

Плюсы и минусы удаленной документации:

сравнение опыта работы с языками Дагестана, Сибири и Дальнего Востока

Н. Б. Аралова^{4,5}, Е. М. Будянская¹, М.-Э. А. Винклер¹, Т. И. Давидюк^{1,2},
А. П. Евстигнеева¹, Ю. Б. Коряков^{1,3}, Д. Д. Мордашова^{1,2}, В. С. Харитонов¹

¹Институт языкознания РАН, ²МГУ им. М. В. Ломоносова, ³НИУ «Высшая школа экономики»,
⁴Westfälische Wilhelms-Universität Münster, ⁵Christian-Albrechts-Universität zu Kiel

О понятии и теме доклада

Удаленная документация — такой вид сбора и/или обработки данных, когда лингвист работает с участниками исследования дистанционно, не находясь непосредственно в месте проживания языкового сообщества, т.е. с помощью различных мессенджеров, электронной почты или платформ в Интернете.

Цели доклада:

- поделиться нашим опытом;
- обсудить факторы, влияющие на дизайн проекта и конкретную методологию;
- оценить преимущества и ограничения в представленных проектах.

План доклада

1. Удаленная документация в мировой практике

2. Наш опыт
 - a. анклухский и амухский диалекты цудахаро-сирхинского языка
 - b. татский язык
 - c. мансийский язык
 - d. уильтинский (орокский) язык
 - e. северноюкагирский & якутский языки

3. Обсуждение и некоторые выводы

Удаленная документация

- Английский термин: *remote documentation*, позднее *distributed documentation*
- Активное развитие и обсуждение методологии в научном сообществе в последние годы (вызванное в том числе пандемией Ковид-19):
 - в 2021 г. ELDP организовал серию воркшопов ELDP Remote Fieldwork workshop series (использование скриптов для обработки звуковых файлов в WhatsApp, использование YouTube для перевода текстов и др.)
 - появляются публикации: Leemann et al. 2020, Williams et al. 2021, Grzech, Shaca 2022, Gabber & Vondiziano 2023 и др.
- Развивается также и в России в работе с малыми языками, также особенно бурно в последние несколько лет.

