М. Г. Гриф¹, О. О. Королькова², Ю. С. Мануева¹

¹ Новосибирский государственный технический университет
пр. Карла Маркса, 20, Новосибирск, 630073, Россия

² Новосибирский государственный педагогический университет
ул. Виллюйская, 28, Новосибирск, 630126, Россия

grifmg@mail.ru, ookorol@mail.ru, manueva.js@gmail.com

РАЗРАБОТКА АЛГОРИТМИЧЕСКОГО И ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ВЫБОРА ЗНАЧЕНИЯ МНОГОЗНАЧНОГО СЛОВА И ОМОНИМА В СИСТЕМЕ КОМПЬЮТЕРНОГО СУРДОПЕРЕВОДА РУССКОГО ЯЗЫКА НА ОСНОВЕ СЕМАНТИЧЕСКОЙ МОДЕЛИ

Проведен сравнительный анализ существующих систем компьютерного перевода со звучащего на жестовый язык. Выявлены преимущества и недостатки рассмотренных систем, на этом основании сделан вывод о необходимости учета семантической информации в процессе перевода. Рассмотрена технология семантического анализа. Описана разработанная модель выбора значения многозначного слова и омонима на основе семантического словаря В. А. Тузова и системы автоматической обработки текста «Диалинг». Приводится описание программного обеспечения и алгоритмическая часть программного обеспечения, направленного на выполнение семантического анализа и перевода. Приведены примеры использования программного обеспечения. Рассмотрены вопросы тестирования работоспособности модуля семантического анализа. Для решения этой задачи система семантического анализа была внедрена в систему компьютерного сурдоперевода «Сурдофон». Для анализа корректности работы семантического модуля проведено сравнение с определением семантических значений системами «Яндекс Переводчик» и «Google Переводчик». Разработанная система показала свое преимущество в ряде сложных случаев. База жестов РЖЯ была дополнена знаками, названиями которых являются омонимы и многозначные слова русского языка; выявлены особенности исполнения таких знаков.

Ключевые слова: русский жестовый язык, компьютерный сурдоперевод, семантический анализ, морфологический анализ, синтаксический анализ, омонимы, многозначные слова, синтаксические конструкции, альтернативы.

Введение

Лица с ограниченными возможностями здоровья по-прежнему остаются одной из наиболее социально уязвимых групп населения. По данным Министерства здравоохранения Российской Федерации, количество людей с ограниченными возможностями здоровья по слуху в России составляет около 13 млн человек. В качестве помощи данной категории граждан государством предусмотрено предоставление услуг сурдопереводчика в размере 40 часов в год. Эти меры не решают проблему коммуникации людей с дефектами слуха с обществом и государственной властью в силу недостаточности предоставляемого объема услуг сурдопереводчика. В связи с этим возникает необходимость использования мультимедийных компьютерных систем, которые будут переводить звучащую на русском языке речь на русский жестовый язык (РЖЯ) и с русского жестового языка на звучащую речь на русском языке.

Гриф М. Г., Королькова О. О., Мануева Ю. С. Разработка алгоритмического и программного обеспечения выбора значения многозначного слова и омонима в системе компьютерного сурдоперевода русского языка на основе семантической модели // Вестн. Новосиб. гос. ун-та. Серия: Лингвистика и межкультурная коммуникация. 2018. Т. 16, № 3. С. 32–44.

Большинство реализованных компьютерных сурдопереводчиков предназначено для перевода с английского языка на английский жестовый язык. Примерами таких систем являются: система перевода Zardoz, система машинного перевода TEAM, проект ViSiCAST [Гриф и др., 2014; Wakefield, 2002; Zhao et al., 2000; Veale, Conway, 1994]. Стратегия перевода в данных системах основана на морфологическом и синтаксическом анализе. Только в системе Zardoz дополнительно учитывается семантическая информация [Veale, Conway, 1994]. Для осуществления качественного перевода недостаточно морфологической и синтаксической информации, необходимо проводить семантический анализ предложения. Примером отечественной разработки в области компьютерного сурдоперевода с русского языка на РЖЯ является программный продукт «Сурдофон».

Использование семантических значений для перевода на жестовый язык обсуждается в работах [Гриф, 2012; Гриф, Тимофеева, 2012]. Существенный вклад в разработку компьютерного сурдопереводчика внесен компанией «Сурдофон», но данная разработка не учитывает семантическую составляющую русского языка и РЖЯ.

