

От звучащей речи — к жестовой

А.Л. Воскресенский

Г.К. Хахалин

В статье даётся предварительное описание подхода к созданию системы автоматизированного сурдоперевода. Приводится обоснование необходимости в создании такой системы, описание некоторых её особенностей, возможных путей разрешения проблем.

Введение

Необходимость в создании систем автоматизированного сурдоперевода диктуется не только требованиями мирового сообщества по обеспечению равных прав для всех [1]. Не все глухие в достаточной степени понимают текст сообщений на информационных табло с выводом текста «бегущей строкой», что связано с меньшим (по сравнению со слышащими) объёмом их активного словаря.

В данной работе представляются как результаты исследований, проводившихся в течение ряда лет, так и описание подхода к решению возникающих в ходе работы задач. В отличие от предшествующих публикаций (например, [2, 3]), основное внимание уделяется проблемам поиска конкретных значений понятий, возникающих при переводе звучащей речи в жесты. При этом используются примеры, зафиксированные при разработке толкового словаря русского жестового языка RuSLED [4, 5].

Основные сложности при переводе текста (который может быть результатом работы подсистемы распознавания звучащей речи) связаны с разрешением омонимии, нахождением необходимого значения полисемичного слова, а также с преобразованием фраз русского языка, имеющих свободный порядок слов, в жестовые выражения, в которых порядок жестов значительно более строг.

По своим функциям и характеристикам система перевода текста в жесты может быть отнесена к системам искусственного интеллекта [6], при этом решаемые задачи поиска требуемого значения слова или совокупности слов в некоторых случаях сложнее, чем при переводе с одного словесного языка на другой.

1. Краткое описание словаря RuSLED

Словарь русского жестового языка RuSLED (Russian Sign Language Explanatory Dictionary) включает в себя функции толкового словаря как для введённого слова, так и для его жестового представления. На вход словаря подаётся произвольная форма слова, а на выходе демонстрируются варианты жестового толкования данного слова.

Словарь содержит 2372 слова (с толкованиями их значений) и 2537 видеоизображений жестов (включая различные варианты исполнения), передающих значения этих слов. Для 1592 жестов (63% от общего числа, вошедших в словарь) даны дополнительные пояснения, относящиеся к манере исполнения жеста или описывающие смысловые нюансы, передаваемые жестом.

В словаре представлены жесты, используемые в Санкт-Петербурге и его окрестностях. Частично представленные в словаре жесты совпадают с жестами, используемыми в Москве, но в целом расхождения достаточно велики, что дало повод назвать данный словарь «Петербургский диалект».

Видеоряд словаря составлен на основе видеокурса, изданного Межрегиональным центром реабилитации (МЦР), г. Павловск [7]. В данной версии словаря для демонстрации жестов используются оцифрованные фрагменты видеозаписи сурдопереводчиков, заимствованные из видеокурса. Использование для просмотра жеста элемента ActiveX Windows Media Player позволяет:

- просмотреть этот же жест повторно при нажатии кнопки плеера },
- приостановить выполнение жеста в требуемом месте при нажатии кнопки плеера ||,
- просмотреть любую фазу выполнения жеста, передвинув мышью движок плеера в соответствующую позицию (рис. 1).



Рис. 1. Экранная форма словаря RuSLED

Поставленная ранее цель — использование для демонстрации жестов виртуального персонажа (аватара) — пока не достигнута из-за сложности представления мимики, сопровождающей жесты и выполняющей весьма важную роль в жестовом языке глухих. Так, например, слова *милый*, *симпатичный* передаются одним жестом, но отличаются движениями губ, проговаривающих фрагменты соответствующих слов.

При составлении пояснений к некоторым жестам использовались пояснения из словаря «Говорящие руки» Фрадкиной [8], составленного на основе московского варианта жестового языка.

При составлении пояснений к словам использованы более 30 словарей и энциклопедий, доступ к которым осуществлялся через Интернет, с использованием, по большей части, службы «Словари» портала Яндекс, за исключе-

нием нескольких словарей — в частности, одной из версий Толкового словаря русского языка Ушакова, размещенной на портале ГРАМОТА.РУ.

