в речи данных информантов не встречалось, что свидетельствует о лексикализации упомянутых традиционных диалектных наименований. Таким образом, можно говорить о присутствии в русских говорах Удмуртии следов чоканья и цоканья, но не о живом фонетическом явлении. В связи с этим стоит отметить, что сопоставление данных корпуса русских говоров Удмуртии, зафиксированных в разные годы, помогает увидеть направление изменений местного диалектного языка, сохранение одних и нивелировку других языковых явлений.

Среди грамматических особенностей, связанных с влиянием финно-угорских языков, можно отметить выражения *да чо да, ли чо ли* (*Три ли чо ли года*), отмеченные в устной речи диалектоносителей и являющиеся характерной чертой русских говоров Удмуртии. Заметим, что эти сочетания можно найти и в письменной части корпуса, а также услышать в речи жителей республики всех возрастов.

Другой грамматической особенностью, которая объясняется влиянием финно-угорского языка, в данном случае коми, является употребление форм именительного падежа существительных вместо родительного в конструкциях, указывающих на количество или отсутствие предметов. Это явление в русских говорах севера Карсвайского района (на северо-востоке) Удмуртии описал В.Ф. Барашков (Барашков: 1957). В корпусе письменных записей все отмеченные примеры зафиксированы также

в этой части республики. В мультимедийном подкорпусе обнаружено единственное употребление такой конструкции: *Трое дети были*, – также записанное в с. Карсовай, что подтверждает присутствие в русских говорах Удмуртии данного явления и его территориальную ограниченность.

Таким образом, тексты подкорпуса аудиозаписей русских говоров Удмуртии можно считать объективным источником информации о фонетических и грамматических чертах данных говоров, которые при его помощи можно исследовать в различных аспектах.

Описание записей в корпусе содержит указание на год, фамилию информанта и собирателя. Другие сведения об информантах, а также тематика разговора обычно представлены в начале аудио.

Благодаря разметке мультимедийного подкорпуса, соответствующей общим целям и задачам, стоящим перед разработчиками ЛГИС «Диалект», слова из аудиозаписей обнаруживаются при запросе в лексикографическом модуле ЛГИС «Диалект», дополняя список примеров употребления и уточняя географию распространения слова. Также лексемы, отмеченные в аудиозаписях, попадают в выборку по соответствующему вопросу при построении лингвистической карты, дополняя ее. Так, при разметке в записи, сделанной в д. Кабачигурт Игринского района, отмечено диалектное наименование брюквы *калега* (рис. 1).

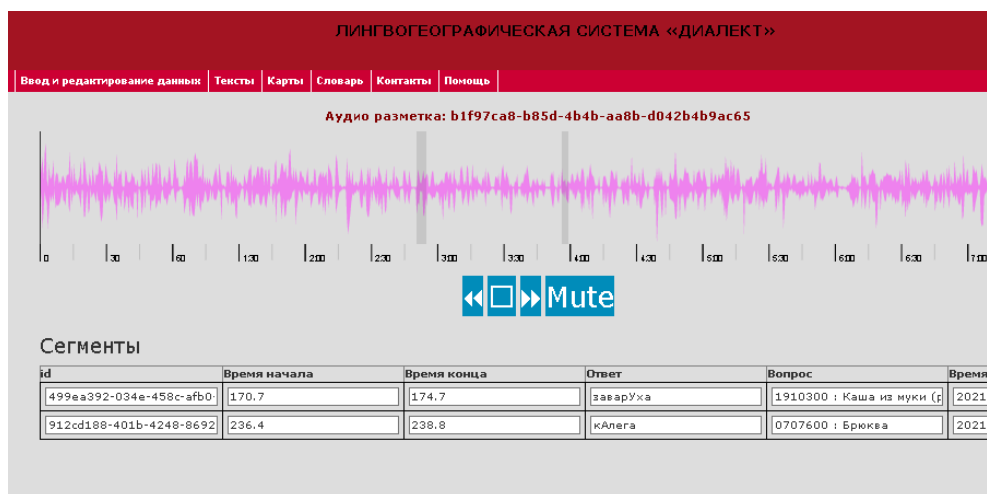


Рис. 1. Результаты аудиоразметки записи

При запросе в лексикографическом модуле ЛГИС «Диалект» можно увидеть эту фиксацию среди других (рис. 2).



