

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPUTER ENGINEERING
AND CONTROL

УДК 004.82

DOI: 10.17212/1814-1196-2019-1-41-60

**Разработка предикатной модели выбора
лексических значений при переводе
на русский жестовый язык на основе
семантического словаря В.А. Тузова и системы
автоматической обработки текста «Диалинг»***

И.А. КАГИРОВ^{1,а}, Ю.С. МАНУЕВА^{2,б}

¹ 199178, РФ, Санкт-Петербург, 14-я линия Васильевского острова, 39, Санкт-Петербургский институт информатики и автоматизации РАН

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

^а kagirov@iias.spb.su ^б manueva.js@gmail.com

Рассмотрены существующие системы семантического анализа русского языка: многоцелевой лингвистический процессор «ЭТАП-3», технология ABBYY Comreno, технология Texterra, статистические и нейросетевые подходы в «Яндекс Переводчике» и «Google Переводчике». На основе анализа преимуществ и недостатков рассмотренных систем сделан вывод о необходимости разработки системы семантического анализа для повышения качества перевода и внедрения в систему перевода со звучащего русского языка на русский жестовый язык «Сурдофон». Целью данной статьи является описание разработанной предикатной модели семантического анализа системы компьютерного сурдоперевода русского языка на основе семантического словаря В.А. Тузова. Для определения корректного значения многозначного слова и омонима в предложении применяются система «Диалинг», словарь В.А. Тузова и список жестов русского жестового языка. Рассмотрена модель выбора значения многозначного слова и омонима, основанная на логике предикатов первого порядка. При выполнении семантического анализа предложения значение trueMeaning слова word считается корректным, если предикат P (word, trueMeaning) имеет значение «истина». Для повышения качества перевода и разрешения ситуаций, в которых отсутствует возможность однозначного определения значения многозначного слова, предложен статистический алгоритм на базе корпуса русского языка. Определен критерий оценки корректности найденного значения многозначного слова. Проведен сравнительный анализ разработанного модуля семантического анализа с существующими системами «Яндекс Переводчик» и «Google Переводчик». Сделан вывод, что разработанная система семантического анализа проводит более точный перевод при наличии лексической неоднозначности. Разработанная предикатная модель с учетом статистического алгоритма позволяет в сравнении с семантическим словарем В.А. Тузова решить проблему выбора лексического значения в случае, когда истинны условия для двух и более альтернативных значений.

* Статья получена 05 сентября 2018 г.

В данной ситуации определяется лучшее сочетание лексических значений по критерию близости к корпусу русского языка.

Ключевые слова: русский жестовый язык, компьютерный сурдоперевод, семантический анализ, предикатная модель, омонимы, статистический анализ

ВВЕДЕНИЕ

Лица с ограниченными возможностями здоровья являются одной из наиболее социально уязвимых групп населения. По данным Министерства здравоохранения Российской Федерации, количество людей с ограниченными возможностями здоровья по слуху составляет порядка 13 миллионов человек. В качестве помощи данной категории граждан государством предусмотрено предоставление услуг сурдопереводчика в размере 40 часов в год. Приведенные меры не решают в полном объеме проблему коммуникации людей с дефектами слуха с обществом и государственной властью, так как предоставляемого объема услуг переводчиков недостаточно. В связи с этим возникает необходимость в использовании мультимедийных компьютерных систем, которые будут переводить звучащую русскую речь на русский жестовый язык и обратно. 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 [1–3]. Стратегия перевода в данных системах основана на морфологическом и синтаксическом анализе. Только в системе Zardoz дополнительно учитывается семантическая информация. Для осуществления корректного перевода недостаточно морфологической и синтаксической информации, необходимо проводить семантический анализ предложения. Данные работы не могут быть в чистом виде применены для русского языка и русского жестового языка, так как рассчитаны только на американский жестовый язык [20–22].

1. ПОСТАНОВКА ЗАДАЧИ

В последние годы в России проблеме коммуникации людей с ограниченными возможностями по слуху стали уделять больше внимания. Применительно к определению лексических значений для перевода на жестовый язык можно привести работы М.Г. Грифа [20–23]. Существенный вклад в разработку компьютерного сурдопереводчика внесен компанией «Сурдофон», число используемых жестов русского жестового языка в котором составляет около 3500, но данная разработка не учитывает семантическую составляющую. Работа В.А. Тузова «Компьютерная семантика русского языка» посвящена описанию семантики только русского звучащего языка и представляет собой семантический словарь, содержащий 163 901 словарную статью [19]. В работе ученого не рассматривается применимость к русскому жестовому языку. На основе соответствия лексических значений слов русского языка жестам русского жестового языка А.Л. Воскресенским разработан словарь RuSLED, включающий 2372 слова и 2537 видеоизображений жестов. В данном словаре не осуществляется перевод предложений, а рассматриваются отдельные слова [24].