Соответствия между семантикой слов русского языка и жестами РЖЯ были положены в основу словаря RuSLED, разработанного А. Л. Воскресенским [2008], однако этот словарь служит только для рассмотрения отдельных слов, а не для перевода предложений.

Работа В. А. Тузова «Компьютерная семантика русского языка» [2003] посвящена описанию семантики только русского звучащего языка, но применимость предложенного метода для представления семантики РЖЯ не рассматривается.

Постановка задачи

Цель статьи заключается в описании разработанного метода выбора значения многозначного слова и омонима при переводе с русского языка на РЖЯ.

Для достижения поставленной цели были сформулированы и решены следующие задачи.

1. Разработана модель выбора значения многозначного слова и омонима на основе семантического словаря В. А. Тузова и системы автоматической обработки текста «Диаллинг».

2. Разработаны и реализованы алгоритмы, направленные на определение значения многозначного слова и омонима.

3. Разработана база данных на основе семантического словаря В. А. Тузова и базы жестов сибирского диалекта РЖЯ.

4. База жестов РЖЯ дополнена знаками, названиями которых являются омонимы и многозначные слова русского языка, выявлены особенности исполнения таких знаков.

5. Разработаны и реализованы алгоритмы перевода предложений на РЖЯ, использующие результаты определения значений слов.

6. Разработано программное обеспечение для перевода на РЖЯ.

Общая схема компьютерного перевода русской речи (текста) на русский жестовый язык

Представим предложение *Sentence* в виде функции контекста f над словами *Word*:

$$Sentence = f(Word_1, Word_2, \dots, Word_n),$$

где n – количество слов в предложении.

Каждое слово *Word* в предложении *Sentence* имеет набор различных значений *Meaning*. Представим слова в виде набора значений, составленного на основе семантического словаря В. А. Тузова:

$$Word_i = \{Meaning_1, Meaning_2, \dots, Meaning_m\},$$

где m – количество значений слова.

Далее представим каждое значение *Meaning* посредством следующего набора атрибутов:

$$Meaning_i = \{Class_i, Phraseologism_i, Case_i, Function_i, Class1_i, Case1_i, Class2_i, Case2_i, Gesture_{ij}\},$$

где

$Class_i$ – класс слова;

$Phraseologism_i$ – признак вхождения во фразеологический оборот, следующая часть фразеологического оборота;

$Case_i$ – падежная форма (определена для предлогов, каждая модель управления вынесена в отдельное значение);

$Function_i$ – признак необходимости для данного значения двух аргументов (большее количество аргументов в статье не рассматривается);

$Class1_i$ – класс первого подчиненного слова;

$Case1_i$ – падеж первого подчиненного слова;

$Class2_i$ – класс второго подчиненного слова;

$Case2_i$ – падеж второго подчиненного слова;

$Gesture_i$ – соответствующий значению жест русского жестового языка.

Слова в предложении связаны семантическими отношениями, полученными при обработке системой Диалинг:

$$R(A, B),$$

где

R – название семантического отношения;

A – зависимый член отношения;

B – управляющий член отношения.

Кроме того, каждое слово $Word$ имеет набор морфологических и синтаксических характеристик:

$$Word_i = \{MorthCharacteristic_i, SyntCharacteristic_i\},$$

где

$MorthCharacteristic_i$ – морфологические характеристики слова;

$SyntCharacteristic_i$ – синтаксические характеристики слова.

Морфологические характеристики $MorthCharacteristic$ содержат следующую информацию:

$$MorthCharacteristic_i = \{SpeechPart_i, Case_i\},$$

где

$SpeechPart_i$ – часть речи, морфологические характеристики слова;

$Case_i$ – падеж.

При выполнении семантического анализа предложения значение $Meaning_i$ слова $Word_i$ считается корректным при наличии признака необходимости для данного значения двух аргументов, если выполняется следующее условие. Значение $Meaning_i$ слова $Word_i$ считается корректным, если существуют два слова $Word_j$ и $Word_k$, состоящие в семантических отношениях с $Word_i$, а классы значений и падежи данных слов совпадают с соответствующими классами и падежами слова $Word_i$.

Если признак необходимости для данного значения двух аргументов отсутствует, то для признания значения $Word_i$ корректным необходимость наличия второго слова с соответствующими характеристиками отсутствует. Таким образом, значение $Meaning_i$ слова $Word_i$ считается корректным, если существует слово $Word_j$, которое находится в семантическом отношении с $Word_i$, а также класс $Class_j$ и падеж $Case_j$ совпадают с классом и падежом подчиненных слов для значения $Meaning_i$ слова $Word_i$.