**ЛИНГВОГЕОГРАФИЧЕСКАЯ СИСТЕМА «ДИАЛЕКТ»**

**Тексты** | **Карты** | **Словарь** | **Контакты** | **Помощь**

**кАлега (52)**

**Брюквя**

республика Удмуртская, Алнашский район, село Гарга ( [1999-11-03..1999-11-09](#) )  
 Балезинский район, село Киршонки ( [2009-07-10..2009-07-21](#) )  
 Завьяловский район, село Завьялово ( [2001-01-01..001-01-01](#) [1999-01-01..001-01-01](#) )  
 село Люк ( [1998-08-26..001-01-01](#) )  
 село Совхозный ( [2009-07-07..2009-07-18](#) [2009-07-07..2009-07-13](#) )  
 Игринский район, поселок Игра ( [2007-01-01..001-01-01](#) )  
 Киясовский район, село Первомайский ( [2007-08-21..2007-08-25](#) )  
 Красногорский район, село Красногорское ( [2000-01-01..001-01-01](#) )  
 Сарапульский район, село Девятково ( [1998-07-05..1998-07-26](#) )  
 Селтинский район, село Узи ( [1998-08-04..001-01-01](#) )  
 Сюмсинский район, поселок Кильмезь ( [2001-01-01..001-01-01](#) [1999-07-01..001-01-01](#) )  
 село Сюмси ( [2003-06-25..001-01-01](#) )  
 Якшур-Бодьинский район, село Чур ( [2003-01-01..001-01-01](#) )  
 село Якшур-Бодья ( [1999-01-01..001-01-01](#) )  
 Удмуртская Республика, Воткинский район, село Перевозное ( [1984-01-01..001-01-01](#) )  
 Завьяловский район, село Гольяны ( [1978-01-01..001-01-01](#) )  
 Игринский район, деревня Кабачигурт ( [2021-01-02..001-01-01](#) )  
 Сарапульский район, деревня Горбуново ( [1973-01-01..001-01-01](#) )  
 Шарканский район, деревня Кочки ( [1985-01-01..001-01-01](#) )  
 село Зюзино ( [1985-01-01..001-01-01](#) )  
 Якшур-Бодьинский район, село Старые Зятцы ( [1975-01-01..001-01-01](#) )

Медиа файлы: [Россия, Удмуртская Республика, Сарапульский район, деревня Горбуново, 1973, 33 Гор.jpg](#)  
[Россия, Удмуртская Республика, Шарканский район, село Зюзино, 1985, 20 Зюз.jpg](#)  
[Россия, Удмуртская Республика, Завьяловский район, село Гольяны, 1978, 7 Голь.jpg](#)  
[Россия, Удмуртская Республика, Игринский район, деревня Кабачигурт, 2021, shirkova 4.ogg](#)  
[Россия, Удмуртская Республика, Воткинский район, деревня Двигатель, 1983, 16 Двиг.jpg](#)  
[Россия, Удмуртская Республика, Воткинский район, село Перевозное, 1984, Image \(11\).jpg](#)  
[Россия, Удмуртская Республика, Якшур-Бодьинский район, село Старые Зятцы, 1975, 10 СЗ.jpg](#)  
[Россия, Удмуртская Республика, Шарканский район, деревня Кочки, 1985, 17 Коч.jpg](#)  
[Россия, Удмуртская Республика, Шарканский район, деревня Кочки, 1985, 7 Коч.jpg](#)  
[Россия, Удмуртская Республика, Сарапульский район, деревня Горбуново, 1973, 51 Горб.jpg](#)

Рис. 2. Запрос в лексикографическом модуле

Также данное слово отражено на лингвистической карте, показывающей названия брьюквы, распространенные на территории Удмуртии (рис. 3).