По рекомендациям сурдопедагогов, обеспечена возможность фильтрации словника словаря по грамматическим категориям (существительные, глаголы, прилагательные, наречия, предлоги, частицы, числительные, местоимения). Для просмотра всего содержимого словаря нужно выбрать категорию «Все слова».

Программная оболочка словаря зарегистрирована Госкоорцентром информационных технологий (ОФАП Минообразования и науки РФ) №10727 от 30.05.2008.

Дистрибутив словаря на DVD выполнен и распространяется ООО НПП «Дериа Графикас» (г. Санкт-Петербург).

Отличием 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].

Поле «Введите слово» позволяет вводить произвольные словоформы или выбирать из списка лексемы, имеющиеся в словаре. В список «Исходная форма слова» выводится соответствующее основе значение лексемы или несколько значений, если по результатам морфологического анализа выбрано несколько записей.

При выборе пользователем нужной лексемы в поле «Наименование жеста» выводится наименование жеста (как правило, совпадающее с лексемой) или (если данной лексеме соответствуют несколько жестов) список наименований жестов. Для каждого из значений слова выдаётся только то значение жеста, семантика которого соответствует значению выбранного из списка слова [4].

2. Примеры неоднозначности слов и соответствующие процедуры обработки контекста

Поскольку между жестами и словами нет однозначного соответствия, при переводе текста в жесты необходимо не только разрешать проблемы омонимии (которые в ряде случаев могут быть сняты лингвистическими средствами путём анализа морфологических форм слов и синтаксиса фраз, в которых они встречаются), но и осуществлять тщательный отбор нужного значения полисемичного слова из соответствующего ряда синонимов.

Наблюдения сурдопедагогов [10] показывают, что абстрактно-логический уровень мышления у глухих формируется позднее, чем у слышащих. В результате у глухих преобладает предметно-образный уровень мышления. Поэтому, как показано ниже, в ряде случаев использование синонима вместо точного значения допустимо при переводе с одного словесного языка на другой (слушающий подсознательно подставляет вместо услышанного слова нужное значение), тогда как при переводе на жестовый язык мы должны найти точное значение слова, иначе мы не сможем подставить в формируемое жестовое выражение нужный жест.

Использованные ниже примеры основаны на словах, имеющих в словаре RuSLED.



2.1. Омография некоторых форм слов

В русском языке написания слова *вино* в родительном падеже единственного числа и слова *вина* в именительном падеже совпадают. Эти примеры могут быть продолжены: например, совпадают написания существительного *весть* в именительном падеже множественного числа и родительном падеже единственного числа, а также глагола *вести*.

Здесь для выявления нужного значения слова достаточно использовать синтаксический анализ локального контекста (ближайшего окружения слова, зачастую меньшего, чем предложение в целом), позволяющий выбрать нужную лексему из вариантов, предлагаемых морфологическим анализатором. При этом учитываются согласованность прилагательных и существительных и связность предложения, включающего анализируемые цепочки слов [11].

2.2. Некоторые случаи омонимии

Приведём несколько примеров. Так, словом *лук* в русском языке обозначаются как съедобное растение, так и вид метательного оружия; словом *автомат* обозначаются как вид огнестрельного оружия, так и устройство, работающее по заданной программе.

Здесь для выявления нужного значения подчеркнутого слова уже не достаточно использовать синтаксический анализ локального контекста. Необходимо использовать контекст, выходящий за пределы предложения [12]. При этом необходимо учитывать частотные характеристики встречаемости слов в рассматриваемом контексте [13, 14], не исключая из рассмотрения предлоги [14], которые часто относятся к категории «стоп-слов», не учитываемых при анализе. Таким образом, помимо достаточно обширного словаря и знания грамматики, система обработки текста должна иметь примеры употребления слов, входящих в её словарь, имеющие ссылки на соответствующие семантические классы.