В настоящее время существуют различные системы перевода со звучащего языка на язык жестов [1–3]. Существующие системы компьютерного сурдоперевода (система Zardoz, система TEAM, система ViSiCAST, система «Сурдофон») обладают основным недостатком – отсутствие учета семантической составляющей звучащего языка и жестового языка. Для осуществления качественного перевода необходимо учитывать специфику двух языков. Рассмотрены возможные подходы для решения данной проблемы: многоцелевой лингвистический процессор «ЭТАП-3», который содержит около 100 000 лексических единиц, технология ABYY Compreno, технология Texterra, статистические и нейросетевые подходы в «Яндекс Переводчике», в котором количество поддерживаемых языков составляет 95, и «Google Переводчике», работающем со 105 языками. На основе анализа преимуществ и недостатков данных систем сделан вывод о необходимости разработки системы семантического анализа для повышения качества перевода и внедрения разработанной системы в программу «Сурдофон».

Исходными данными является предложение на русском языке, которое получено после этапа распознавания речи. Повышение качества перевода достигнуто путем внедрения в систему «Сурдофон» модуля семантического анализа.

Целью данной статьи является описание разработанной предикатной модели семантического анализа системы компьютерного сурдоперевода русского языка на основе словаря семантических отношений В.А. Тузова, а также статистического алгоритма, определяющего лексическое значение слова в случае, когда истинны условия для двух и более альтернативных значений.

2. ОБЗОР СИСТЕМ СЕМАНТИЧЕСКОГО АНАЛИЗА РУССКОГО ЯЗЫКА

Основным недостатком существующих систем компьютерного сурдоперевода является отсутствие учета семантической составляющей звучащего и жестового языков. Для осуществления качественного перевода необходимо учитывать специфику двух языков. Рассмотрим возможные подходы для решения данной проблемы: многоцелевой лингвистический процессор «ЭТАП-3», технологию ABYY Compreno, технологию Texterra, статистические и нейросетевые подходы в Яндекс и Google переводчиках.

Концепция многоцелевого лингвистического процессора «ЭТАП-3» принадлежит И.А. Мельчуку, который частично разработал правила и словари для анализа и синтеза текстов [25]. Научными сотрудниками Института проблем передачи информации им. А.А. Харкевича разработано программное обеспечение «ЭТАП-3», представляющее модель «Смысл-Текст» [4, 25]. Цель программы – анализ и синтез текстов для определения его смысла. Лексические функции решают задачи разрешения синтаксической омонимии и лексической неоднозначности, идиоматического перевода. Пример синтаксической омонимии представлен в примере «контроль правительства»: первый случай – правительство контролирует кого-то, второй – правительство контролируется. Процесс разрешения синтаксической омонимии заключается в определении роли слова «контроль» (осу-

ществлять или быть под, находиться под, подвергаться). Система «ЭТАП-3» способна также решить проблему лексической неоднозначности. Корректное значение определяется в случае, когда глагол употреблен в контексте существительного (например, держать слово, держать пари, держать зачет) [4].

Рассмотрим другую технологию анализа текстов на естественном языке ABBYU Compreno [5]. В отличие от систем, основанных на статистике и правилах, ABBYU Compreno выполняет полный семантико-синтаксический анализ текста, создает его универсальное представление, извлекает сущности, события и связи между ними. Рассмотрим этапы работы ABBYU Compreno (рис. 1) [5].

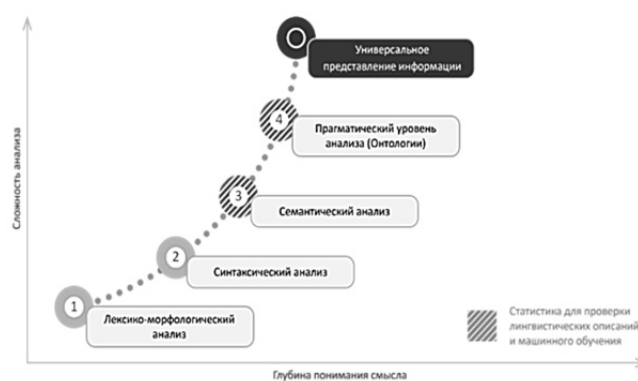


Рис. 1. Этапы работы ABBYU Compreno

Fig. 1. Stages of ABBYUCompreno work

Работа ABBYU Compreno начинается с этапа лексико-морфологического анализа. На первом этапе исходный текст разбивается на абзацы, предложения и слова. Затем программа определяет части речи и морфологические признаки слов, такие как род, число, падеж. На втором этапе проводится полный синтаксический анализ. Семантический анализ определяет значения каждого слова, далее разрабатывает семантическую структуру предложения на основе связей, которые были установлены на предыдущем этапе. Следующий этап заключается в прагматическом уровне анализа. На этом этапе накладывается прагматический слой анализа текста, применяются онтологии (терминология для конкретной предметной области анализа) и правила извлечения нужных объектов. Основным достоинством технологии ABBYU Compreno является определение смысла многозначных слов с помощью разрешения проблемы омонимии. К преимуществам данной технологии относятся точный анализ, быстрый запуск проекта и качественная работа с информацией на русском языке [5].