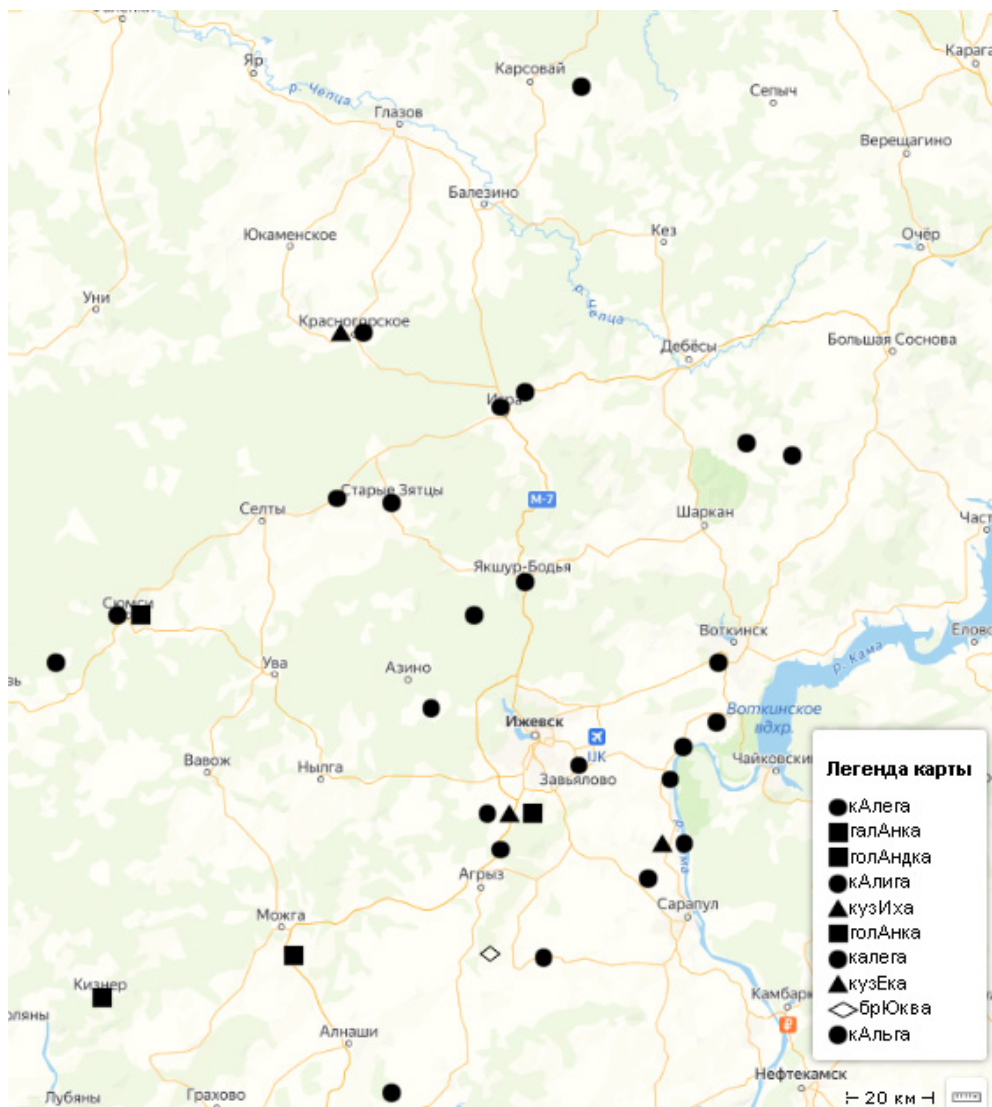


Рис. 3. Отражение искомого слова на лингвистической карте

Диалектные наименования, отмеченные в мультимедийном подкорпусе, могут послужить для уточнения характеристики слов, зафиксированных в других материалах корпуса русских говоров Удмуртии, а также карт, составленных на их основе.

В записи, сделанной в с. Черёмушки Можгинского района, отмечено слово *девятерик*: *Стоит пенёк... и вылетает оттуда этот, мы звали девятерик, это как оса*. При поиске в лексикографическом модуле было найдено еще одно употребление этого слова в значении «шмель» в тексте, записанном в д. Русская Коса Кизнерского района в 1985 г. При обращении к контексту было обнаружено,

что значение «шмель» было приписано диалектному слову приблизительно, поскольку прямого определения и указания на обозначаемый предмет, как и в первом случае, контекст не содержит: *(внучка информанта): баушка, девятерик (почему его так называют?)* [в скобках даны примечания собирателя]; – *говорят, девять жал у него, как шмель, токо большой, вон летат*. Судя по определению, в данном случае информант говорит скорее о шершне, чем об осе, однако возможно, что носители говоров не различают этих насекомых, называют их одним и тем же словом, поэтому, принимая во внимание отсутствие вопроса «шершень» в вопроснике ЛГИС «Диалект», можно считать слово *девятерик* наименованием осы и шмеля. Заметим, в словаре русских народных говоров слово *девятерик* приведено как название неопределенного насекомого в говорах Вятской губернии (1896 г.) (СРНГ 7: 323), в состав которой в то время входили территории современных Можгинского и Кизнерского районов Удмуртии. В материалах к Лексическому атласу русских народных говоров слово *девятерик* отмечено как название шмеля в Куженерском и Медведевском районах республики Марий Эл и в Кизнерском районе Удмуртии (но не отражено на карте), а также приводятся записи, свидетельствующие о неразличении диалектоносителями шершней и шмелей, шершней и ос (ЛЯРНГ 2: 666, 672). Таким образом, запись устной речи, сделанная в 2000 году, помогла уточнить значение диалектного слова, зафиксированного в более ранних записях.

## ВЫВОДЫ

Корпусы устной русской речи, в том числе диалектной, являются важным источником для изучения народных говоров. Более того, роль диалектных корпусов увеличивается по мере того, как традиционные говоры уходят в прошлое и записи остаются одним из немногих источников, достоверно передающих звучание, грамматическое строение и лексико-семантическое своеобразие диалектной речи, а инструментарий корпуса облегчает поиск информации для исследования различных сторон диалектного языка.

Мультимедийный подкорпус корпуса русских говоров Удмуртии в ЛГИС «Диалект» существует и развивается с 2018 года. Он пополняется новыми записями диалектной речи, производится разметка введенных записей, продолжают попытки конвертировать звуковые файлы в письменный текст для облегчения работы с ними. К сожалению, доступные системы распознавания устной речи пока не справляются с этой задачей на необходимом уровне.

