СОПОСТАВИТЕЛЬНОЕ ЛЕКСИКОГРАФИЧЕСКОЕ ОПИСАНИЕ СЛОВ РУССКОГО ЯЗЫКА И ЖЕСТОВ ЯЗЫКА ГЛУХИХ РОССИИ В СЛОВАРЕ RUSLED

COMPARATIVE LEXICOGRAPHIC DESCRIPTION OF RUSSIAN WORDS AND GESTURES OF RUSSIAN SIGN LANGUAGE IN RUSLED DICTIONARY

Воскресенский А.Л. *(AVoskresenskij@college.mesi.ru)*
Колледж МЭСИ, Москва

Представляется лексикографическое описание жестов глухих России в сопоставлении со словами русского языка, имеющими те же или близкие значения, в мультимедийном толковом словаре RuSLED, предназначенном для изучения особенностей употребления жестов и слов русского языка.

Введение

Двуязычный словарь является одним из мест встречи двух культур, языки которых представлены в словаре. Он должен не только дать справку об употреблении речевых единиц (в данном случае слов и жестов) соответствующих языков, но и представить семантическую структуру отдельных речевых единиц с учётом оттенков и переходов в их значениях и употреблении. При этом справочная информация словаря должна быть понятна пользователю с учетом существующих культурных различий.

Различия между социальными группами слышащих и глухих достаточно велики [1]. Учитывая это, а также то, что между словами русского языка и жестами языка глухих России во многих случаях нет однозначного соответствия [2], нужно признать, что существующие словари жестового языка (см., например, [3, 4]) не дают достаточно ясного (для человека, не владеющего языком жестов) описания особенностей применения жестов. Кроме того, для многих глухих было бы полезно иметь толковый словарь русского языка, в котором слова дополнительно пояснялись бы жестами.

Нужно признать, что составители словаря RuSLED не являются специалистами в лексикографии, и данный словарь является их первым опытом в данной области. Но это имеет и положительные черты. Например, составители словарей жестов обычно дают варианты жестов без пояснений, являются ли эти варианты диалектными формами или несут различные оттенки значений. Для них это очевидно, поэтому необходимые для новичка в данной области пояснения опускаются. С другой стороны, при описании слов русского языка, в связи с многообразием предметных областей, в которых фигурируют слова-омонимы, даже опытные специалисты иногда ошибаются [5], особенно в случаях, когда решение кажется очевидным. В этом случае действует механизм, сходный с «ложными друзьями переводчика» при переводе с одного языка на другой [6].

Для разработчиков словаря RuSLED все в данной области деятельности является новым, поэтому даже в очевидных для специалистов случаях у них возникают «детские» вопросы *«почему?», «чем это объясняется?»* и т.п. Может быть именно поэтому словарь RuSLED не имеет известных аналогов, в которых такое внимание уделялось бы толкованиям слов и жестов. При всех своих недостатках словарь вызвал интерес у представителей общины глухих Москвы и даже такие высокие оценки, как «это то, что нужно».

Описание словаря RuSLED

Словарь русского жестового языка включает в себя функции толкового словаря, как для введенного слова, так и для его жестового представления. На вход словаря подается произвольная форма слова, а на выходе демонстрируются варианты жестового толкования данного слова.

Для нормализации входной словоформы (получения лексемы) разработан морфологический анализатор, в основу которого положен морфологический анализатор [7], использованный ранее в разработках TULIPS, TULIPS-2 и др. В качестве основного источника допустимых словоизменений (также как в [7] и в большинстве отечественных систем морфологического анализа) используется словарь А.А. Зализняка [8].

Морфологический анализатор включен в словарь как элемент будущей системы перевода текста в жесты.

Отличием разрабатываемого словаря является то, что для каждого семантического значения лексемы (и жеста) используется отдельный вход словаря – отдельная запись в таблице базы данных. Это значительно удобнее для пользователя, является очевидным решением для электронных толковых словарей и рекомендуется лексикографами [9].

В соответствии с этим изменены по сравнению с [7] как структура базы данных, так и запросы на выборку данных, реализованные на языке SQL. Использование SQL облегчит предполагаемую в дальнейшем реализацию мультимедийного толкового словаря русского жестового языка в виде Web-сервиса.

