

## **ПРОГРАММА ELAN ДЛЯ ИССЛЕДОВАНИЯ ДИАЛЕКТОВ КОМИ ЯЗЫКА**

**В.П. Чупров,**

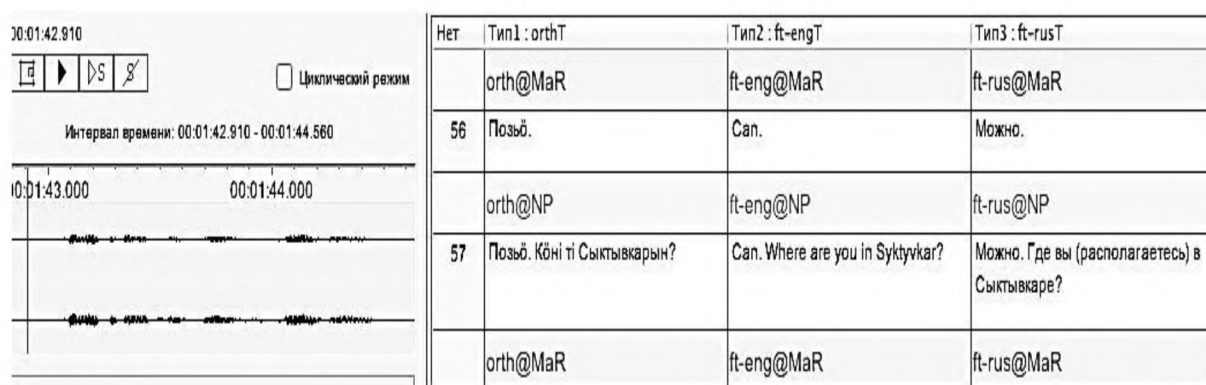
магистрант института гуманитарных наук  
ФГБОУ ВО «СГУ им. Питирима Сорокина»,  
переводчик Центра инновационных языковых технологий КРАГСИУ

Важными составляющими в диалектологических и фольклорных исследованиях является сбор, систематизация и хранение полевого материала. В век современных информационных технологий важно использовать на этих стадиях различные компьютерные технологии. Одна из таких современных программ – ELAN. Она разработана Институтом психолингвистики Макса Планка в Неймегене (Голландия) специально для архивирования записей языков, находящихся под угрозой исчезновения, в рамках проектов Фонда Фольксвагена. Цель этой программы – обеспечить надёжную технологическую основу для аннотации и эксплуатации мультимедийных записей. ELAN поддерживает:

- визуализацию аудио– и/или видеосигналов одновременно с полученными аннотациями;
- временную привязку аннотаций к медийному потоку;
- сложные связи аннотаций друг с другом;
- неограниченное количество задаваемых пользователем уровней аннотации;
- различные шрифты и кодировки;
- экспорт данных в виде текстовых файлов табличного вида (tab-delimited text);
- импорт и экспорт между ELAN, Praat, ToolBox, Shoebox и другими популярными лингвистическими программами;
- развитые поисковые опции.

Программу ELAN можно использовать и для анализа поведения людей (жестов, мимики, характера речи информантов, различных их действий в коммуникации). Создаются так называемые мультимодальные корпуса, т.е. корпуса, включающие запись звука и видео, собираются видеозаписи поведения людей в реальных эмоциональных ситуациях. Для исследования поведения информирующих корпус сопровождается разметкой: записывается текст, произнесённый людьми в кадре, их мимика, движение рук и тела.

В зависимости от числа участников коммуникации выделяются аннотации для орфографической транскрипции и переводов. В программе ELAN выделяют два типа аннотаций: независимые и зависимые. Независимые связаны со звуковым файлом во времени, зависимые связаны не со звуком, а с другой аннотацией. Например: орфографическая транскрипция относится к независимому типу аннотации, так как она напрямую связана со звуком, перевод – зависимый тип аннотации, поскольку он связан с транскрибированным предложением. В нашем случае независимый тип аннотации – это расшифрованный текст на ижемском диалекте коми языка, а зависимый тип – это переводы на русский и английский языки. На экране монитора аннотации записей выглядят следующим образом:



The screenshot shows the ELAN software interface. On the left, there is a timeline with a time range of 00:01:42.910 to 00:01:44.560. Below the timeline are two audio waveforms. On the right, there is a table with four columns: 'Нет', 'Тип1 : orthT', 'Тип2 : ft-engT', and 'Тип3 : ft-rusT'. The table contains two rows of annotations, with the second row highlighted in blue.