На рис. 1 представлена модель выбора значения многозначного слова и омонима на основе семантического словаря В. А. Тузова и системы автоматической обработки текста «Диалинг» [Сокирко, 2000].

Для определения корректного значения многозначного слова и омонима в предложении применяются система «Диалинг», словарь В. А. Тузова и список жестов сибирского диалекта РЖЯ. На первом этапе введенное предложение обрабатывается системой «Диалинг» [Там же]. Выходной информацией из данной системы будет являться следующее: начальные формы слов в предложении, морфологические и синтаксические характеристики, семантические отношения. На втором этапе был проанализирован семантический словарь В. А. Тузова,

Рис. 1. Модель выбора значения многозначного слова и омонима

на основе которого была спроектирована база данных. Словарь, разработанный В. А. Тузовым, представляет собой текстовый документ формата «.txt» [Тузов, 2003]. Словарные статьи состоят из заголовочного слова и его семантического описания. Для работы со словарем необходимо разбить словарную статью на перечисленные выше атрибуты, с которыми в дальнейшем будет работать модуль семантического анализа (класс слова, признак вхождения во фразеологический оборот, следующая часть фразеологического оборота, падежная форма и др.). На третьем этапе в базу данных был добавлен список жестов сибирского диалекта РЖЯ. Также проведена работа по определению соответствия между жестами и значениями слов в словаре. На четвертом этапе в базе данных осуществляется поиск значений слов по начальным формам, полученным в результате работы системы «Диалинг». Затем полученные альтернативы обрабатываются модулем семантического анализа.

Основной этап призван решить проблему омонимии с использованием семантического словаря В. А. Тузова [2003]. После разрешения проблемы многозначности на данном этапе определяются соответствующие жесты. Результатом работы модуля являются единственные значения слов и соответствие «слово – жест». В более компактном виде модель выбора значения многозначного слова и омонима представлена на рис. 2. Процесс семантического анализа можно условно разбить на две части: первичный семантический анализ и основной семантический анализ. Задача первичного семантического анализа заключается в определении начальных форм слова, морфологических и синтаксических характеристик, а также семантических отношений. На основе полученной информации основной семантический анализ определяет значения слов и соответствие жестам РЖЯ.

Определим характеристики слова, необходимые для функционирования системы семантического анализа: начальная форма слова; класс слова, который, в свою очередь, делится на номер класса и, при наличии, слово, от которого произведено описываемое слово, и специальный признак производности лексемы; набор морфологических характеристик, часть речи; признак вхождения во фразеологический оборот, следующая часть фразеологического оборота; соответствующая падежная форма (для предлогов); признак необходимости для данного значения двух аргументов; классы и падежи подчиненных слов.

Рис. 2. Схема работы модуля семантического анализа

Алгоритм разбиения словарной статьи семантического словаря В. А. Тузова состоит из следующих этапов. На первом этапе из словарной статьи выделяется начальная форма слова. На следующем этапе определяются характеристики лексемы: относится она к базовой или производной. Затем делается вывод о том, входит или нет рассматриваемое слово во фразеологический оборот. В соответствии с полученными данными заполняются соответствующие характеристики для фразеологизма. Далее определяется часть речи анализируемого слова. Если это глагол, то выполняется проверка на тип семантической функции. В случае успешного прохождения проверки на тип функции с двумя аргументами в классы и падежи зависимых слов будут записаны два значения соответственно. Иначе будет заполнено по одному значению в классе и падеже зависимых слов. Отдельно обрабатываются предлоги: для этой части речи определяются предложно-падежные формы. Каждому варианту управления предлога соответствует отдельное значение. Для других частей речи заполняется значение класса и падежа, вторые значения не заполняются. На заключительном этапе определяется набор морфологических характеристик.

Рассмотрим пример разбиения для имени существительного. В качестве примера базисной лексики рассмотрим омоним *шляпа*¹ в значении 'головной убор'. Словарная статья имеет вид: «ШЛЯПА \$121366(РОД:!Род) == 0^06 {ж1 364} <1>». Начальная форма «шляпа» указана в начале словарной статьи. Класс слова \$121366 заключен после символа «\$». Так как *шляпа* является базисной лексемой, то в данном случае класс совпадает с номером класса, и отсутствует слово, от которого произведено описываемое слово. Лексема имеет признак *базовая*. Морфологические характеристики заключены в фигурные скобки {ж1 364}. Так как отсутствует символ «N», то данная лексема не входит во фразеологический оборот. Признак вхождения во фразеологический оборот заполняется значением «0». Признак необходимости для данного значения двух аргументов равен «0». Класс слов, с которыми взаимодействует слово *шляпа*, в данном значении входит в 1213. Остальные характеристики для базисной лексики отсутствуют.

