

ПРОЕКТ СПЕЦИАЛИЗИРОВАННОГО КОРПУСА РУССКОГО ЖЕСТОВОГО ЯЗЫКА ДЛЯ ПРИМЕНЕНИЯ В ЧЕЛОВЕКО-МАШИННЫХ ИНТЕРФЕЙСАХ¹

*Ильдар Амирович Кагиров
(Санкт-Петербургский институт информатики
и автоматизации Российской академии наук)*

Введение. Настоящая статья посвящена описанию некоторых аспектов базы данных русского жестового языка (РЖЯ), работа над которой ведется в лаборатории речевых и многомодальных интерфейсов СПИИРАН с 2018 года.

Задача создания подобной базы данных появилась в ходе создания многомодального интерфейса на основе жестов и речи для управления ассистивным мобильным информационным роботом. В качестве роботизированной платформы была выбрана роботележка для супермаркета. Общая концепция подобного устройства подразумевает двустороннюю коммуникацию с покупателями (понимание запросов, навигация) посредством речевого и жестового каналов. Канал ввода информации поддерживается контроллером Microsoft Kinect v2.0, канал вывода представлен 3D-аватаром.

Основным предметом данной статьи является система транскрибирования и «хранения» единиц РЖЯ в электронной базе данных, позволяющая компьютеру (и, как следствие, пользователю) эффективно работать с ней.

РЖЯ: существующие описания и базы данных. К сожалению, исследования по РЖЯ довольно немногочисленны, и научное лингвистическое исследование РЖЯ началось, по большому счету, совсем недавно. Не существует комплексного описания грамматики и лексики, равно как и полноценного корпуса РЖЯ. Данные по РЖЯ редко включаются в типологически ориентированные работы по жестовым языкам в силу малодоступности первых [1], [2].

На сегодня самым крупным (и единственным) корпусом² РЖЯ является «Корпус русского жестового языка», создаваемый Новосибирским государственным университетом [3]. Корпус содержит

¹ Данные исследования проводятся при поддержке Минобрнауки России в рамках Соглашения № 14.616.21.0095 (идентификатор RFMEFI61618X0095).

² Под корпусом в настоящей статье понимается собрание текстов, снабженное лингвистической разметкой, или аннотацией (помимо некоторых других характеристик, см. [9, 10]).

свыше 230 текстов от 43 носителей РЖЯ и снабжен переводом и аннотацией в среде ELAN [4]. Возможен поиск по корпусу — как по лексемам, так и по грамматическим показателям. Также в корпусе содержатся сведения об информантах, присутствует и сортировка текстов по метаданным. В целом корпус представляет собой эффективный инструмент для изучения РЖЯ. Стоит отметить, однако, что большая часть информантов является носителями сибирского и московского диалектов РЖЯ (Новосибирск, Москва), что снижает ценность этого корпуса для исследований других диалектов, в частности, петербургского.

Помимо онлайн-корпуса РЖЯ, существует ряд электронных видеословарей и учебных курсов, в которых реализована возможность поиска по лексемам. К сожалению, объем настоящей статьи не позволяет остановиться на них подробно, поэтому мы ограничимся простым перечислением: [5: 128; 6–8].

Предполагается создание корпуса со следующими характеристиками:

- Двухязычный, параллельный: тексты на РЖЯ снабжены переводом на русский язык.
- Устный: в силу специфики функционирования текстов на жестовых языках.
- Жанровая принадлежность текстов: литературная и разговорная.
- Полнотекстовый.
- Размеченный: метатекстовые, морфологические, синтаксические, экстралингвистические данные.
- Объем корпуса и, как следствие, его репрезентативность на настоящий момент не могут быть четко обозначены.

Специфика задачи, вкратце представленной во Введении, подразумевает, что заявленный в настоящей работе корпус должен быть пригоден для применения в человеко-машинных интерфейсах, в частности, в нем должны присутствовать элементы разметки, позволяющие осуществлять эффективное машинное обучение и распознавание жеста. Таким образом, текущей задачей создателей корпуса является разработка концепции разметки и структурирования данных. Текущей задачей, решение которой прямо влияет на объем корпуса, является разработка концепции разметки и хранения данных.