Форма доступа к данным демонстрационной версии словаря, выполненной на СУБД MS Access, приведена на рис. 1.

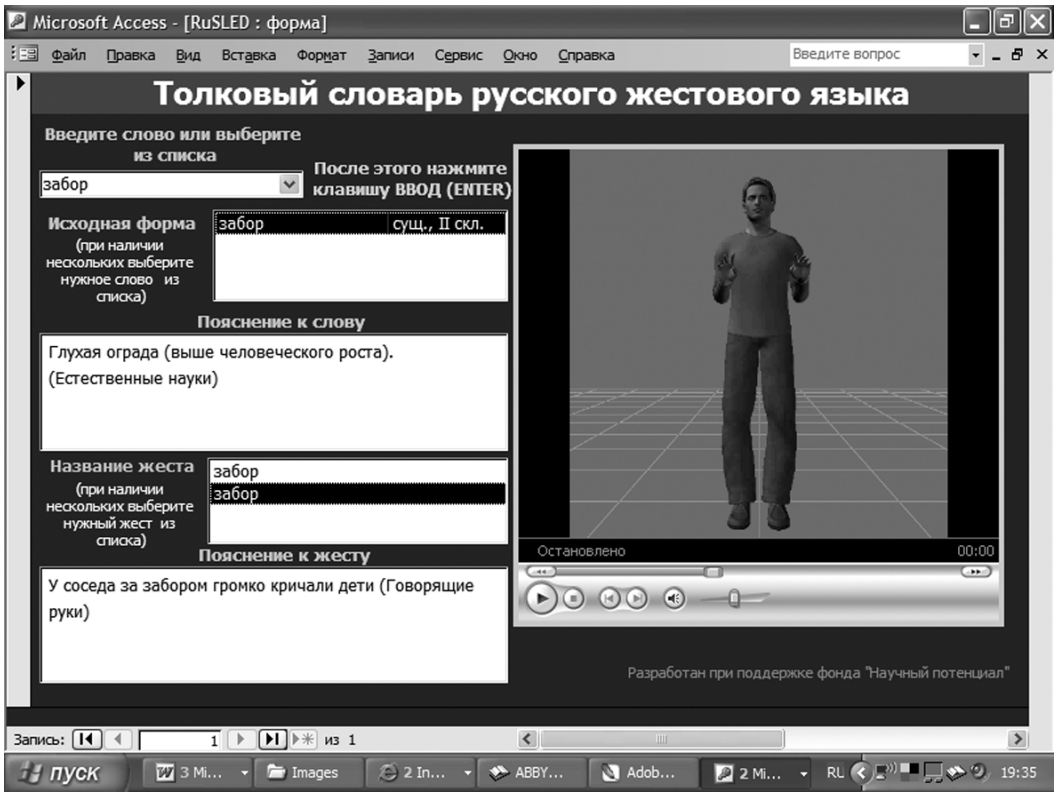

Рис. 1. Словарь RuSLED. Форма доступа к данным.

Поле «Введите слово» позволяет вводить произвольные словоформы или выбирать из списка лексемы, имеющиеся в словаре. В список «Исходная форма» выводится соответствующее основе значение лексемы или несколько значений, если по результатам морфологического анализа выбрано несколько записей, рис. 2, на котором приведен случай ввода пользователем словоформы «бора». Она соответствует именительному падежу слова «бора», обозначающего сильный ветер в приморских районах, где невысокие горы подступают непосредственно к побережью, а также родительному падежу слова «бор», имеющему несколько значений: еловый или сосно-

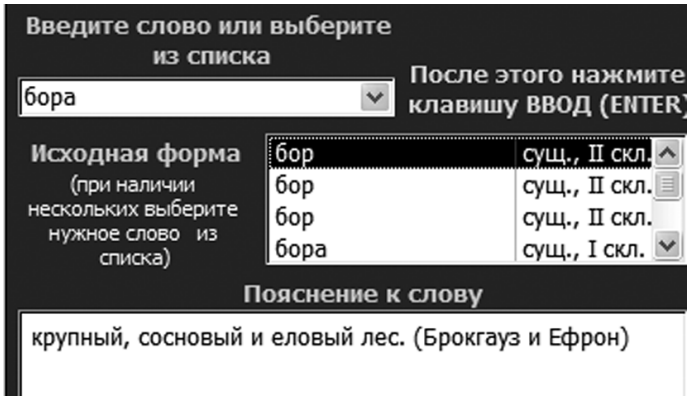

Рис. 2.