Значение омонима *шляпа*² – 'глупый человек'. Словарная статья в данном случае имеет вид: «ШЛЯПА \$12413440/0/15(РОД:!Род,!Max) == 0^16 {M1o 364} <1>». Лексема также является базисной, класс отличается от класса предыдущего значения и равен «12413440/0/15». Остальные характеристики определяются по аналогии с предыдущим примером.

Рассмотрим дальнейшие преобразования, которым подвергается полученная на этапе семантического анализа информация (рис. 3).

Рис. 3. Модель перевода на жестовый язык

На вход модуля перевода поступает следующая информация: исходное предложение, информация, полученная в результате семантического анализа: соответствие «слово – жест», информация, полученная системой «Диаллинг» (морфологические и синтаксические характе-

ристики). Модель перевода заключается в изменении порядка слов в предложении в соответствии с требованиями русского жестового языка. Синтаксические конструкции, допустимые в жестовом языке, бывают двух типов: последовательность «субъект, глагол, объект» и «субъект, объект, глагол». С учетом семантической информации результатом работы системы является корректная последовательность жестов РЖЯ. Таким образом, предложение с русского языка переведено на РЖЯ.

На основе словаря В. А. Тузова посредством разбиения словарной статьи была спроектирована и реализована база данных. Она включает в себя 13 основных и вспомогательных таблиц. Также были проведены работы по определению соответствия значений слов жестам сибирского диалекта РЖЯ. В качестве исходной информации была взята база жестов сибирского диалекта РЖЯ [Королькова, 2014]. Далее были проведены работы по анализу значений жестов и приведения в соответствие со значениями слов в словаре В. А. Тузова. Для решения данной задачи в разработанную на основе семантического словаря структуру базы данных были добавлены таблицы, содержащие список жестов и список соответствия «значение слова – жест» РЖЯ. Рассмотрим пример составления такой таблицы. В РЖЯ для передачи значений слова *аудитория* имеется два различных жеста. Первый жест передает значение ‘помещение для чтения лекций’, второй – значение ‘слушатели’. В словаре В. А. Тузова также приведено два значения для слова *аудитория*: помещение и слушатели. Таким образом, двум разным значениям слова *аудитория* соответствуют два различных жеста.

Система семантического анализа русского текста

Рассмотрим описание программного обеспечения и алгоритмическую часть программного обеспечения, направленного на выполнение семантического анализа и перевода.

Рассмотрим работу программы на примере предложения «Ирина заплетала русую косу» и проведем его семантический анализ (рис. 4). Работа модуля семантического анализа завершается заполнением поля «Результат». В поле «Результат» представлено соответствие «слово – жест». Кроме того, если у слова отсутствует свой жест, то в поле «Результат» отображается указание, что данное слово необходимо показывать с использованием дактиля. Также отмечается способ поиска жеста в случае отсутствия полного совпадения: с помощью синонимов или антонимов.

Рис. 4. Вкладка «Перевод»

На вкладке «Первичный семантический анализ» отображаются результаты работы предварительного этапа разбора предложения. На данном этапе определяются морфологические, синтаксические характеристики и определяются семантические отношения (рис. 5).

Рис. 5. Вкладка «Первичный семантический анализ»

Рис. 6. Вкладка «Семантический анализ»

Рис. 7. Синтаксические преобразования

Рис. 8. Переведенное предложение с первым типом базового порядка слов

На вкладке «Семантический анализ» представлен «Список исходных альтернатив» и «Список итоговых альтернатив» (рис. 6). Первоначально мы имели три омонима: *коса*¹ (волосы), *коса*² (инструмент) и *коса*³ (берег). После проведения семантического анализа количество альтернатив сократилось до одной – *коса*¹ (волосы).

Выполним синтаксические преобразования. В разработанной системе доступно два варианта базового порядка слов (рис. 7).

Выберем вариант базового порядка слов «субъект, глагол, объект» (рис. 8). На вкладке «Результат» после слов «Синтаксические преобразования» добавлено исходное предложение с измененным порядком слов. После этого выводится перевод предложения на жестовый язык с учетом особенностей синтаксиса жестового языка.