Словник. В качестве «пилотного исследования» был записан словник по тематике «товары из супермаркета». Словник включает 164 отдельных слова, словосочетания и предложения. В качестве информантов были привлечены студенты и преподаватели МЦР. Запись проводилась при помощи сенсора Microsoft Kinect 2.0, стимулы выводились

на экран компьютера и демонстрировались информантами по 5 раз минимум в целях дальнейшего компьютерного обучения. Большая часть дикторов¹ не владеет литературной нормой РЖЯ, поэтому все переводы были предварительно стандартизированы специалистами из МЦР.

Наполнение словника производилось путем отбора наиболее частотных (на взгляд создателей) наименований продукции, для которых РЖЯ имеет соответствующие жестовые обозначения. Так, слова вроде «фасоль» и «креветки» были исключены в силу того, что носители прибегают к дактилированию при передаче этих понятий.

Кроме того, в словник вошли фразы на РЖЯ, помогающие пользователю управлять роботизированной тележкой посредством жестового или голосового ввода. Это наиболее «естественные команды», связанные с запросом местонахождения продукции, передвижения по залу и т. п.

Аннотирование жестов. Основной задачей при обработке словника оказалась задача транскрибирования жестов. Транскрибирование представляется необходимым, во-первых, для удобства работы с корпусом, во-вторых, для выделения основных смыслообразительных признаков каждого жеста.

Вслед за работой [11] жест принято рассматривать не как цельную, неделимую сущность, но как совокупность трех элементов:

- форма и ориентация руки;
- место исполнения жеста;
- характер движения.

Зачастую ориентация руки считается отдельным элементом. Таким образом, «классический» подход к описанию жеста оперирует набором из трех (или четырех) характеристик, оказывающихся минимальными дифференциальными признаками жеста. Набор реализаций этих признаков конечен (в [12] говорится о 45 формах руки, 25 локализациях и 12 типах выполнения жеста для американского жестового языка). Такая трактовка жеста позволяет говорить о фонологии применительно к жестовым языкам мира. Несмотря на то, что в термине «фонология» заложено указание на звук (др.-греч. *phōnē* — «звук»), этот термин все же удачен в силу функционального сходства фонем звучащих языков и описанных выше дифференциальных признаков жеста. Впрочем, в ранних работах по жестовым языкам использовался термин «хиология» (др.-греч. *kheir* — «рука»), а вместо термина «фонема» — термин «хирема» [11].

¹ В англоязычной литературе используется специальный термин *signer*, для которого нет точного перевода на русский язык. Поэтому в настоящей статье информанты, владеющие РЖЯ, названы дикторами. В [18] предложен термин «говорящий».

В настоящее время «классическая» модель Стоуки, подразумевающая декомпозицию фонемы на отдельные синхронно исполняемые элементы, была расширена за счет введения еще одного параметра жеста — линейности [13–15]. Так, в модели Р. Джонсона и С. Лиделла [15] жест представлен как последовательность сегментов движения (*movement*) и остановок (*hold*). В модели В. Сэндлер [16] конфигурация рассматривается как отдельный автосегментный параметр, т. е. принадлежащий самостоятельному уровню и ставящийся в соответствие сегментным единицам.

В настоящее время существует несколько систем фиксации жеста [5: 129–135]. Самой подробной и разработанной (но не самой популярной ввиду ее громоздкости) из них является HamNoSys [17], позволяющая очень точно описать и, в случае с 3D-аватарами, представить любой жест. HamNoSys, как и прочие системы транскрипции жестовых языков, основывается на классической модели Стоуки, оперируя в конечном итоге формой руки, локацией и характером движения. Кроме того, HamNoSys представляет «фонетическую», а не «фонематическую» транскрипцию жеста, не давая представления об основных смысловоразличительных признаках. Это создает определенные неудобства для автоматической обработки и классификации единиц РЖЯ. Так, запись одного и того же жеста в представлении двух разных информантов (МОЛОКО, Рис. 1) выглядит по-разному:



Рис. 1. Различия в произношении у разных дикторов

При одинаковом характере движения (обе руки по очереди поднимаются и опускаются с характерным изменением формы кисти) реализации жеста у двух разных информантов отличаются в таких деталях, как локализация рук, ориентация кисти и даже форма кисти (положение большого пальца). Система HamNoSys предполагает фиксацию всех особенностей, что делает запись и громоздкой, и вводящей в заблуждение рядом «внешних» различий.

Представляется, что для аннотации леммы слова в корпусе требуется не фиксация «абсолютных» характеристик руки информанта, а фиксация основных дифференциальных признаков, формирующих данный жест. В связи с этим удобной будет декомпозиция жеста не в соответствии с «классической» моделью Стоуки, а представление его как последовательности движений и остановок.