Нет	Тип1 : orthT	Тип2 : ft-engT	Тип3 : ft-rusT
	orth@MaR	ft-eng@MaR	ft-rus@MaR
56	Позьб.	Can.	Можно.
	orth@NP	ft-eng@NP	ft-rus@NP
57	Позьб. Кбні ті Сыктывкарын?	Can. Where are you in Syktyvkar?	Можно. Где вы (располагаетесь) в Сыктывкаре?
	orth@MaR	ft-eng@MaR	ft-rus@MaR

Разметка корпуса – это информация о содержании видеофайла: слова, которые говорят люди в кадре, данные о жестах, мимике и движении людей. По разметке в корпусе данных можно найти интересующие исследователя определённые элементы, например найти все случаи «определённого жестикулирования» и определить, с чем это может быть связано, с какой интонацией произношения зафиксировано или какими значениями. Исследования, построенные на основе разметки, позволяют узнать, для чего или почему человек выполняет тот или иной жест или движение в диалоге. Такие исследования позволяют судить о паттернах, стандартных последовательностях, выражениях эмоций у человека ([http://www.mpi.nl/tools/elan/tp/how-to/ELAN\\_handout\\_Russian.pdf](http://www.mpi.nl/tools/elan/tp/how-to/ELAN_handout_Russian.pdf)). Всё это даёт возможность исследователю более полно изучить материал, охватив экстралингвистические факторы, культурную и бытовую ситуацию коммуникации. Не исключено, что цифровые записи диалектного материала можно изучать также и с помощью других электронных исследовательских программ. Например, с помощью программ FONETIKA и ELAN можно провести обширные фоностатистические исследования. Программа FONETIKA разработана в Лаборатории кафедры иностранных языков МарГУ (об этом: Савиных 1996, Пунегова 2002). Методика анализа фоностатистических данных с помощью программы FONETIKA была использована при изучении фонетики финно-угорских языков, в том числе и фонетических особенностей коми языка в трудах Г.В. Пунеговой (Пунегова 1996; 1998; 2002). Известно, что изучение аллофонного варьирования звуков слуховым методом часто сопровождается инструментальными методами исследования. Программа ELAN на основе текстов цифровых записей, преобразованная их в фонетическую транскрипцию, имеет широкие возможности исследовать материал на уровне сегментной и супraseгментной фонетики. Так, с помощью программы можно распознать звуки и их варианты, характерные для исследуемой речи диалекта, а также с помощью звуковой волны, отражающей на экране монитора, выявить особенности интонации, длительности сегментных единиц и их формантные значения. Эта программа, к примеру, может служить и для выявления формантных показателей гласных в ижемском диалекте коми языка и таким образом дополнить ранее уже описанные характеристики аллофонового варьирования гласных, изученных как на экспериментально-фонетическом материале (Терентьев 1970, Игушев 1972, Канева 2003 и др.), так и описательным методом исследования (Лыткин 1970, Сахарова, Сельков 1976, Игушев 1996 и др.).

Какие возможности для исследователя-лингвиста даёт эта программа? Самым большим преимуществом программы является то, что она даёт большие возможности для поиска в корпусе. Задав определённые требова-

ния к поиску, можно находить в корпусе необходимые для исследователя лексемы или морфемы. Например, мы решили посмотреть, как в корпусе используется послелог *сай-*, который может выступать в форме *сайын*, *сайэ*, *сайсянь* и т. д. В поиске по корпусу эти формы можно легко найти, причём программа ELAN даст возможность рассмотреть их в речевом контексте. Более подробно о возможностях поиска можно прочитать на сайте [wiki.fu-lab.ru](http://wiki.fu-lab.ru). Для грамматистов, интересующимся формами того или иного диалекта, можно создать корпус диалектных особенностей грамматических форм, которые могут храниться в архиве данной программы и использоваться учёными в любое время: при описании как отдельных диалектов языка и в сравнительно-типологическом плане.

Также очень важно, что ELAN позволяет работать с текстами в едином пространстве: видео-, аудиозаписи, расшифровки текстов (аннотации), их переводы, разделение аннотаций по участникам коммуникации находятся в одном рабочем поле. Во время расшифровки текста программа даёт возможность подробно прослушать тот или иной фрагмент текста, так как разрешает снижать скорость речи участника коммуникации. Таким образом, речь информанта можно рассматривать, поочерёдно разбивая на наименьшие фонетические единицы – от фразы до сегментной единицы. Это позволяет, к примеру, рассмотреть и установить особенности звучания на уровне фразы, фонетической синтагмы, фонетического слова, слога и звука.

Как уже отмечалось выше, ELAN позволяет составлять несколько аннотаций: расшифровку речи, его фонетическую транскрипцию, перевод на другие языки и т.д. Эта возможность интересна исследователю с точки зрения сопоставительно-типологического метода: можно сопоставлять коми языковые элементы (лексемы, реалии, устойчивые выражения и т.д.) с русским, английским языком. Параллельные тексты (переводы) также могут являться объектами исследования по вопросам проблем перевода устных текстов.

Программу ELAN активно используют участники проекта «Коми-ижемцы: создание аннотированного электронного корпуса для дальнейших исследований по коми языковым особенностям российского Крайнего Севера». Проект реализуется с 2014 г. Основной целью проекта является социолингвистическое исследование всех территорий проживания коми-ижемцев, сравнительное исследование локальных говоров ижемского диалекта, создание электронного архива аннотированных записей (видео, аудио, расшифрованные тексты на коми языке, в переводе на русский и английский языки) (Блокланд Р., Рийсслер М., Партанен Н. и др.). Этот проект направлен на создание своего рода электронного корпуса диалектных

текстов. В диалектологических исследованиях значимость таких корпусов отмечается многими учёными. Так, Г.А. Некрасова пишет, что «основу ЭБКД должен составить программно обеспеченный электронный текстовый корпус, являющийся наиболее надёжной формой хранения диалектных текстов» (Некрасова, 2010).

