

## ДАнные ДОКУМЕНТАЦИОННЫХ ПРОЕКТОВ ЗА ПРЕДЕЛАМИ ДОКУМЕНТАЦИИ ЯЗЫКА\*

Н. М. Стойнова

И. А. Хомченкова

ИРЯ им. В. В. Виноградова РАН / ИЯз РАН

ИРЯ им. В. В. Виноградова РАН / ИЯз РАН / МГУ

В процессе документации малых языков России исследователи (вполне закономерно) фокусируются на сборе аудио- и видеоматериалов, содержащих исключительно исследуемый язык и призванных обеспечить его полноценное описание. При этом лингвисту приходится задумываться и о том, что полученные им эксклюзивные данные в дальнейшем могут понадобиться для решения задач, далеко выходящих за пределы его непосредственных интересов, в особенности при работе с труднодоступным языком или языком, находящемся под угрозой исчезновения.

В нашем докладе мы хотим обсудить опыт работы с «побочным продуктом» документационных проектов: аудиозаписями на русском языке и записями с переключением кодов. Нашей задачей было, во-первых, описание нестандартного русского языка билингвоносителей малых языков Севера Сибири и Дальнего Востока, во-вторых, исследование его взаимодействия с автохтонным языком при переключении и смешении кодов в горномарийском, мокшанском, нанайском и ульчском языках. В докладе мы приведем практические соображения о том, насколько подходящим оказывается материал текущих и недавних проектов по документации малых языков (изначально не ориентированных на такие задачи) для подобного рода исследований.

Тексты на русском языке и тексты с большим количеством русских вкраплений, оказавшиеся в фокусе нашего внимания, как правило, признаются «негодными» или как минимум наименее ценными для документационного проекта. Они записываются часто случайно (например, вместо ожидаемого текста на исследуемом языке); к тому же их обрабатывают, описывают и хранят менее тщательно, чем основные данные проекта. В нашем докладе мы обсудим, в частности, следующие проблемы, с которыми мы столкнулись при работе с такими текстами:

а) лакуны в социолингвистической информации о носителях: как правило, очень мало или нет информации о времени и процессе освоения русского языка, о ситуациях его использования;

б) процесс записи: стоит ли прерывать человека, если он перестает говорить на своем языке и переходит на русский, стоит ли прерывать аудиозапись; насколько удобной для наших целей оказывается «сплошная» многочасовая запись vs. запись отдельных коротких фрагментов текста;

в) обработка записей: часть ценной для исследования переключения кодов и русского языка билингвов информации теряется при «нарезке» записей, удаления «лишних» фрагментов по краям текста и особенно в середине записи;

г) метаданные к текстам: далеко не всегда тексты на русском языке и с большими русскими вкраплениями снабжаются метаданными, далеко не всегда для текстов на автохтонном языке в метаданных указано, что в них есть фрагменты с переключением кодов и что для них есть параллельная версия на русском языке; наконец, практически никогда при наличии параллельных текстов на автохтонном языке и на русском не указывается, в каком порядке они были записаны (что достаточно важно при исследовании интерференции).

Таким образом, в докладе будут рассмотрены технические принципы документации языков, которые влияют на решение важных задач, вторичных по отношению к основному проекту. Материал доклада может быть полезен и в более общей дискуссии о том, как обеспечить максимальную универсальность полевых данных, насколько это возможно и в какой мере необходимо.

---

\* Исследование поддержано грантом РНФ #17-18-01649.