Pro et contra

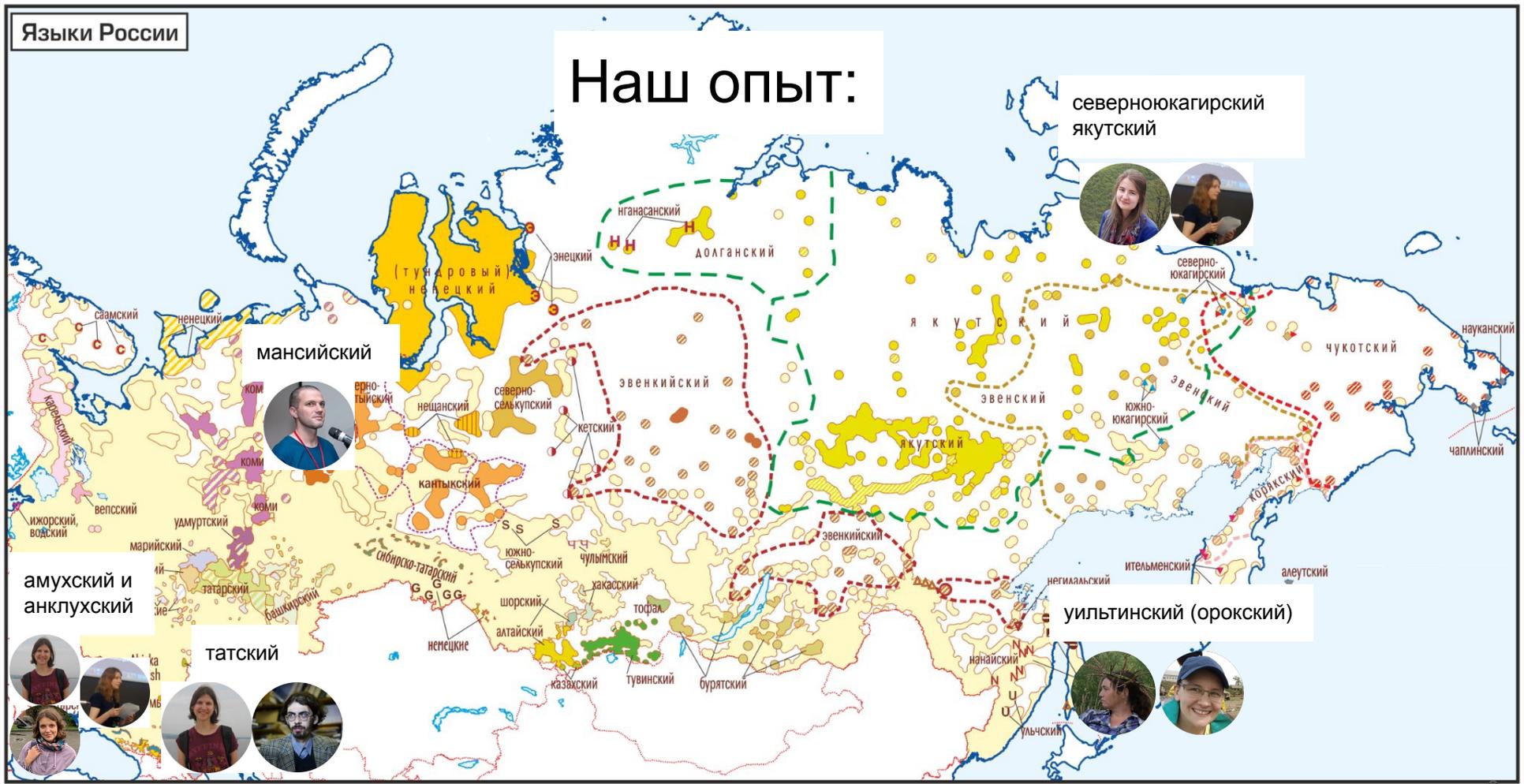


- Поездки сложно (и иногда невозможно) организовать, связь по Интернету дешевле и во многом проще;
- когда осталось всего несколько носителей, не хочется ограничивать работу с ними только временем экспедиций;
- возможно совмещать полевую работу и другую академическую деятельность;
- большая свобода для представителей языкового сообщества в выборе, когда и сколько работать;
- движение в сторону “документации силами языкового сообщества”;
- возможность финансово поддержать коллег в деле документации языка в течение года, а не только во время экспедиции.
- ...



- Сложно (но можно!) наладить контакт без предварительной работы;
- отсутствие Интернета в местах проживания языкового сообщества;
- отсутствие соответствующей техники и навыков работы на ней;
- качество звука (если записываем звуковые файлы, много факторов, которые надо проконтролировать + предыдущий пункт);
- рабочая сессия ограничена парой часов;
- личное общение приятнее.
- для некоторых задач не подходит
- ...

Наш опыт:



Параметры

Про каждый проект мы расскажем в следующем порядке:

- статус витальности языка (проект “Языки России” [Коряков и др. 2022]¹)
- задачи
- ход работы и инструменты
- наши помощники
- достоинства метода
- ограничения метода
- технические требования

¹Доступна по ссылке: [http://jazykirf.iling-ran.ru/\(2022\)_Spisok_jazykov_Rossii_Monograph.pdf](http://jazykirf.iling-ran.ru/(2022)_Spisok_jazykov_Rossii_Monograph.pdf)

Анклухский и амухский диалекты цудахаро-сирхинского языка (< даргинские языки): статус 2А (прерванные)

Задачи: разбор записанных в экспедиции текстов, словаря и разговорника

Ход работы и инструменты: лингвист присылает помощнику записанный материал, помощник записывает его при помощи даргинского алфавита и переводит и отправляет на электронную почту

Наш помощник: Османов Рамазан Гасангусейнович, пенсионер, носитель анклухского диалекта. Самостоятельно выучил даргинскую письменность.



Анклухский и амухский диалекты цудахаро-сирхинского языка

Достоинства: постоянная связь с носителем (в т.ч. обсуждение лингвистических моментов), монетизация знания языка, удобный распорядок работы

Ограничения: влияние литературной нормы; отложенное уточнение неточностей расшифровки; необходимость предварительной подготовки записей (нарезки и увеличения громкости) в силу индивидуальных ограничений помощника

Технические требования: компьютер, навык работы на компьютере, Интернет

Татский язык: статус ЗА(Р) (локализованные)

Задачи: разбор записанных в экспедиции текстов

Ход работы: лингвист присылает помощнику записанный материал, помощник записывает его при помощи алфавита и переводит и отправляет текстовый файл в мессенджер

Наш помощник: Агарагимов Амир Аразович, 11-классник, носитель джалганского диалекта татского языка, самостоятельно предложил вариант письменности для татского языка, который затем был доработан с помощью лингвистов



Татский язык

Достоинства: налаженная связь с носителем (в т.ч. обсуждение лингвистических моментов), монетизация знания языка

Ограничения: неустойчивость письменности; отложенное уточнение неточностей расшифровки; значительные перерывы в работе

Технические требования: компьютер, навык работы на компьютере, Интернет

Мансийский язык: статус 2Б (прерывающиеся)

Задачи: пересказы (видеороликов, мультфильмов) устные и письменные, запись нарративов, анкеты (элицитация).