Именно программа ELAN является базовой в работе проекта «Коми-ижемцы: создание аннотированного электронного корпуса для дальнейших исследований по коми языковым особенностям российского Крайнего Севера», так как предоставляет широкие возможности для изучения ижемского диалекта коми языка. Во время экспедиций участники проекта записывают видео- и аудиофайлы, затем эти записи обрабатываются программой ELAN – создаются файлы для дальнейшей работы. На сегодняшний день более 50% записанного участниками проекта материала в той или иной степени готово к использованию: записи диалектного полевого материала аннотированы на коми язык с переводом на русский и английский языки.

После завершения транскрибирования и перевода текста на отдельные языки файл сохраняется и становится готовым для использования в исследованиях.

Исследования, которые можно проводить на основе собранного материала, могут иметь не только чисто лингвистический, но и социолингвистический, культурологический характер, так как собранный материал имеет богатую тематику. Тематически можно выделить несколько разделов, указывающих на различные сведения:

- биографические сведения;
- сведения о языковых связях с соседними народами (коми-русские, коми-ненецкие, коми-саамские);
- сведения о языковой ситуации в регионе (информация о статусе языка, о том, обучают ли его в школе);
- сведения о традиционных занятиях коми-ижемцев (оленоводство, охота, рыболовство, скотоводство);
- сведения о традиционной духовной культуре народа (праздники, обряды, традиции, народные песни, частушки);
- сведения о материальной культуре (одежда, традиционное жилище);
- топонимические сведения (названия деревень, сёл, ручьёв, рек, лугов и т.д.);
- исторические сведения (заселение территорий, исторические события).

В заключение хотелось бы отметить, что программа ELAN – это современная программа, которая открывает широкие возможности для систематизации, хранения и исследования диалектологического корпуса собранно-

го в полевых условиях материала, позволяющая обращаться к этим материалам исследователям и анализировать его с различных точек зрения, различными методами и принципами.

\*\*\*

Блокланд Р., Рийссерс М., Партанен Н., Федина М., Чемышев А. Использование цифровых корпусов и компьютерных программ в диалектологических исследованиях: теория и практика // Актуальные проблемы диалектологии языков народов России : материалы XIV Всероссийской научной конференции. – Уфа: ИИЯЛ УНЦ РАН, 2014. – С. 252.

Игушев Е.А. Экспериментальные данные по долготе гласных ижемского диалекта // Советское финно-угроведение. – 1972. – № 1. – С. 55–57.

Игушев Е.А. Коми диалектология. – Сыктывкар, 1996. – 94 л.б.

Канева С.А., Пунегова Г.В. Аллофонное варьирование звука /e/ в ижемском диалекте коми языка (на материале экспериментального анализа фонетически представительного текста) // Символ в системе культуры: проблемы интерпретации текста. – Сыктывкар, 2003. – С. 119–121.

Лыткин В.И. Несколько слов о вокализме ижемского диалекта коми языка // СФУ. – 1970. – № 4. – С. 259–261.

Некрасова Т.А. Электронный диалектный корпус как ресурс сохранения и изучения коми диалектов // Финно-угорский мир. – Саранск, 2010. – № 1 (16). – С. 13–16.

Пунегова Г.В. Фонетическая интерференция в условиях коми-русского двуязычия (на материале экспериментально-фонетического исследования согласных) : дис. ... канд. филол. наук. – Сыктывкар, 2002. – 291 с.

Пунегова Г.В. Фоностатистические характеристики фонем в литературном коми языке // Фонетический фонд финно-угорских языков. – Йошкар-Ола; СПб., 1996. – Вып. 1. – С. 90–112.

Пунегова Г.В. Функциональная нагрузка фонологических единиц в разных стилях литературного коми языка // Фонетический фонд финно-угорских языков. – Йошкар-Ола; СПб., 1999. – Вып. 3. – С. 64–126.

Савиных И.В. Программа FONETIKA для исследования фоностатистических данных финно-угорских языков // Фонетический фонд финно-угорских языков. – Йошкар-Ола, СПб., 1996. – Вып. 1. – С. 175–177.

Сахарова М.А., Сельков Н.Н. Ижемский диалект коми языка. – Сыктывкар, 1976. – 288 с.

Терентьев В.А. О долготах гласных и согласных в ижемском диалекте коми-зырянского языка // Советское финно-угроведение. – 1970. – № 4. – С. 255–257.

[http://www.mpi.nl/tools/elan/tp/how-to/ELAN\\_handout\\_Russian.pdf](http://www.mpi.nl/tools/elan/tp/how-to/ELAN_handout_Russian.pdf) (дата обращения – 08.02.2016).