Рассмотрим подробнее технологию Textterra, которая представляет собой набор средств для автоматического построения онтологий, а также проведения семантического анализа предложений, разработанный в Институте системного программирования В.П. Иванникова [6]. В качестве решаемых задач с использованием данной технологии приводятся следующие: анализ отзывов, поиск документов и построение баз знаний. В основе технологии лежат два основных метода: методы получения данных и методы компьютер-

ной лингвистики. На первом этапе происходит разбиение текста на отдельные предложения, затем на отдельные слова. На втором этапе выполняются алгоритмы морфологического анализа, которые определяют морфологические характеристики и начальные формы слов. На третьем этапе проводится синтаксический анализ для определения связей между словами и выполняются алгоритмы, направленные на разрешение кореферентности. Также модуль содержит алгоритм исправления ошибок. В системе реализованы алгоритмы определения именованных сущностей, которые встречаются в текстах с указанием конкретных имен, и алгоритмы, осуществляющие поиск соответствий между понятиями и фрагментами текстов, определяющих семантику данных понятий [6].

Переводчик российской компании «Яндекс» представляет собой сервис, предназначенный для автоматического перевода слов, фраз, целых текстов. Работа сервиса основана на собственной разработке технологии машинного перевода. Количество поддерживаемых языков составляет 95 языков. Для данной статьи наибольший интерес представляет механизм перевода, а в частности, семантический анализ переводимых текстов. С 2011 года в «Яндекс Переводчике» работает система статистического перевода. В основе статистического подхода лежит загрузка словарей и грамматик различных языков, множество параллельных текстов, которые являются переводами друг друга. В 2017 году в компании «Яндекс» в дополнение к статистическому переводу внедрена нейронная сеть, которая, как и система статистического анализа, также обучается на текстах. Наилучший результат достигается при совместной работе этих двух подходов.

В статистическом подходе исходное предложение разделяется на отдельные слова и фразы. Затем проводится перебор всех вариантов и происходит оценка вероятности каждого из них. Оценка вероятности производится на основе того, сколько раз данный вариант встречался в загруженных текстах. Например, английское слово *spring* имеет два основных значения – «весна» и «пружина». В рассмотренном случае если рядом стоит слово *early*, то скорее всего имеется в виду значение «весна». Если в предложении стоит слово *metal*, то наиболее вероятен перевод слова *spring* как «пружина». Основное преимущество статистической модели – это возможность обрабатывать редкие и сложные слова, а также фразы. У данного подхода есть и существенный недостаток, который заключается в отсутствии построения качественной полной картины текста.

Нейросетевой подход также применим к таким видам информации, как речь и изображение. Для системы, основанной на нейронной сети, перевод текста является одной из возможных для решения с ее помощью задач. Обучение нейронной сети проводится на текстовых массивах, которые представляют собой текст оригинала и текст перевода. В процессе перевода нейросеть оперирует не словами и фразами, как при статистическом переводе, а целыми предложениями. В качестве входной информации в данном случае является предложение на одном языке, а выходная информация – это переведенное предложение. Нейросетевой подход успешно применяется даже в случае, если слова, которые передают смысл предложения, находятся в разных его частях. Пример перевода предложения с использованием статистического и нейросетевого подхода представлен на рис. 2.

Подход, основанный на нейронных сетях, имеет следующие недостатки. Если нейросеть не сможет перевести предложение, то результат перевода будет некорректным. Также для нейронной сети сложностью является перевод редких слов, таких как редкие имена, топонимы. В данной ситуации на помощь придет статистический перевод, если данное слово встречалось в его текстах. Таким образом, в системе перевода компании «Яндекс» используется гибридный подход на основе систем нейросетевого и статистического перевода. Введенный текст обрабатывается одновременно обеими системами, затем проводится оценка перевода. Гибридная система применяется только для перевода с английского на русский. На данный момент для остальных языков работает только статистическая модель [7, 8].

исходное предложение:	Can I have a banana?
статистический перевод:	Я могу есть банан?
нейросетевой перевод:	Можно банан?

Рис. 2. Пример перевода с использованием статистического и нейросетевого подхода

Fig. 2. Example of translation using the statistical and neural network approach

«Google Переводчик» – бесплатный многоязыковой сервис машинного перевода текстов, разработанный компанией Google. «Google Переводчик» поддерживает более чем 100 языков [9]. В конце 2016 года переводчики стали применять технологию нейронных сетей (Google Neural Machine Translation) [10]. В «Google Переводчике» используются два вида перевода: статистический перевод и перевод, основанный на нейронных сетях. С 2006 года основным методом перевода является статистический перевод, в основе алгоритмов которого лежат исследовательские работы Франца Оча [11, 12]. В исследованиях в качестве необходимых компонент для качественной работы статистического перевода для новой пары языков указаны параллельные тексты на двух внедряемых языках, насчитывающие более 150–200 миллионов слов, и отдельные тексты на каждом языке, размеры которых должны быть около миллиарда слов. Затем полученный языковой корпус используется алгоритмами статистического перевода. В системе Google отсутствует прямой перевод: сначала выполняется перевод на английский язык и затем на язык перевода [13]. В процессе перевода запускается алгоритм поиска в миллионах документов для улучшения качества перевода. В случае обнаружения документа, переведенного человеком-переводчиком, система Google выдвигает предположения в области лучшего перевода. С 2016 года в компании ведутся разработки в области перевода, основанного на нейронных сетях (Google Neural Machine Translation) [14]. В частности, используются сети с короткой памятью. Нейронная сеть улучшает качество перевода за счет использования метода перевода на основе примера, в котором система «учится на милли-

онах примеров». Затем полученный результат преобразуется для получения более естественного перевода с правильной грамматикой [15]. Первыми языками для тестирования работоспособности нейронной сети стали английский, французский, немецкий, португальский, испанский, турецкий, китайский, японский и корейский языки. В 2017 году были добавлены хинди, русский и вьетнамский языки. В сравнении со статистическим переводом нейронные сети переводят напрямую без использования языка-посредника. Основным недостатком системы является зависимость качества перевода от используемых языков.