В самом деле, демонстрация жестов информантами зачастую разнится в деталях: немного отличаются конфигурация кисти, локализация, однако характер выполнения жеста, т. е. движение чаще всего остается неизменным. Поэтому эффективная нотация, позволяющая осуществлять поиск по корпусу и распознавание жеста компьютером, должна учитывать не все гипотетически возможные характеристики жеста, а те и только те, что позволяют отличить один жест от другого.

Для заявленной базы данных такими признаками оказываются движение, форма кисти и — для некоторых жестов — локализация. В Таблице 1, представляющей собой фрагмент таблицы одноручных жестов, верхняя строка показывает типы движения руки (используются символы из нотации HamNoSys), левая колонка — типы конфигурации кисти. Цифрами даются ссылки на соответствующий жест из словника, там же указывается локализация жеста.

Таблица 1

Пример классификации элементов русского жестового языка

	↘↙↔	?+	↔+	↔+	↘↙↔	↓
↘	1					10
↘ ²		3,5				
↘				6		
↘ ³			7			
↘ ⁵					8	
↘	9					

Не исключено, что при дальнейшем расширении списка слов некоторые детали предложенной системы классификации будут меняться.

Заключение. Данная работа является лишь очерком, пилотным исследованием проблемы, и представляет собой постановку задачи. Подытожим ее основные пункты:

1. Выбор системы транскрибирования жестов. Использование подробных, но громоздких систем нотации представляется неоправданным в силу их избыточной информативности и плохого соответствия принципам автоматической обработки жеста.

2. Принцип классификации жестов. Как было показано выше, в основу классификации был положен признак движения, в качестве дополнительных признаков используются форма кисти и локализация. Возможно, с расширением объема корпуса этот принцип будет несколько видоизменен.

Список использованной литературы

1. The Study of Signed Languages: Essays in Honor of William C. Stokoe. Washington, D.C.: Gallaudet University Press, 2002.
2. Sign Languages. Cambridge: Cambridge University Press, 2010.
3. Корпус русского жестового языка // [Электронный ресурс] — Электронные данные — Режим доступа: URL: <http://rsl.nstu.ru/> свободный.
4. The Language Archive // [Электронный ресурс] — Электронные данные — Режим доступа: URL: <https://tla.mpi.nl/forums/software/elan/> свободный.
5. Карпов А. А., Кагиров И. А. Формализация лексикона системы компьютерного синтеза языка жестов // Тр. СПИИРАН. 2011. Вып. 1 (16). С. 123–140.
6. Воскресенский А. Л. Сопоставительное лексикографическое описание слов русского языка и жестов языка глухих в России в словаре RuSLED // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог». 2008. № 7(14). С. 91–96.
7. Spread the Sign // [Электронный ресурс] — Электронные данные — Режим доступа: URL: <https://www.spreadthesign.com/ru.ru/search/> свободный.
8. Проект «Сурдосервер 2.0» // [Электронный ресурс] — Электронные данные — Режим доступа: URL: <http://surdoserver.ru> свободный.
9. McEnery T, Wilson A. Corpus linguistics. Edinburgh: Edinburgh University Press, 1996.
10. Кононев М. В. Введение в корпусную лингвистику: учебное пособие для студентов лингвистических и филологических специальностей университетов. Прага: Animedia Company, 2014.
11. Stokoe W. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf // Studies in Linguistics: Occasional papers. 1960. No. 8. Buffalo: Dept. of Anthropology and Linguistics, University of Buffalo.

12. *Battison R.* Lexical borrowing in American Sign Language. Silver Spring, MD: Linstok Press; 1978.
13. *Newkirk D.* On the temporal segmentation of movement in American Sign Language. La Jolla. 1981.
14. *Liddell S.* Think and Believe: Sequentiality in American Sign Language // Language. 1984. № 60. С. 372–399.
15. *Liddell S., Johnson R.* American Sign Language: The phonological base // Sign Language Studies. 1989. № 64. P. 195–277.
16. *Sandler W.* The spreading hand autosegment of American Sign Language // Sign Language Studies. 1986. № 50. P. 1–28.
17. DGS-Korpus // Электронный ресурс [<https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys-97.html>].
18. *Буркова С. И., Филимонова Е. В.* Редупликация в русском жестовом языке // Русский язык в научном освещении. 2014. № 2. С. 202–258.