В то же время уже в существующем формате корпус аудиофайлов ЛГИС «Диалект» дает исследователям русских говоров Удмуртии широкие возможности:

1. Прослушивание полных записей подкорпуса позволяет выявлять и анализировать фонетические и грамматические особенности диалектной речи. В отличие от письменных записей, где речь может быть записана отрывочно и неточно затранскрибирована, аудиозаписи дают более достоверное представление о строении диалектной речи.

2. Удмуртия является мультилингвальным регионом. В этом плане аудиозаписи корпуса помогают обнаружить признаки взаимодействия языков, распространенных в республике, подтвердить или опровергнуть предположения о влиянии других языков на русские народные говоры.
3. Слова, отмеченные в аудиозаписях в результате разметки, отражаются в электронном словаре русских говоров Удмуртии, а также на картах, которые может составить пользователь ЛГИС «Диалект». Таким образом, данные слова дополняют и расширяют представление о языковом ландшафте Удмуртии.

Речь диалектоносителей, представленная в виде аудиозаписей, а также результаты разметки этой речи являются новым языковым материалом, уточняющим сведения о лексико-семантических особенностях русских говоров Удмуртии.

## БИБЛИОГРАФИЯ

- REGINA A. VERNYAEVA, EKATERINA A. ZHDANOVA (2018): "The corpus of Russian dialects of the Republic of Udmurtia: Presentation of a multimedia material", *El'Manuscript. 7th International Conference on Textual Heritage and Information Technologies: Abstracts, Participants, Programme (Vienna and Krems, Austria. 14–18 September 2018)*. Eds. Victor Baranov, Patricia Engel, Jurgen Fuchsbauer, Heinz Miklas. Vienna: Verlag Holzhausen. Scientific Publishing, 2018. P. 67. ISBN 978-3-903207-28-8.
- БАРАНОВ, В. А., ВЕРНЯЕВА, Р. А., ЖДАНОВА, Е. А. (2020): «Мультимедийный корпус русских говоров Удмуртии: разработка и возможности использования». *Cuadernos de Rusística Española, T. 16, с. 39-54*.
- БАРАШКОВ, В. Ф. (1957): Русский говор северной части Карсвайского района Удмуртской АССР: дис. ... канд. наук, Москва, 312 с.
- ЖДАНОВА, Е. А. (2022): «Корпус русских говоров Удмуртии как инструмент для изучения фонетических и грамматических диалектных особенностей». *Социально-экономическое управление: теория и практика, Т. 18, 1, с. 89-94*.
- Лексический атлас русских народных говоров (ЛАРНГ): т. 2. Животный мир.* (2022): Москва; Санкт-Петербург, Нестор-История.
- Словарь русских народных говоров (СРНГ)* (1965–2018): Вып. 1–51, Москва; Ленинград, Издательство «Наука».

## BIBLIOGRAPHY

- REGINA A. VERNYAEVA, EKATERINA A. ZHDANOVA (2018): "The corpus of Russian dialects of the Republic of Udmurtia: Presentation of a multimedia material", *El'Manuscript. 7th International Conference on Textual Heritage and Information Technologies: Abstracts, Participants, Programme (Vienna and Krems, Austria. 14–18 September 2018)*. Eds. Victor Baranov, Patricia Engel, Jurgen Fuchsbauer, Heinz Miklas. Vienna: Verlag Holzhausen. Scientific Publishing, 2018. P. 67. ISBN 978-3-903207-28-8.

- BARANOV, V. A., VERNYAEVA, R. A., ZHDANOVA, E. A. (2020): “Mul'timediinyi korpus russkikh govorov Udmurtii: razrabotka i vozmozhnosti ispol'zovaniya”. *Cuadernos de Rusística Española*, T.16, pp. 39-54.
- BARASHKOV, V. F. (1957): “Russkii govor severnoi chasti Karsovaiskogo raiona Udmurtskoi ASSR: dis. ... kand. nauk”, Moskva, 312 p.
- ZHDANOVA, E. A. (2022): “Korpus russkikh govorov Udmurtii kak instrument dlya izucheniya foneticheskikh i grammaticheskikh dialektnykh osobennostei”. *Sotsial'no-ekonomicheskoe upravlenie: teoriya i praktika*, T. 18, 1, pp. 89-94.
- Leksicheskii atlas russkikh narodnykh govorov (LARNG): T. 2. Zhivotnyi mir.* (2022): Moskva; Sankt-Peterburg, Nestor-Istoriya.
- Slovar' russkikh narodnykh govorov (SRNG)* (1965–2018): Vyp. 1–51, Moskva; Leningrad, Izdatel'stvo «Nauka».