3. СИСТЕМА СЕМАНТИЧЕСКОГО АНАЛИЗА ДЛЯ ПЕРЕВОДА НА РУССКИЙ ЖЕСТОВЫЙ ЯЗЫК

На рис. 3 представлена схема работы разработанной системы семантического анализа.



Рис. 3. Схема выбора значения многозначного слова и омонима

Fig. 3. Scheme for choosing the meaning of a multi-valued word and a homonym

Для определения корректного значения многозначного слова и омонима в предложении применяется система «Диалинг» и используются словарь В.А. Тузова и список жестов русского жестового языка [16–18]. На первом этапе введенное предложение обрабатывается системой «Диалинг» [17]. В качестве выходной информации из данной системы будут являться начальные формы слов в предложении, морфологические и синтаксические характеристики, семантические отношения. На втором этапе был проанализирован семантический словарь В.А. Тузова, на основе которого спроектирована база данных. Словарь, разработанный В.А. Тузовым, представляет собой текстовый документ формата «.txt» [19]. Для работы со словарем необходимо разбить словарную статью на компоненты, с которыми в дальнейшем будет работать модуль семантического анализа. Большинство слов в словаре имеет более одного значения. На третьем этапе в базу данных был добавлен список жестов русского жестового языка [16]. Проведена работа по определению соответствия между жестами и значениями слов в словаре. На данном этапе завершаются подготовитель-

ные исследования, необходимые для работы модуля семантического анализа. На четвертом этапе в базе данных осуществляется поиск значений слов по начальным формам, полученным в результате работы системы «Диалинг». Затем значения обрабатываются модулем семантического анализа. Основной этап работы модуля семантического анализа – решение проблемы омонимии с использованием семантического словаря В.А. Тузова. После разрешения многозначности на данном этапе определяются соответствующие жесты. Результатом работы модуля являются единственные значения слов и соответствие слово–жест. 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 русского жестового языка.

Анализ русского текста состоит из следующих этапов: морфологического анализа слов в предложении, синтаксического анализа структуры предложения и семантического анализа слов в предложении [2].

Морфологический и синтаксический анализ выполняется на основе системы «Диалинг» [17], в которой морфологический анализ состоит из морфоанализа и лемматизации словоформ. Под лемматизацией будем понимать приведение различных форм слова к словарным, а под морфоанализом – определение морфологических характеристик слова. Работа морфологического модуля основана на использовании трех морфологических словарей: большого словаря, который базируется на грамматическом словаре А.А. Зализняка, словаря имен собственных и словаря географических слов. На этапе лемматизации происходит определение начальной формы слова, необходимой для дальнейшей работы со словарями. Морфологическая часть речи определяется традиционным образом. При анализе реальных текстов необходимо учитывать соседние слова, потому что от выбора конкретной интерпретации зависит выбор интерпретации другого слова. Объем морфологического словаря составляет более 130 тысяч лексем. Если в словаре отсутствует данная словоформа, то применяется алгоритм, который ищет в словаре словоформу, максимально совпадающую с конца со входной словоформой.

В качестве входной информации синтаксического анализа поступают результаты морфологического анализа. На выходе получаем набор семантических отношений. Определим семантическое отношение как универсальную связь, которая усматривается носителем языка в конкретном контексте. Любой текст представляет собой набор различных отношений. Главный недостаток отношений заключается в том, что одни отношения похожи на другие, но определить общие черты отношений и их различия непросто. Так как в существующих системах перевода отсутствует модуль, выполняющий семантический анализ, то далее будет более подробно рассмотрена система семантического анализа на основе предикатной модели.

4. ПРЕДИКАТНАЯ МОДЕЛЬ ВЫБОРА ЛЕКСИЧЕСКИХ ЗНАЧЕНИЙ ПРИ ПЕРЕВОДЕ НА РУССКИЙ ЖЕСТОВЫЙ ЯЗЫК

Рассмотрим модель выбора значения многозначного слова и омонима, основанную на логике предикатов первого порядка. Для проведения семантического анализа из морфологических характеристик, полученных на предыдущем этапе, рассмотрим только часть речи и падеж.

Морфологические характеристики слова в предложении представим в виде предиката Mch :

$$Mch(word, [speechPart, case]), \quad (1)$$

где $word$ – исходное слово (входная переменная); $speechPart$ – часть речи слова (выходная переменная); $case$ – падеж слова (выходная переменная).

Семантический словарь В.А. Тузова представим как предикат V :

$$V(word, [l_1, ..., l_n]), \quad (2)$$

где $word$ – исходное слово (входная переменная); $l_1, ..., l_n$ – лексические значения слова (выходная переменная); n – количество лексических значений слова (выходная переменная).