вый лес, химический элемент, инструмент стоматолога и т.д. При выборе пользователем элемента списка «Исходная форма» в поле «Пояснение к слову» выводится толкование выбранной лексемы.

На рис. 3, 4 приведен случай ввода пользователем словоформы «стекла». Морфологический анализатор в поле «Исходная форма» выводит существительное «стекло» и глагол «стечь». Пример показывает, что морфологический анализатор словаря учитывает чередование согласных в различных словоформах лексемы. Учитываются пропуски букв («беглые гласные»), добавление возвратных частиц («-ся», «-сь»).

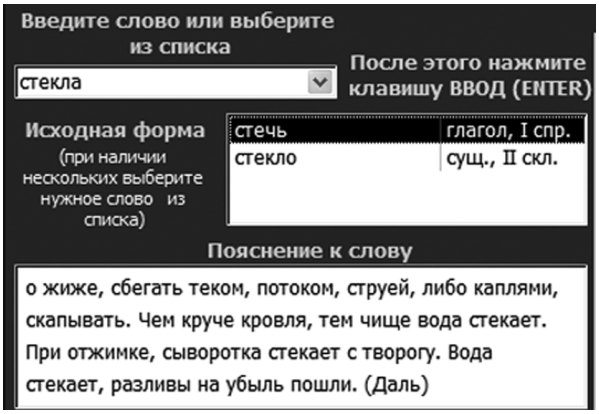

Рис. 3.

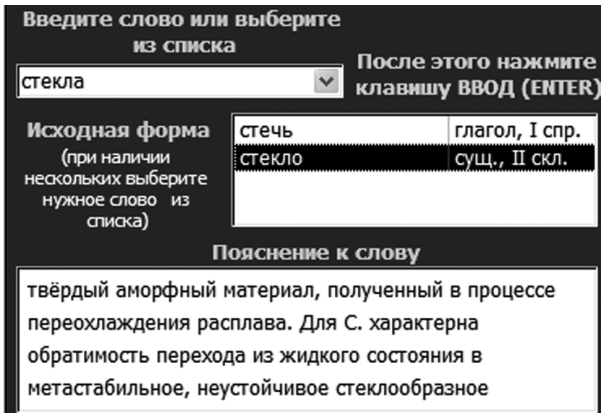

Рис. 4.

При выборе пользователем нужной лексемы в поле «Название жеста» выводится наименование жеста (как правило, совпадающее с лексемой) или (если данной лексеме соответствуют несколько жестов) список наименований (рис. 1). При выборе пользователем нужного жеста его изображение выводится в окне плеера, а в поле «Пояснение к жесту» выводится поясняющий текст. В ходе разработки осуществлена оцифровка [4], что позволяет использовать фрагменты этого курса в словаре RuSLED (рис. 4, 5). Постепенно осуществляется замена видеофрагментов анимированными изображениями с целью перехода к компоновке жестовых высказываний из комбинаций жестов с использованием единого демонстратора жестов – виртуального персонажа.

На рис. 5 и 6 представлены результаты выбора пользователем различных значений слова «лук». Для каждого из значений этого слова выдается только то значение жеста, семантика которого отвечает выбранной лексеме. Подробнее этот механизм будет рассмотрен ниже.