Разработанная программа семантического анализа была внедрена в систему компьютерного сурдоперевода «Сурдофон». В рамках решаемой задачи рассмотрена архитектура системы, приведено описание серверного и клиентского модуля, а также их взаимодействие. Проведено тестирование работы системы «Сурдофон» с внедренным модулем семантического анализа. Для анализа корректности работы семантического модуля проведено сравнение с определением семантических значений системами «Яндекс Переводчик» и «Google Переводчик». Разработанная система показала свое преимущество в более сложных случаях.

Цель интеграции модуля семантического анализа в систему компьютерного перевода заключается в показе введенного текста корректными жестами, отражающими смысл сказанного. В качестве демонстрации приведено два примера: фразы «Шляпа проворонил автобус» и «Егор услышал хорошие вести». Для отражения работоспособности интегрированного модуля семантического анализа представлены изображения показа жестов без семантического анализа и с включенным семантическим блоком. Рассмотрим работу системы на примере предложения «Шляпа проворонил автобус» (рис. 9). На рис. 10, а фраза отображена некорректно, жестом со значением 'головной убор'. Таким образом, заложенный в предложение смысл оказался полностью потерян, и, для того чтобы его воспроизвести, нужно приложить колоссальные усилия, а при быстром показе это маловероятно. Кроме того, даже одно неправильно показанное предложение может полностью нарушить восприятие целого текста. При включении семантического анализа (рис. 10, б) слово *шляпа* отображается другим жестом, означающим 'неумный человек'.

Рис. 9. Ввод предложения в систему «Сурдофон»

Рис. 10. Перевод жестами слова *шляпа*:

а – без семантического анализа предложения;

б – при наличии семантического анализа

Проведем сравнительный анализ работы модуля семантического анализа с существующими лидерами в электронном переводе: «Яндекс Переводчик» и «Google Переводчик». Значения исследуемых слов в указанных системах определим с помощью компьютерных переводчиков для перевода на английский. Для корректной оценки результата перевода, выполненного этими системами, значение слова на английском языке будет определено с помощью онлайн словаря Мультигран.

В русском языке существуют омонимы *лисичка*¹ (животное) и *лисичка*² (гриб). В английском языке словам *лиса*, *лисичка*¹ (животное) соответствует слово *fox*, слову *лисичка*² (гриб) – *chanterelle*. Проанализируем корректность определения семантического значения слова *лисичка* в предложении «Лисичка убежала в лес» (рис. 11–14).

Рис. 11. Перевод системой «Яндекс Переводчик»

Рис. 12. Перевод системой «Google Переводчик»

Рис. 13. Перевод системой семантического анализа

Рис. 14. Перевод системой семантического анализа (список альтернатив)

Значение слова *лисичка* корректно определено системой «Яндекс Переводчик» и модулем семантического анализа. Переводчик компании Google перевел *лисичку* в значении ‘гриб’.

В 14 контрольных примерах система Яндекс допустила 4 ошибки, система Google – 5 ошибок. Процент ошибок сравниваемых систем достиг 26 %, т. е. перевод будет ошибочен почти в каждом пятом случае. С помощью разработанного модуля семантического анализа контрольные примеры были переведены правильно. Таким образом, можно сделать вывод, что разработанная система семантического анализа проводит более точный перевод при наличии лексической неоднозначности. Кроме того, система семантического анализа соотносит с каждым семантическим значением соответствующий жест РЖЯ, а порядок слов в предложении приводит в соответствие с требованиями жестового языка. Возможность использования Яндекс и Google переводчиков для перевода на жестовый язык на данный момент отсутствует.

Расширение базы жестов в системе компьютерного сурдоперевода

Список жестов сибирского диалекта, на который мы опирались в процессе разработки программного обеспечения, не является исчерпывающим, т. е. включающим все существующие в настоящее время жесты РЖЯ, так как, во-первых, русский жестовый язык, как и все естественные языки, постоянно развивается и пополняется новыми языковыми единицами, а во-вторых, сибирский диалект – это один из диалектов РЖЯ.