2.3. Полисемия

Слово *земля* в русском языке имеет ряд значений, из которых в словаре RuSLED встречаются значения *планета*, *почва*, *берег*. Рассмотрим последний случай (рис. 2).

Для жеста, передающего значение *берег*, в словаре [8] приводится пояснение: «"Земля!" — закричали матросы». Различные программы-переводчики, доступные в Интернете, дают следующие варианты перевода (примеры а, б, с):

- (а) «Ground!» — *sailors cried (Cognitive Translator, <http://cs.isa.ru:10000/ct/>);*
- (б) «The Earth!» — *sailors have cried (PROMT® Translator, <http://www.translate.ru/>);*
- (с) «Land!» — *cried the sailors (Translator Google®, <http://translate.google.com/>).*

Общаясь с помощью словесной речи, мы каждый раз решаем задачу распознавания информации, передаваемой нам собеседником. При этом происходит

подстановка значений слов, хранящихся в нашей памяти, т.е. воспринятый смысл текста не является точным аналогом слов, составляющих фразы текста. Там, где это возможно, воспринятое содержание фразы внутренне дополняется (и корректируется) в соответствии с общим содержанием текста и имеющимися знаниями об окружающем мире, не вызывая проявляемых внешне затруднений и протеста. Поэтому варианты (а) и (с) могут быть признаны допустимыми для случая словесного языка, а вариант (b) — нет, поскольку «The Earth» означает планету Земля, которую матросы не могут увидеть как цельный объект ни при каких обстоятельствах.

Но отметим, что ни в одном из случаев не получено значение coast (берег), необходимое для задачи сурдоперевода. То есть система сурдоперевода должна самостоятельно решать задачу выбора и подстановки нужных значений слов, исходя из общего содержания текста. Эти значения не всегда, как показывают приведённые примеры (а)–(с), будут очевидными, поэтому такая задача с полным основанием может считаться интеллектуальной.

Поясним ход рассуждений системы в данном случае, приводящих к распознаванию ситуации [15]: матросы находятся на корабле, пребывающем в открытом море → корабль со всех сторон окружён водой → граница воды и суши (земли) называется *берег* → если матросы закричали «Земля!», это означает, что они увидели границу между водой и сушией (землёй), т.е. *берег (coast)* (рис. 3).