Перекрёстные и множественные пересказы — для оценки вариативности в дискурсе.

Ход работы: активная фаза — февраль–май 2021

Сейчас (весна 2023): обучение Elan для разбора (фонетического, морфологического, перевода)

Наши помощники: Галина Дунаева, Андрей Дунаев, Евгений Анямов, Софья Власенкова и др.

Мансийский язык

На скриншоте —
Галинина работа по
расшифровке.

Учится глоссировать

The screenshot displays the ELAN 0.4 software interface. The top menu bar includes 'Файл', 'Правка', 'Аннотация', 'Слой', 'Тип', 'Поиск', 'Вид', 'Параметры', 'Окно', and 'Справка'. Below the menu is a toolbar with tabs for 'Таблица', 'Полный текст', 'Субтитры', 'Лексикон', 'Комментарии', 'Распознаватель', 'Метаданные', and 'Регулировка'. The main area features two sliders: 'Громкость' (Volume) and 'Скорость' (Speed), both set to 100. Below these is a timeline with playback controls and a selection tool. The bottom section shows a transcription table with two rows: 'Mansi' and 'Russia'. The 'Mansi' row contains the transcription 'да? porse xotal' an' высыпали | am at wajlum jatapaln jaleim sunseim tax xotal' - nan jaln n | Коля kitelas laj rde un't'. The 'Russia' row contains 'ать сюда. | да? мусором куда-то сейчас высыпали. | я не знаю на ту сторону схожу посмотрю. - Ты сходи но. | Коля спрашивал ..где'.

Language	Transcription
Mansi	да? porse xotal' an' высыпали am at wajlum jatapaln jaleim sunseim tax xotal' - nan jaln n Коля kitelas laj rde un't
Russia	ать сюда. да? мусором куда-то сейчас высыпали. я не знаю на ту сторону схожу посмотрю. - Ты сходи но. Коля спрашивал ..где'

Мансийский язык

Достоинства: удобное время (без созвонов), легко придумывать задания, вклад в обучение

Ограничения: плохая фонетика (качество записи)

Технические требования: (не обязательно стабильный) интернет; смартфон; листочек и ручка - для письменных работ; мессенджеры

Уильтинский (орокский) язык: статус 2А- (прерванные)

Задачи: 1.расшифровка ранее записанных текстов, 2. сбор фольклорных образцов малой формы (загадки, “выражения воспитывающего характера”)

Ход работы: 1. лингвист подготавливает транскрипцию в ELAN’е, созванивается с помощниками по ZOOM’у, вместе слушают, исправляют подготовленную транскрипцию и переводят -- наиболее приближенно к работе в поле.

2. Помощник-носитель записывает голосовыми сообщениями в WhatsApp короткие образцы фольклора, присылает фото рукописи и перевода.

Наши помощницы: Л.Х. Минато (южный диалект, преподает язык в школе) и Е.А. Бибилова (северный диалект, активно сотрудничает с учеными, занимается публикацией книг на родном языке), обе участвовали в создании уильтинского букваря

ELAN 6.3 - LH_palochki_21.eaf

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Lexicon Comments Recognizers Metadata Controls

XXX.Text-txt-oaa

Nr	Annotation	Begin Time	End Time	Duration
1	əri ulta purilni xupikəni mō mōma xupikə	00:00:...	00:00:...	00:00:...
2	bi nočika biḡəḡə hupixəmbi biči bakka	00:00:...	00:00:...	00:00:...
3	əri jewa mō	00:00:...	00:00:...	00:00:...
4	ziḡdoju dūdoju ilādoju tudmō bini ēdu hupičči	00:00:...	00:00:...	00:00:...
5	čoči nari puril tundajū xasu xasuju nuḡuju puril təsiči gəḡə mār	00:00:...	00:00:...	00:00:...
6	dappəni mōwa gum ḡālaži	00:00:...	00:00:...	00:00:...

00:00:20.561 Selection: 00:00:20.561 - 00:00:26.652 6091

LH_palochki_21.wav 1:20.000 00:00:21.000 00:00:22.000 00:00:23.000 00:00:24.000 00:00:25.000 00:00:26.000 00:00:27.000

Interlinear-title-en [0]

XXX.Text-txt-oaa [60] zḡḡdoju dūdoju ilādoju tudmō bini ēdu hupičči | čoči na

XXX.Translation-gls-r [3]



Лена Будянская



Natalia Aralova



Елена Бибикина

Уильтинский язык

Достоинства: активное участие носителя, постоянная связь с информантом, процесс не ограничен временем экспедиции, сессия в zoom дает возможность уточнять вопросы быстро, как при очной сессии

Ограничения: техническая оснащенность информантов (тихие записи не слышно без наушников), разница во времени