Следующим этапом работы над компьютерным сурдопереводом станет его проверка на языковом материале, полученном в результате сопоставительного анализа материалов словарей трех диалектов русского (российского) жестового языка [Королькова, 2014]. На первом этапе исследования методом сплошной выборки были выделены жесты и сформированы их списки в алфавитном порядке из четырех словарей РЖЯ: ВIDEOSловарь русского жестового языка¹, словарь И. Ф. Гейльмана [1975–1979]; Словарь русского жестового языка [2009], словарь Р. Н. Фрадковой [2001]. Общий объем жестов составил почти 13 000 языковых единиц. На втором этапе исследования была составлена сопоставительная таблица жестов, включенных в эти лексикографические источники, проведен сопоставительный анализ выделенных языковых единиц. В результате исследования нами был составлен общий список жестов, содержащихся в этих словарях, в который вошло около 6 200 единиц.

Дальнейший анализ полученного языкового материала позволил уточнить количество жестов, названиями которых являются омонимы и многозначные слова русского языка [Королькова, 2013], а также расширить представления об особенностях исполнения жестов, названиями которых являются омонимы русского языка [Королькова, 2015].

Исследование лексического материала позволило выявить существование 54 пар жестов, названиями которых являются омонимы русского языка, а также определить некоторые особенности их исполнения. В отличие от звучащих омонимов, жесты, названиями которых являются омонимы, исполняются по-разному, однако это исполнение позволяет точно передать соответствующее значение. К особенностям исполнения следует отнести использование в большинстве случаев не самостоятельного, а существующего жеста РЖЯ (например, для показа жеста ДУШИТЬ (духами) используется жест ДУХИ), реже используется контурный, рисующий (например, ЛИСТ (кусок материала) и др.), указательный (ПОЛ (настил) и ЯЗЫК (орган речи)) или символический (общепринятый) жест (например, ДА (согласие) и др.), а также жест, имитирующий соответствующее действие (например, БРАК (супружеский) и др.), возможно дактилирование некоторых жестов (например, РАК (онкологическое заболевание) и др.).

Нами выделено 280 жестов, названиями которых являются многозначные слова русского языка. С помощью этих жестов можно передать многие значения, которые имеют их звучащие аналоги. Основное отличие жестов, аналогичных многозначным словам, от их звучащих аналогов заключается в том, что их различное исполнение позволяет определить значение

¹ ВIDEOSловарь русского жестового языка // Институт социальной реабилитации НГТУ: сайт. Новосибирск, 2011. URL: <http://www.nisor.ru/snews/oa/>.

жеста без опоры на контекст. Ряд жестов этой группы относится к числу имитирующих. Некоторые жесты имеют исполнение, сходное с невербальным компонентом, сопровождающим соответствующие аналоги русского языка.

Таким образом, нами была дополнена база жестов РЖЯ, названиями которых являются омонимы и многозначные слова русского языка, и выявлены особенности их исполнения.

Заключение

Представленные в данной статье исследования составляют теоретическую и практическую основу для компьютерного сурдопереводчика русского языка. Программное обеспечение внедрено в систему компьютерного сурдоперевода русского языка «Сурдофон». Протестирована работоспособность разработанного программного обеспечения. Проведен сравнительный анализ корректности определения семантических значений в сопоставлении с результатами работы систем «Яндекс Переводчик» и «Google Переводчик».

Таким образом, в ходе исследования определено, что разработанная система семантического анализа проводит более точный перевод при наличии лексической неоднозначности. Разработанное программное обеспечение соотносит с семантическим значением слова соответствующий жест сибирского диалекта русского жестового языка. Возможность перевода системами Яндекс и Google на жестовый язык пока отсутствует.

Проведено тестирование работоспособности программного модуля семантического анализа и перевода на русский язык жестов. Для решения этой задачи система семантического анализа была внедрена в систему компьютерного сурдоперевода «Сурдофон». В рамках этой задачи рассмотрена архитектура системы, в которой приведено описание серверного и клиентского модуля, а также их взаимодействие. Проведено тестирование работы системы «Сурдофон» с внедренным модулем семантического анализа. Для анализа корректности работы семантического модуля проведено сравнение с определением семантических значений системами «Яндекс Переводчик» и «Google Переводчик». Разработанная система перевода показала свое преимущество в более сложных случаях.

Список литературы

Воскресенский А. Л. Сопоставительное лексикографическое описание слов русского языка и жестов языка глухих России в словаре RuSLED // Компьютерная лингвистика и интеллектуальные технологии: Материалы ежегодной Международной конференции «Диалог». М.: Изд-во РГГУ, 2008. № 7 (14). С. 91–96.

Гейльман И. Ф. Специфические средства общения глухих. Дактилология и мимика: Словарь. Л., 1975–1979. Ч. 1–4.