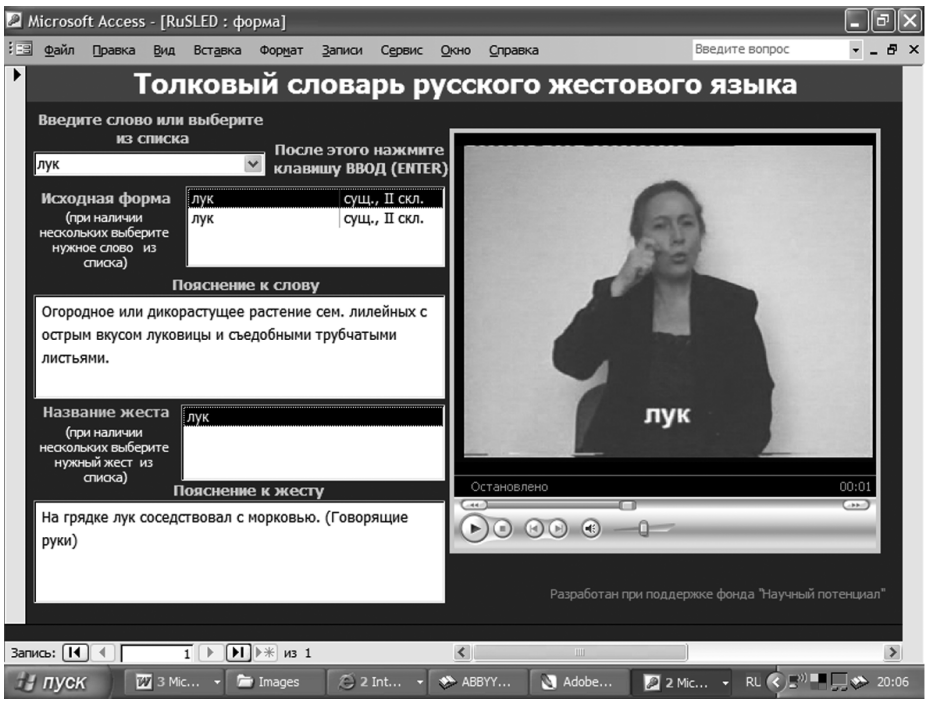

Рис. 5. Жест «лук» (растение).

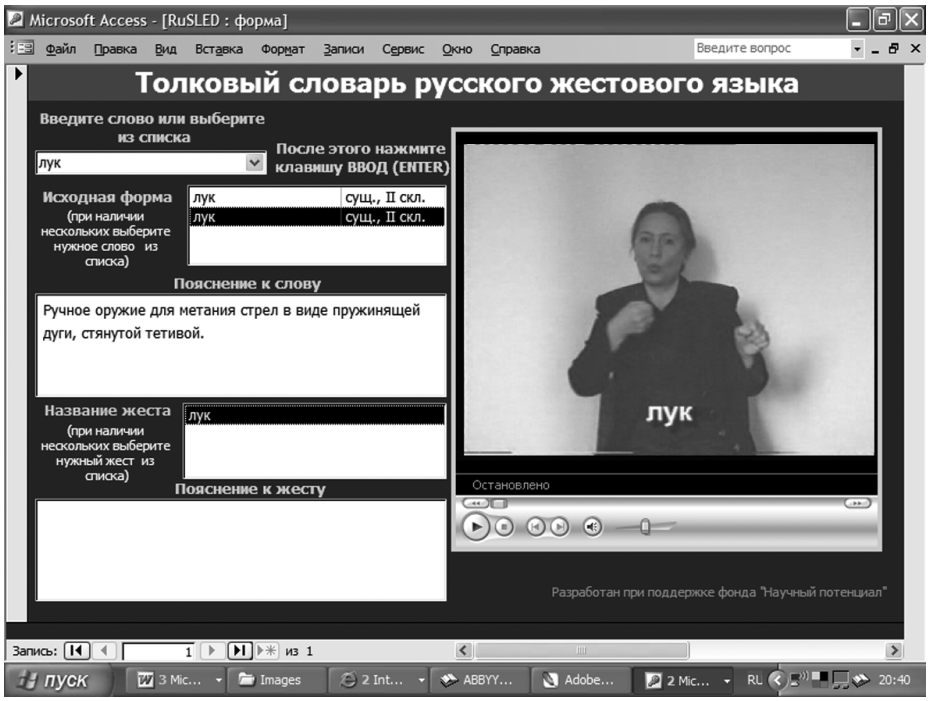

Рис. 6. Жест «лук» (оружие).

Источник словника словаря RuSLED

Одной из важнейших (и сложных) задач является определение размера и состава словника словаря. Поскольку предполагается, что данный словарь должен служить пособием при изучении русского жестового языка, словник (и *жестовник* ?) должны обеспечивать хотя бы минимальные требования коммуникации.