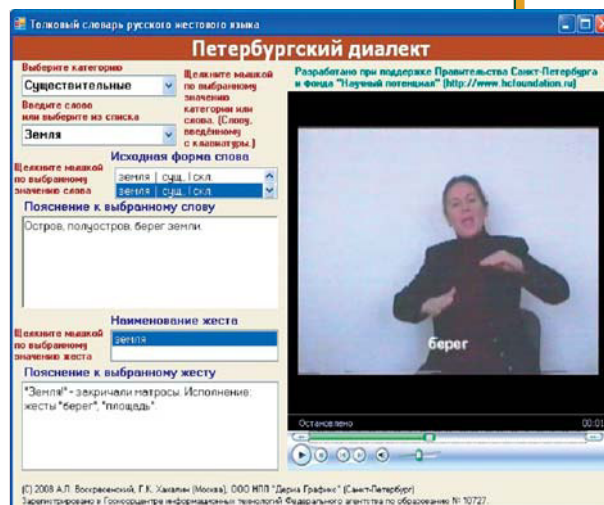


Рис. 2. Жест «берег» в ряду синонимов слова «земля» в словаре RuSLED

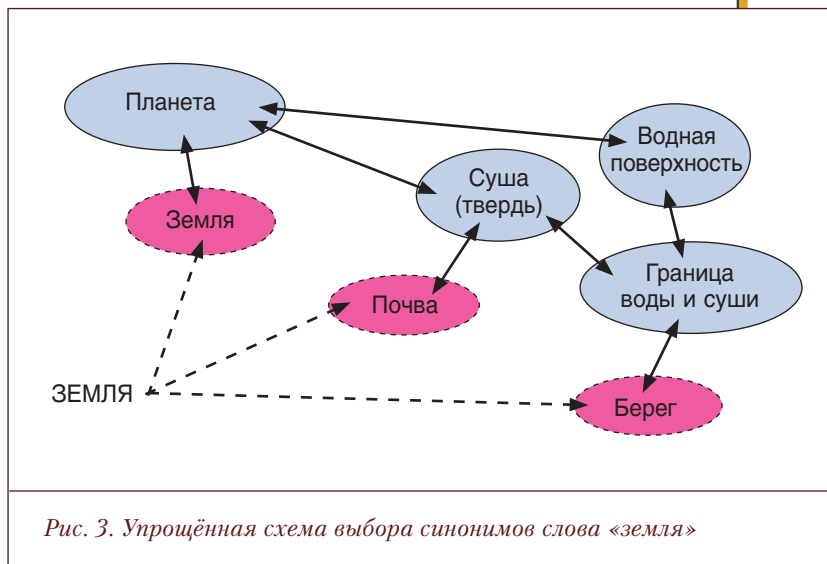


Рис. 3. Упрощённая схема выбора синонимов слова «земля»

Представленные рассуждения соответствуют традиционной системе логических умозаключений, известной со времён Аристотеля. Из известных прототипов систем искусственного интеллекта, использующих подобный подход, можно назвать, например, системы NARS и Novamente [16].

Для выполнения подобных рассуждений система должна иметь обширные знания об окружающем мире (или, по крайней мере, о тематике обрабатываемых текстов), форми-



рующие внутреннюю онтологию системы [6]. Эти знания могут пополняться не только за счёт содержания обрабатываемых текстов, но и из внешних источников, например, из сети Интернет.

При этом система должна «иметь собственное мнение» об окружающем мире и значениях слов, описывающих этот мир, поскольку семантическая разметка внешних источников информации может не отвечать требованиям решаемой задачи.

3. Возможный путь повышения качества распознавания речи

Система автоматизированного сурдоперевода, чтобы действительно быть полезной (например, во время публичных выступлений, лекций и т.п.), должна иметь подсистему распознавания речи. Соответственно, качество сурдоперевода будет во многом зависеть и от качества распознавания речи на входе системы.

Отвлечёмся от значений качества распознавания речи имеющихся систем и прототипов. Несомненно, что ко времени создания средств перевода текста в жесты они возрастут. Но чем ближе качество распознавания будет приближаться к идеалу, тем труднее будет движение к нему, тем больше затрат будет приходиться на каждую долю процента повышения качества работы системы распознавания.

Похожая ситуация была с системами оптического распознавания символов (OCR), которые используются для ввода в компьютер текста. В начале 90-х годов на российском рынке было представлено довольно много подобных систем, конкурирующих между собой. В рамках выполнения работ по созданию каталога российских коллективов, работающих в области обработки текста и речи [17], была поставлена задача оценки различных систем OCR. Практически все разработчики давали сведения о точности распознавания, равной 95–97%.

Но распознанный системой OCR текст ещё не является конечным продуктом. Для окончательной обработки необходимо выявить и устранить ошибки распознавания. Частично такая обработка выполняется путём автоматической проверки полученного текста с помощью словаря, включённого в систему OCR. Оставшиеся ошибки (и добавленные за счёт ошибочных срабатываний системы орфо коррекции) исправляются вручную. В [18] была предложена методика выявления ошибок распознавания, позволившая определить наиболее эффективные системы. Позднее данный прогноз оправдался.

Поскольку результаты работы системы распознавания речи во многих случаях должны представлять текст, целесообразно для коррекции ошибок распознавания использовать методы обработки полученного текста (морфологические, синтаксические, семантические).

Заключение

Работа по созданию системы автоматизированного сурдоперевода находится в начальной стадии. Имеющиеся предварительные результаты показывают, что такая система должна не просто конвертировать речь (и текст) из одной формы в другую, но и вести интеллектуальную обработку. Фактически система должна понимать обрабатываемые сообщения.

Таким образом, эта система должна решать задачи не менее, а возможно, и более сложные, чем системы автоматизированного перевода с одного словесного языка на другой.