Фразеологический оборот будет иметь вид предиката Ph :

$$Ph(word, [p_1, ..., p_n], [sw_1, ..., sw_n]), \quad (3)$$

где $word$ – исходное слово (входная переменная); $p_1, ..., p_n$ – признак нахождения слова во фразеологическом обороте (выходная переменная); $sw_1, ..., sw_n$ – второе слово фразеологического оборота (выходная переменная); n – количество лексических значений слова (выходная переменная).

Семантические отношения слова, полученные системой «Диалинг», определим, как предикат Sr :

$$Sr(word, sentence, [sr_1, ..., sr_m]), \quad (4)$$

где $word$ – исходное слово (входная переменная); $sentence$ – исходное предложение (входная переменная); $sr_1, ..., sr_m$ – слово предложения, состоящее в семантических отношениях с анализируемым словом (выходная переменная); m – количество слов, находящихся в семантических отношениях с анализируемым словом (выходная переменная).

Класс слова представим в предикате C следующего вида:

$$C(word, [c_1, ..., c_n]), \quad (5)$$

где $word$ – исходное слово (входная переменная); $c_1, ..., c_n$ – классы лексических значений (выходная переменная); n – количество лексических значений слова (выходная переменная).

Для работы семантического модуля также необходимо учесть признак необходимости для лексического значения двух подчиненных слов, который представим в виде предиката Fl :

$$Fl(word, [fl_1, ..., fl_n]), \quad (6)$$

где $word$ – исходное слово (входная переменная); $fl_1, ..., fl_n$ – признак необходимости для лексического значения двух подчиненных слов (принимает значение истина/ложь) (выходная переменная); n – количество лексических значений слова (выходная переменная).

В формуле (7) представлен предикат Cth_i , определяющий класс i -го первого подчиненного слова:

$$Cth_i(word, sr_i, [cth_{i1}, ..., cth_{ik}]), \quad (7)$$

где $word$ – исходное слово (входная переменная); sr_i – i -е слово предложения, состоящее в семантических отношениях с анализируемым словом (входная переменная); $cth_{i1}, ..., cth_{ik}$ – классы лексических значений i -го подчиненного слова (выходная переменная); k – количество лексических значений слова (выходная переменная).

Класс j -го второго подчиненного слова будет иметь вид предиката Cth_j :

$$Cth_j(word, sr_j, [cth_{j1}, ..., cth_{jr}]), \quad (8)$$

где $word$ – исходное слово (входная переменная); sr_j – j -е слово предложения, состоящее в семантических отношениях с анализируемым словом (входная переменная); $cth_{j1}, ..., cth_{jr}$ – классы лексических значений j -го подчиненного слова (выходная переменная); r – количество лексических значений слова (выходная переменная).

Для определения лексического значения рассмотрим предикат решения $Decision$. Для случая, когда для подтверждения корректности лексического значения необходимо одно зависимое слово, и данное слово не входит во фразеологический оборот, то предикат $Decision$ будет иметь вид

$$Decision(word, trueMeaning, [l_1, ..., l_n], [c_1, ..., c_n], [speechPart, case], [fl_1, ..., fl_n], [p_1, ..., p_n], [sw_1, ..., sw_n], [sr_1, ..., sr_m], [cth_{j1}, ..., cth_{jr}]), \quad (9)$$

где $word$ – исходное слово (входная переменная); $trueMeaning$ – корректное лексическое значение слова в данном контексте (выходная переменная); $l_1, ..., l_n$ – лексические значения слова (входная переменная); $c_1, ..., c_n$ – классы лексических значений (входная переменная); n – количество лексических значений слова (выходная переменная); $speechPart$ – часть речи слова (входная переменная); $case$ – падеж слова (входная переменная); $fl_1, ..., fl_n$ – при-

знак необходимости для лексического значения двух подчиненных слов (принимает значение истина/ложь) (входная переменная); p_1, \dots, p_n – признак нахождения слова во фразеологическом обороте (входная переменная); sw_1, \dots, sw_n – второе слово фразеологического оборота (входная переменная); sr_j – j -е слово предложения, состоящее в семантических отношениях с анализируемым словом (входная переменная); $cth_{j1}, \dots, cth_{jr}$ – классы лексических значений j -го подчиненного слова (входная переменная); r – количество лексических значений слова (входная переменная).

Иными словами, при выполнении семантического анализа предложения значение *trueMeaning* слова *word* считается корректным при отсутствии признака необходимости для данного значения двух аргументов, если предикат $P(word, trueMeaning)$ имеет значение «истина»:

$$\begin{aligned} P(word, trueMeaning) \leftarrow & Mch(word, [speechPart, case]) \wedge V(word, [l_1, \dots, l_n]) \wedge \\ & \neg Ph(word, [p_1, \dots, p_n], [sw_1, \dots, sw_n]) \wedge Sr(word, sentence, [sr_1, \dots, sr_m]) \wedge \\ & C(word, [c_1, \dots, c_n]) \wedge \neg Fl(word, [f_1, \dots, f_n]) \wedge Cth_i(word, sr_i, [cth_{i1}, \dots, cth_{ik}]) \wedge \end{aligned} \quad (10)$$

$$\begin{aligned} & Decision(word, trueMeaning, [l_1, \dots, l_n], [c_1, \dots, c_n], [speechPart, case], \\ & [fl_1, \dots, fl_n], [p_1, \dots, p_n], [sw_1, \dots, sw_n], [sr_1, \dots, sr_m], [cth_{i1}, \dots, cth_{ir}]). \end{aligned}$$