Первоначально словарь предполагалось создавать на основе набора жестов, представленных в [4], добавив к ним дополняющие жесты из [3]. Но, при демонстрации прототипа словаря представителям общины глухих Москвы, оказалось, что многие жесты, представленные в [4], не совпадают с жестами, используемыми в Москве для обозначения тех же понятий. При этом была высказана просьба не включать в словарь варианты жестов, чтобы «не было путаницы».

В настоящее время словарь формируется на основе наименований жестов, представленных в [4], но выполненных в манере, принятой в Москве. При этом создаваемый словарь не является нормативным, скорее это дескриптивный словарь, описывающий московский диалект русского жестового языка. Но не следует забывать, что литературный русский язык возник на основе московского диалекта…

Учитывая, что большинство предполагаемых будущих пользователей словаря – молодые люди, школьники, дальнейшее пополнение словаря должно вестись на основе современной лексики. В качестве источника этой лексики избран сайт http://youngwriters.ucoz.ru/, на котором представлены литературные произведения школьников и студентов младших курсов институтов, объединенных молодежным пресс-центром «Метаморфозы» московского района Строгино.

Содержание поля «Пояснение к слову»

Пояснения к словам должны передавать смысловые оттенки слов и особенности их употребления для лучшего усвоения норм русского языка глухими, для которых русский язык во многих случаях может считаться вторым языком [2]. Показанные на рис. 3 и 4 примеры, заимствованные, соответственно, из словаря В. Даля и Большой советской энциклопедии, представленных в разделе «Словари» портала Яндекс (www.yandex.ru), свидетельствуют, что тексты пояснений должны тщательно редактироваться, чтобы быть понятным пользователям, чей активный словарь достаточно ограничен. В то же время эти пояснения должны, по возможности, передавать существенную информацию, чтобы способствовать развитию кругозора пользователей словаря. Наиболее эффективно решение этих противоречивых задач может быть осуществлено при участии педагогов школ глухих и слабослышащих и психологов, специализирующихся на работе с глухими.

Содержание поля «Пояснение к жесту»

В настоящее время поясняющие тексты к жестам взяты из словаря [3]. Но примеры, показанные на рис. 1 и 5, свидетельствуют, что эти пояснения (во многих случаях отсутствующие) слишком кратки и не выполняют возлагаемой на них роли. В связи с этим заполнение этого поля словарных статей, описывающих жесты, ведется с привлечением носителей русского жестового языка. При этом особое внимание уделяется случаям, когда одному слову русского языка соответствуют несколько жестов, передающих различные оттенки смысла. Например, слову «*этаж*» соответствуют различные жесты «*этаж*», соответствующие в одном случае описанию одного из этажей многоэтажного дома, в другом – указывающие, что речь идет о многоэтажном доме.

Использование толковых статей словаря для поиска семантических связей

Семантическая связь лексем и жестов осуществляется путем включения в записи таблиц Words и Signs полей, названных Sem. В настоящее время данные в эти поля заносятся вручную. Выборка нужного жеста (ссылка на соответствующий мультимедийный файл) производится из записи таблицы Signs, имеющей совпадающие значения полей Name и Sem со значениями полей Lexema и Sem таблицы Words, соответственно. В дальнейшем, при накоплении лексикографических данных, заполнение полей Sem планируется автоматизировать, при этом предполагается использовать методы дискриминантного анализа для выделения слов и словосочетаний, наиболее значимых в данной предметной области для определения границ смысловых полей.

Так, например, для приведенных на рис. 5 – 6 примерах, можно считать, что для лексемы «лук (растение)» такими словами будут *«расти», «сорвать», «срезать», «посадить», «острый»*. Эти термины относятся к таксономическому классу «*растение*». Для лексемы «*лук (оружие)*» такими словами будут *«натянуть», «взять», «согнуть», «тугой»*. Эти термины относятся к таксономическому классу «*инструмент*». Конечно, эти примеры являются упрощенными. Для хранения фраз, словосочетаний и отдельных слов, наиболее свойственных контекстам, в которых проявляются семантические значения слов и наименований жестов, в словаре имеются скрытые поля, недоступные пользователю.