Мы надеемся, что успешное завершение начатой работы окажется полезным не только для нужд сурдоперевода, но и для более широкого круга применений.

Литература

1. UN Resolution A/RES/48-96 (Part II, Rule 5, paragraph 7). <http://www.un.org/documents/ga/res/48/a48r096.htm>.
2. Voskresenski A. «Computer bank of sign languages», Conference abstracts, WISTCIS Outlook Conference «Information Society Priorities: New Prospects for European CIS Countries». Moscow, Russia, 20–21 November, 2003.
3. Voskresenski A. Signs and speech: two forms of human communication.. // Proceedings of the Ninth International Conference «Speech and Computer» SPECOM'2004, Saint-Petersburg, Russia, 2004. P. 666–669.
4. Voskresenskij A., Khakhalin G. Semantic Search Engine in a Multimedia Russian Sign Language Dictionary // Proceedings of the XIth International Conference «Speech and Computer» SPECOM'2007, October 15–18, 2007, Moscow, Russia. P. 739–744.
5. Voskresenskiy A.L., Gulenko I.E., Khakhalin G.K. RuSLED Dictionary as Tool for Semantic Study // Proceedings of XV International Conference «Dialogue-2009» on Computer Linguistics and Intellectual Technologies, May 27–31, 2009, Moscow, Russia.
6. Voskresenskij A. Text Disambiguation by Educable AI System.. // The First Conference on Artificial General Intelligence / P. Wang et al. (Eds.), AGI-08, 1–3 March, 2008, Memphis, USA. IOS Press, 2008.
7. Специфические средства общения глухих. Видеокурс. В 3 частях // СПб — Павловск: МЦР, 2002.
8. Фрадкина Р.Н. Говорящие руки: Тематический словарь жестового языка глухих России. // М.: Изд-во «Сопричастность» ВОИ, 2001. 598 с.
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.
10. Шиф Ж.И. Усвоение языка и развитие мышления у глухих детей. М., 1968.
11. Хахалин Г.К., Воскресенский А.Л. Контекстное фрагментирование в лингвистическом анализе // Десятая национальная конференция по искусственному интеллекту с международным участием КИИ-2006 (25–28 сентября 2006 г., Обнинск): Труды конференции. В 3 т. Т. 2. М.: Физматлит, 2006. С. 479–488.
12. 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.
13. Жигалов В.А., Жигалов Д.В., Жуков А.А., Кононенко И.С., Соколова Е.Г., Толдова С.Ю. Система ALEX как средство для многоцелевой автоматизированной обработки текстов // Компьютерная



лингвистика и интеллектуальные технологии: Труды Международного семинара Диалог'2002. Т. 2: Прикладные проблемы. М.: Наука, 2002. С. 192–208.

14. Воскресенский А.Л., Хахалин Г.К. Кластерный анализ контекста // Международная конференция «Математическая теория систем» МТС-2009 (26–30 января 2009 г., Москва, Россия): Труды конференции. М.: ИСА РАН, 2009. С. 102–106.

15. Леонтьева Н.Н. Автоматическое понимание текстов. Системы, модели, ресурсы. М.: Издательский центр «Академия», 2006. 304 с.

16. Artificial General Intelligence. /B. Goertzel, C. Pennachin (eds). Springer, 2007.

17. Воскресенский А.Л., Воскресенский В.А., Флюр-Семенова В. Интернет-версия Каталога российских коллективов и разработок в области автоматизированной обработки речи и текстов на естественном языке // Труды Международного семинара «Диалог-2001». Т. 2. Аксаково, 2001.

18. Arapov M., Voskresensky A., Semenova V. How to compare and evaluate OCR systems? (Our approach) // Proceedings of ELSNET GO EAST and IMACS Workshop on Integration of Language and Speech. 1995, Moscow, Russia, P. 5–10.

Воскресенский А.Л.
avosj@yandex.ru

Хахалин Г.К.
gkhakhalin@yandex.ru