Если для подтверждения корректности значения слова необходимо наличие двух зависимых слов, и данное слово не входит во фразеологический оборот, то предикат $P(word, trueMeaning)$ примет вид

$$\begin{aligned} P(word, trueMeaning) \leftarrow & Mch(word, [speechPart, case]) \wedge V(word, [l_1, \dots, l_n]) \wedge \\ & \neg Ph(word, [p_1, \dots, p_n], [sw_1, \dots, sw_n]) \wedge Sr(word, sentence, [sr_1, \dots, sr_m]) \wedge C(word, \\ & [c_1, \dots, c_n]) \wedge Fl(word, [f_1, \dots, f_n]) \wedge Cth_i(word, sr_i, [cth_{i1}, \dots, cth_{ik}]) \wedge \end{aligned} \quad (11)$$

$$Cth_j(word, sr_j, [cth_{j1}, \dots, cth_{jk}]) \wedge$$

$$\begin{aligned} & Decision(word, trueMeaning, [l_1, \dots, l_n], [c_1, \dots, c_n], [speechPart, case], [fl_1, \dots, fl_n], \\ & [p_1, \dots, p_n], [sw_1, \dots, sw_n], [sr_1, \dots, sr_m], [cth_{i1}, \dots, cth_{ir}], [cth_{j1}, \dots, cth_{jk}]). \end{aligned}$$

Предикат $P(word, trueMeaning)$ для фразеологического оборота будет иметь вид (формула 12):

$$\begin{aligned}
 P(word, trueMeaning) \leftarrow & Mch(word, [speechPart, case]) \wedge V(word, [l_1, \dots, l_n]) \wedge \\
 & Ph(word, [p_1, \dots, p_n], [sw_1, \dots, sw_n]) \wedge Sr(word, sentence, [sr_1, \dots, sr_m]) \wedge \\
 & C(word, [c_1, \dots, c_n]) \wedge \neg Fl(word, [f_1, \dots, f_n]) \wedge Cth_i(word, sr_i, [cth_{i1}, \dots, cth_{ik}]) \wedge (12) \\
 & Decision(word, trueMeaning, [l_1, \dots, l_n], [c_1, \dots, c_n], [speechPart, case], [fl_1, \dots, fl_n], \\
 & [p_1, \dots, p_n], [sw_1, \dots, sw_n], [sr_1, \dots, sr_m], [cth_{i1}, \dots, cth_{ir}])
 \end{aligned}$$

На основе предложенной предикатной модели разработаны и реализованы алгоритмы, направленные на определение значений многозначных слов и омонимов, работа которых основана на взаимодействии с базой данных, включающей в себя соответствие слов русского языка жестам русского жестового языка.

5. СТАТИСТИЧЕСКИЙ АЛГОРИТМ ОПРЕДЕЛЕНИЯ ЛЕКСИЧЕСКИХ ЗНАЧЕНИЙ

Для повышения качества перевода и разрешения ситуаций, в которых отсутствует возможность однозначного определения значения многозначного слова, предложен статистический алгоритм на основе корпуса русского языка.

Рассмотрим упрощенное описание алгоритма выбора сочетаний альтернативных лексических значений многозначных слов и омонимов. Входное предложение представим в виде последовательности слов:

$$L(\text{последовательность Слов } l_1, \dots, l_n). \quad (13)$$

Кроме того, некоторые слова l_i предложения L могут являться многозначными и иметь несколько значений l_{ij} :

$$l_i \in L. \quad (14)$$

Множество сочетаний таких значений обозначим через M . Имеется множество сочетаний слов русского языка, полученных в рамках корпуса C [5]. Каждое словосочетание c корпуса C имеет длину $D(c)$, которая представляет собой число слов в словосочетании (от 2 до 5) и встречаемость в корпусе $Z(c)$:

$$c \in C; \quad (15)$$

$$0 < Z(c) < 1. \quad (16)$$

Необходимо выбрать наилучшее m :

$$m \in M. \quad (17)$$

Проведем следующую модификацию предложения. Исходное предложение L заменяем на L' путем замены слов $l_i \in L$ в соответствии с $m \in M$.

Для L' определим его покрытие словосочетаниями c . Алгоритм поиска покрытия начинается с первого слова в предложении L' , причем покрытия ищутся в порядке убывания $D(c)$ с продвижением до конца предложения. В общем случае будем иметь следующие покрытия:

1) k_2 покрытий длины 2, соответствующих сочетаниям $c(2)_1, c(2)_2, \dots, c(2)_{k_2}$;

2) k_3 покрытий длины 3, соответствующих сочетаниям $c(3)_1, c(3)_2, \dots, c(3)_{k_3}$;

3) k_4 покрытий длины 4, соответствующих сочетаниям $c(4)_1, c(4)_2, \dots, c(4)_{k_4}$;

4) k_5 покрытий длины 5, соответствующих сочетаниям $c(5)_1, c(5)_2, \dots, c(5)_{k_5}$.

Рассчитаем критерий оценки корректности найденного значения многозначного слова:

$$F(m) = \sum_{j=1}^{k_i} \sum_{i=2}^5 iD(c(i)_j). \quad (18)$$

Если покрытия сочетаниями длины k_3 , $2 \leq k \leq 5$, в L' отсутствуют, то они входят с нулевым весом в $F(m)$. Выбираем такое m , которое максимизирует $F(m)$.