Технические требования: компьютер, навык работы на компьютере, Интернет, в идеале наушники

Северноюкагирский (2А, прерванные) и якутский (3В, ограниченные городские)

Задачи: расшифровка юкагирских текстов, записанных в экспедициях; расшифровка видео 2005 г. от носителя нижнеколымского диалекта якутского

Ход работы: лингвист готовит шаблон для разбора в ELAN с нужным количеством слоев, проводит инструктаж по работе в ELAN'е и разбирает вместе с помощником несколько первых предложений; затем помощник присылает файл с полным разбором текста по эл. почте или иным удобным способом (технические моменты разбора при необходимости обсуждаются в мессенджерах)

Наши помощницы: Самона Николаевна Курилова (северноюкагирский язык, пассивное владение языком, научный сотрудник-лингвист); Анна Платонова (литературный вариант якутского языка, художник, дизайнер)



Северноюкагирский и якутский

Достоинства: процесс не ограничен временем экспедиции, работа в удобном для носителя темпе, оплата труда

Ограничения: необходимо предварительное обучение и заранее подготовленная инструкция; точность расшифровки зависит от носителя (расхождение в якутских диалектах одновременно и плюс, и минус: носитель обращает внимание на диалектные особенности, но не всегда может расшифровать абсолютно всё); для тихих записей нужны наушники

Технические требования: компьютер, навык работы на компьютере, Интернет, в идеале наушники

Северноюкагирский и эвенский (2Б, прерывающиеся): негативный опыт

Северноюкагирский: носительница (около 30-35 лет) на высоком уровне освоила ELAN в ходе очной работы с лингвистом, успешно расшифровывала тексты в присутствии лингвиста, однако дистанционную работу не удалось наладить (см. также доклад [Мордашова, Пупынина & Сидорова 2019]).

Эвенский: носительница получила устную инструкцию о расшифровке (транскрипция и перевод по предложениям в текстовом редакторе), однако не вполне поняла задачу + не смогла уточнить задание из-за плохой связи в населенном пункте и выполнила задачу в формате, отличающемся от необходимого.

Обсуждение

- 1) В каждом проекте свои индивидуальные особенности и своя конфигурация: разнообразие методов позволяет выбрать подходящий формат и гибко подстраиваться под обстоятельства. Неожиданные открытия и удачи:
 - и пожилые могут освоить зум!
 - а в случае слабого интернета можно посылать, например, голосовые сообщения по WhatsApp'у
- 2) Меняется структура проекта, ритм работы и взаимодействия с сообществом:
 - от формата “кабинет–экспедиция–кабинет” мы переходим к регулярной работе в течение года;
 - меняются роли участников: над одним и тем же материалом могут работать участники проекта, которые физически находятся в разных местах (это обычное свойство Интернет-коммуникации, перенесенное на ситуацию полевой работы);
 - вовлечение представителей сообщества в работу с языком на регулярной основе → повышение престижа языка, открытие новых возможностей, освоение новых навыков;
 - пролонгированный контакт.
- 3) Большая зона ответственности у представителей сообщества.

Спасибо за внимание!

Литература и другие источники

Коряков и др. 2022 — Коряков Ю. Б., Давидюк Т. И., Харитонов В. С., Евстигнеева А. П., Сюрюн А. А. Список языков России и статусы их витальности. Монография-препринт. — Институт языкознания РАН, 2022. — 80 с.

Мордашова, Пупынина & Сидорова 2019 — Мордашова Д. Д., Пупынина М. Ю., Сидорова М. А. Тундренный юкагирский в свете ареальных контактов в Нижнеколымском районе. Доклад на конференции “Лингвистический форум 2019: Коренные языки России и мира”.

Gabber & Vondiziano 2023 — Gabber, Shirley, Vondiziano Gregory. 2023. Zooming through Field Methods: Teaching language documentation remotely. *Language Documentation & Conservation* 17: 23–48.

Grzech, Shaca 2022 — Grzech, Karolina, Selena Tisalema Shaca. 2022. Building trust on Zoom: A workflow for language documentation via videoconferencing software. *Language Documentation & Conservation* 16: 79–97.

Leemann et al. 2020 — Leemann, Adrian, Péter Jeszenszky, Carina Steiner, Melanie Studerus & Jan Messerli. 2020. Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing. *Linguistics Vanguard* 6(s3).

Rozhanskiy 2021 — Rozhanskiy F. 2021. *A Manual of Linguistic Fieldwork*. Tartu, 2021. 137 p.

Williams et al. 2021 — Williams, Nicholas, W. D. L. Silva, Laura McPherson & Jeff Good. 2021. COVID-19 and documentary linguistics: Some ways forward. *Language Documentation and Description* 20, 359–377.