Поиск близких по значению записей таблиц Words и Signs предполагается вести с помощью двух процедур. Первая заключается в поиске в полях, содержащих примеры контекстов, одних и тех же слов, однозначно определяющих соответствующий таксон (это может быть и одно слово, встречающееся в соответствующем контексте единично, но с вероятностью, весьма близкой к единице). Если эта процедура не приводит к успеху, то определяются записи таблиц Words и Signs, контекстные поля которых содержат, соответственно, наибольшее число повторяющихся слов и словарных групп.

При использовании одних и тех же тезаурусов для анализа поясняющих текстов и типовых контекстов лексем и наименований жестов можно использовать слова-вершины деревьев отношений «гипероним – гипоним» в качестве наименований таксономических классов.

Дополнение словаря средствами поддержания отношений «часть – целое», «гипоним – гипероним», «синоним – антоним» как для лексем и наименований жестов, так и для словоформ поясняющих статей (т.е. включение в него тезауруса, а позднее развитие этого тезауруса до уровня онтологии) позволит автоматизировать отнесение лексем и наименований жестов к соответствующим таксономическим классам. Соответственно, будет автоматизировано и заполнение полей Sem. Этот подход, как мы надеемся, позволит автоматически тематически связать огромное множество слов со значительно меньшим множеством жестов, используя для этого существующие тезаурусы русского языка, в том числе и русскую версию WordNet.

Заключение

Качественное лексикографическое описание лексем и жестов является весьма критичным для развития создаваемого словаря. Однако, даже из приведенных примеров (рис. 1 – 6) видно, что имеющиеся в настоящее время в словаре тексты пояснений не могут быть признаны удовлетворительными.

Пояснения к словам или слишком кратки (не включают описаний особенностей применения слова) или (при заимствовании из энциклопедий) слишком подробны, но при этом во многих случаях не понятны для лиц, плохо знакомых с русским языком. Кроме того, как выяснилось в ходе работы, для ряда слов (в основном глаголов) отсутствуют (или не найдены) толкования. В ряде случаев толкования даются путем сопоставления с синонимами.

Пояснения к жестам, приведенные в [3], или очень кратки, или вместо пояснения приводится фраза «жесты отличаются по смыслу», при этом не указывается, в чем заключаются различия, каковы особенности применения того или иного жеста. В [4] толкования жестов, которые можно извлечь из примеров применения жесто-

вых фраз, весьма ограничены, для многих жестов отсутствуют. Из-за различия в содержании словарей [3] и [4] во многих случаях поле «Пояснение к жесту» в настоящее время не заполнено (см. рис. 6).

Дальнейшим развитием словаря является включение в него тезаурусных отношений для слов и наименований жестов с целью автоматизации разделения смысловых оттенков. Этот шаг необходим для перехода в дальнейшем к обработке не только отдельных слов, но также и фраз или фрагментов текстов, т.е. для перехода к системе автоматизированного перевода текста в жесты.

Список литературы

1. Базоев В.З., Паленный В.А. Человек из мира тишины // М.: Академкнига, 2002.
2. Зайцева Г.Л. Дактилология. Жестовая речь: Учебное пособие для ВУЗов // М.: Просвещение, 1991.
3. Фрадкина Р.Н. Говорящие руки: Тематический словарь жестового языка глухих России // М., 2001.
4. Специфические средства общения глухих // СПб – Павловск: МЦР, 2002. Видеокурс: В 3 частях.
5. Воскресенский А.Л., Хахалин Г.К. Средства семантического поиска // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» (Бекасово, 31 мая – 4 июня 2006 г.). М.: Изд-во РГГУ, 2006. С. 100 – 104.
6. Акуленко В.В. О «Ложных друзьях переводчика» // Электронный документ: http://linguistic.ru/index.php?id=79&op=content
7. Мальковский М.Г. Диалог с системой искусственного интеллекта // М.: МГУ, 1985.
8. Зализняк А.А. Грамматический словарь русского языка // М.: Русский язык, 1980.
9. Селегей В.П. Электронные словари и компьютерная лексикография // AINEWS. Новости искусственного интеллекта. 2001. № 1 (49). Электронный документ: http://www.lingvoda.ru/transforum/articles/pdf/selegey_a1.pdf.