Таким образом, разработанный алгоритм на основе корпуса русского языка позволяет решить проблему выбора лексических значений в случае, когда истинны условия для двух и более альтернативных значений в семантическом словаре В.А. Тузова.

6. ТЕСТИРОВАНИЕ МОДУЛЯ СЕМАНТИЧЕСКОГО АНАЛИЗА

Для подтверждения работоспособности разработанного модуля проведем сравнительный анализ с существующими лидерами в электронном переводе – «Яндекс Переводчиком» и «Google Переводчиком». Для корректной оценки результата перевода на английский язык системами «Яндекс Переводчик» и «Google Переводчик» значения слов на английском языке будут определены с помощью онлайн-словаря «Мультитран». Поясним используемую процедуру перевода на следующем примере. В русском языке существуют омонимы «лисичка» (животное) и «лисичка» (гриб). В английском языке словам «лиса» и «лисичка» (животное) соответствует слово fox, слову «лисичка» (гриб) – слово chanterelle. Проанализируем корректность определения семантического значения слова «лисичка» в предложении «Лисичка убежала в лес» (рис. 4–7).

Яндекс Переводчик



Рис. 4. Перевод системой «Яндекс Переводчик»

Fig. 4. Translation using the Yandex Translator system

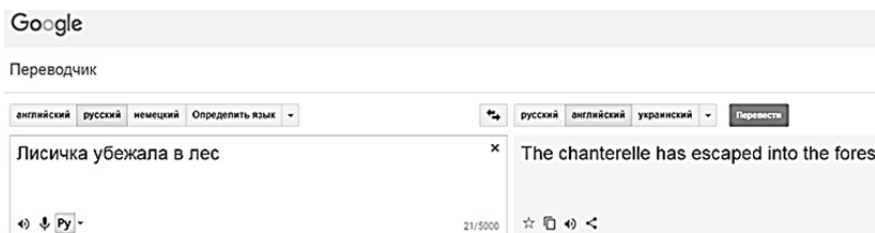


Рис. 5. Перевод системой «Google Переводчик»

Fig. 5. Translation using the Google Translator system

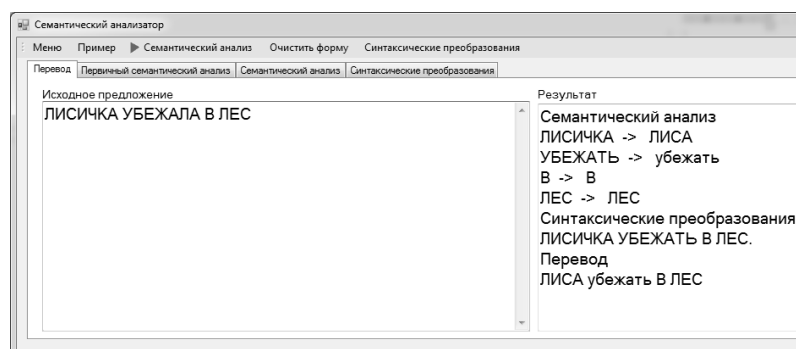


Рис. 6. Перевод системой семантического анализа

Fig. 6. Translation using the semantic analysis system

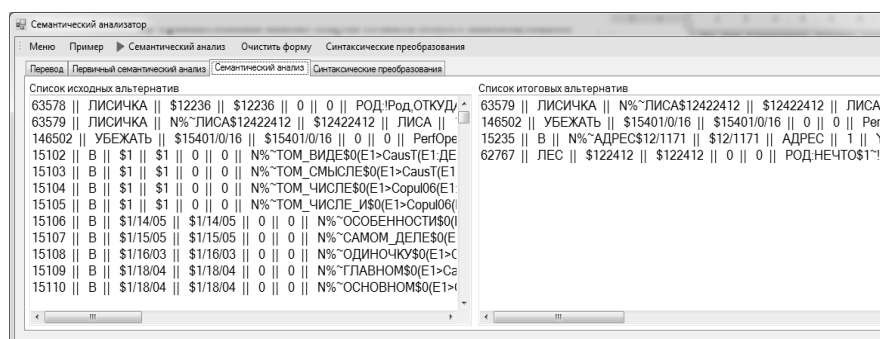


Рис. 7. Перевод системой семантического анализа (список альтернатив)

Fig. 7. Translation using the semantic analysis system (list of alternatives)

Значение слова «лисичка» корректно определено системой «Яндекс Переводчик» и модулем семантического анализа. Переводчик компании Google перевел слово «лисичка» со значением «гриб».

При сравнении использовался модуль семантического анализа без применения статистического алгоритма, рассмотренного в п. 5. Что касается определения безошибочного перевода модулем семантического анализа, то здесь необходимо отметить, что при невозможности выбрать конкретное лексическое значение программа оставляет исходное слово (предложение) без изменений. В этом случае результат работы данного модуля мы также будем считать «правильным». Следовательно, модуль семантического анализа осуществляет перевод с ошибкой, если только осуществляет замену слова на некорректное лексическое значение. Для оценки качества перевода были привлечены квалифицированные лингвисты, специалисты в лексике русского и русского жестового языков. Что касается «Яндекс Переводчика» и «Google Переводчика», то они не имели возможности, в отличие от модуля семантического анализа, оставить предложение неизменным и «были вынуждены» выполнять перевод с русского на английский, рискуя, таким образом, допустить ошибку.