Гриф М. Г., Мануева Ю. С., Козлов А. Н. Построение системы компьютерного сурдоперевода русского языка // Тр. СПИИРАН. 2014. № 6 (37). С. 170–183.

Гриф М. Г., Тимофеева М. К. Проблема автоматизации сурдоперевода с позиции прикладной лингвистики // Сибирский филологический журнал. 2012. № 1. С. 211–219.

Гриф М. Г. Методы и технологии компьютерного сурдоперевода: Учеб. пособие. Новосибирск: Изд-во НГТУ, 2012. 71 с.

Королькова О. О. Особенности омонимии и полисемии в русском жестовом языке (на материале видеословаря русского жестового языка) // В мире научных открытий. 2013. № 5.1. С. 169–184.

Королькова О. О. Определение объема «Полного словаря русского жестового языка» // Современные исследования социальных проблем. 2014. № 3 (19). С. 69–74.

Королькова О. О. Особенности жестов русского жестового языка, названиями которых являются омонимы русского языка // В мире научных открытий. 2015. № 7.8 (67). С. 2931–2942.

Словарь русского жестового языка / Рук. авт. колл. В. З. Базоев. М.: Флинта, 2009. 528 с.

Сокирко А. Семантические словари в автоматической обработке текста (по материалам системы Диалинг). М., 2000. 108 с.

Тузов В. А. Компьютерная семантика русского языка: Учеб. пособие. СПб.: Изд-во СПбГУ, 2003. 391 с.

Фрадкина Р. Н. Говорящие руки. Тематический словарь жестового языка глухих России. М.: Московская городская организация ВОГ, 2001. 598 с.

Veale T., Conway A. Cross modal comprehension in ZARDOZ an English to sign-language translation system. Trinity College, 1994. 326 p.

Wakefield M. VisiCAST. Milestone: Final Report, 2002. 97 p.

Zhao L., Kipper K., Schuler W. A Machine Translation System from English to American Sign Language // Lecture Notes in Computer Science. 2000. Vol. 1934. P. 54–67.

Материал поступил в редколлегию 01.06.2018

Mikhail G. Grif¹, Olga O. Korolkova², Yuliya S. Manueva¹

¹ Novosibirsk State Technical University
20 Karl Marx Ave., Novosibirsk, 630073, Russian Federation

² Novosibirsk State Pedagogical University
28 Vilyuyetskaya Str., Novosibirsk, 630126, Russian Federation

grifmg@mail.ru, ookorol@mail.ru, manueva.js@gmail.com

**A NEW ALGORITHM AND OTHER SOFTWARE
FOR DISAMBIGUATION OF POLYSEMY AND HOMONYMY
FOR COMPUTER TRANSLATION INTO RUSSIAN SIGN LANGUAGE
BASED ON A SEMANTIC PRINCIPLE**

The paper analyses current computer Sign Language translation systems. Their advantages and disadvantages are detected. The main drawback is the lack of original text semantic analysis module capable of solving the task of disambiguation. A general scheme of translation system from phonic Russian to Russian Sign language including a module for semantic analysis is presented. It includes a block of source code analysis, developed by the authors, responsible for handling the semantic component of the Russian language. The semantic module relies on Tuzov's dictionary. The semantic analysis algorithm is also described. The text analysis is completed when each word gets only one semantic description thus solving the problem of ambiguity. The most important developments of the semantic analysis module include the following: expanded collection of gestures, parsing of complex sentences, account in the algorithm analyses predicates classifier of Russian Sign Language. Testing of algorithm is made.

The article compares the existing systems of computer translation from phonic to the sign language. The advantages and disadvantages of the considered systems are revealed and a conclusion is made about the need to take into account the semantic aspect of the translation process. A technology of semantic analysis is suggested. The model to choose an adequate meaning of a polysemic word or homonym on the basis of the automatic text processing system «Dialing» is described. Examples of the use of the software are given. The questions of testing the working capacity of the semantic analysis module are given due attention too. To enhance its efficiency, the system of semantic analysis was added to the translation system «Surdophone». To verify the efficiency of the semantic module's operation, a comparison is made with the definition of some words' semantic meanings by the systems «Yandex Translator» and «Google Translator». The present system showed its advantage in more complex cases. Also, the base of gestures of the RSL whose names are homonyms and polysemic words of the Russian language, were added and the features of their performance were revealed.

Keywords: Russian Sign language, computer sign language translation, semantic analysis, morphological analysis, syntactic analysis, homonyms, polysemous words, syntax constructions, alternatives.

References

Fradkina R. N. Govoryashchie ruki. Tematicheskii slovar' zhestovogo yazyka glukhikh Rossii [Speaking hands. Thematic Dictionary of the Sign Language of the Deaf Russia]. Moscow, Moskovskaya gorodskaya organizatsiya VOG, 2001, 589 p. (in Russ.)

Geyl'man I. F. Spetsificheskie sredstva obshcheniya glukhikh. Daktilologiya i mimika [Specific means of communication for the deaf. Dactylology and facial expressions. Dictionary]. Leningrad, LVTs, 1975–1979, pt. 1–4.

Grif M. G. Metodi i tehnologii kompyuternogo surdoperevoda [Methods and techniques of computer sign language]. A tutorial. Novosibirsk, NSTU Press, 2012, 71 p. (in Russ.)

Grif M. G., Manueva Yu. S., Kozlov A. N. Postroenie sistemy komp'iuternogo surdoperevoda russkogo iazyka [Development of computer interpretation system of Russian language]. *Trudy SPIIRAN [SPIIRAS Proceedings]*, 2014, no. 6 (37), p. 170–183. (in Russ.)

Grif M. G., Timofeeva M. K. Problemi avtomatizatsii surdoperevoda s pozitsii prikladnoi lingvistiki [The problem of automation of sign language from the perspective of applied linguistics]. *Sibirskii filologicheskii zhurnal [Siberian Journal of Philology]*, 2012, no. 1, p. 211–219. (in Russ.)

Korol'kova O. O. Opredeleniye ob'ema «Polnogo slovary russkogo zhestovogo yazyka» [Determination of the volume of «Complete Dictionary Russian Sign Language»]. *Sovremennyye issledovaniya sotsial'nykh problem [Modern Research of Social Problems]*, 2014, no. 3 (19), p. 69–74. (in Russ.)

Korol'kova O. O. Osobennosti omonimii i polisemii v russkom zhestovom yazyke (na materiale videoslovyar russkogo zhestovogo yazyka) [Features homonymy and polysemy in Russian Sign Language (based on videodictionary of Russian Sign Language)]. *V mire nauchnykh otkrytiy [In the World of Scientific Discoveries]*, 2015, no. 5.1, p. 169–184. (in Russ.)

Korol'kova O. O. Osobennosti zhestov russkogo zhestovogo yazyka, nazvaniyami kotorykh yavlyayutsya omonimy russkogo yazyka [Features gestures Russian Sign Language, the name of which is a homonym Russian Language]. *V mire nauchnykh otkrytiy [In the World of Scientific Discoveries]*, 2015, no. 7.8 (67), p. 2931–2942. (in Russ.)

Slovar' russkogo zhestovogo yazyka [Dictionary of Russian Sign Language]. Ruk. avtorsk. koll. V. Z. Bazoev. Moscow, Flinta, 2009. 528 p. (in Russ.)

Sokirko A. Semanticheskie slovari v avtomaticheskoi obrbotke teksta [Semantic dictionary in automated text processing]. Moscow, 2000, 108 p. (in Russ.)

Tuzov V. A. Komputernay semantika russkogo yazuka [Computer semantic of Russian language]. A tutorial. St. Petersburg, 2003, 391p. (in Russ.)

Veale T., Conway A. Cross modal comprehension in ZARDOZ an English to sign-language translation system. Trinity College, 1994, 326 p.

Voskresenskij A. L. Sopostavitel'noe leksikograficheskogo opisanie slov russkogo yazyka i zhestov yazyka gluhih Rossii v slovare RuSLED [A comparative lexicographic description of the words of the Russian language and gestures of the language of the deaf in the dictionary RuSLED]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Materialy ezhegodnoj Mezhdunarodnoj konferencii «Dialog» [Computer Linguistics and Intellectual Technologies: Materials of the annual International Conference «Dialogue»]*. Moscow, RSGU, 2008, no. 7 (14), p. 91–96. (in Russ.)

Wakefield M. VisiCAST. Milestone, Final Report, 2002, 97 p.

Zhao L., Kipper K., Schuler W. A Machine Translation System from English to American Sign Language. *Lecture Notes in Computer Science*, 2000, vol. 1934, p. 54–67.

For citation:

Grif Mikhail, Korolkova Olga, Manueva Yuliya. A New Algorithm and Other Software for Disambiguation of Polysemy and Homonymy for Computer Translation into Russian Sign Language Based on a Semantic Principle. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2018, vol. 16, no. 3, p. 32–44. (in Russ.)