Тестирование проводилось в два этапа. На первом этапе были разработаны 14 специально подобранных контрольных примеров (предложений), которые содержали многозначные слова и омонимы, разбор которых содержался в словаре В.А. Тузова. Система «Яндекс» допустила при их разборе 4 ошибки, а система «Google Переводчик» – 5 ошибок. Процент ошибок сравниваемых систем достиг 26 % [23]. С помощью разработанного модуля семантического анализа все контрольные примеры были переведены правильно.

На втором этапе тестирования модуля семантического анализа были подготовлены 100 случайно выбранных предложений в национальном корпусе русского языка [26]. При подборе данных предложений учитывалось лишь наличие в них многозначных слов и омонимов, которые входят и в словарь В.А. Тузова. Кроме того, данные слова должны были содержаться в двух-трех предложениях из ста.

В результате тестирования переводчики Google и «Яндекс» допустили ошибки в девяти предложениях из ста (модуль семантического анализа все предложения перевел правильно). Таким образом, на данной выборке разработанный модуль семантического анализа продемонстрировал 9 % преимущество в сравнении с переводчиками Google и «Яндекс».

В дальнейшем планируется провести исследование на качество перевода статистического алгоритма в разработанной системе семантического анализа предложений.

ЗАКЛЮЧЕНИЕ

Рассмотрена модель выбора значения многозначного слова и омонима, основанная на логике предикатов первого порядка. Для повышения качества перевода и разрешения ситуаций, в которых отсутствует возможность однозначного определения значения многозначного слова, предложен статистический алгоритм на базе корпуса русского языка. Определен

критерий оценки корректности найденного значения многозначного слова. Проведен сравнительный анализ разработанного модуля семантического анализа с существующими системами «Яндекс Переводчик» и «Google Переводчик». Сделан вывод, что разработанная система семантического анализа проводит более точный перевод при наличии лексической неоднозначности. Разработанная предикатная модель с учетом статистического алгоритма позволяет в сравнении с семантическим словарем В.А. Тузова решить проблему выбора лексического значения в случае, когда истинны условия для двух и более альтернативных значений. В данной ситуации определяется лучшее сочетание лексических значений по критерию близости к корпусу русского языка.

СПИСОК ЛИТЕРАТУРЫ

1. *Veale T., Conway A.* Cross modal comprehension in ZARDOZ an English to sign-language translation system // *Proceedings of the Seventh International Workshop on Natural Language Generation*. – Kennebunkport, Maine, 1994. – P. 249–252.
2. *Andre E.* The generation of multimedia presentations // *Handbook of Natural Language Processing*. – New York: Marcel Dekker, 2000. – Ch. 12. – P. 305–328.
3. *Wakefield M.* VisiCAST Milestone: final report N IST-1999-10500 / *Information Societies Technology*. – [S. l.], 10 December 2002. – 97 p.
4. Многоцелевой лингвистический процессор ЭТАП-3 [Электронный ресурс]. – URL: <http://iitp.ru/ru/researchlabs/922.htm> (дата обращения: 22.03.2019).
5. ABBYY Intelligent Search SDK [Электронный ресурс]. – URL: <https://www.abbyy.com/ru-ru/isearch/compreno/> (дата обращения: 22.03.2019).
6. TEXTERRA. Технология автоматического построения онтологий и семантического анализа текста [Электронный ресурс]. – URL: <http://www.ispras.ru/technologies/texterra/> (дата обращения: 22.03.2019).
7. Яндекс Переводчик. О сервисе [Электронный ресурс]. – URL: <https://translate.yandex.ru/about> (дата обращения: 22.03.2019).
8. Яндекс. Технологии. Машинный перевод [Электронный ресурс]. – URL: <https://yandex.ru/company/technologies/translation/> (дата обращения: 22.03.2019).
9. About Google Translate [Electronic resource]. – URL: <https://translate.google.com/intl/en/about/> (accessed: 22.03.2019).
10. *Baldwin R.* Google introduces Google Translate Chrome Extension for inline translations of text [Electronic resource] // *The Next Web*. – October 16, 2014. – URL: <https://thenextweb.com/google/2014/10/16/google-introduces-google-translate-chrome-extension-inline-translations-text/> (accessed: 22.03.2019).
11. Franz Och, Ph.D., expert in machine learning and machine translation: press release. – La Jolla, CA: Human Longevity, 2014. – 2 p.
12. *Och F.* Statistical machine translation: foundations and recent advances: tutorial at MT Summit 2005 Phuket, Thailand, September 12, 2005 / Google. – Phuket, 2005. – P. 1–4.
13. *Tanner A.* Google seeks world of instant translations [Electronic resource] // *Thomson Reuters*. – 2007. – March 28. – URL: <https://www.reuters.com/article/us-google-translate/google-seeks-world-of-instant-translations-idUSN1921881520070328> (accessed: 22.03.2019).
14. *Le Q., Schuster M.* A neural network for machine translation, at production scale [Electronic resource] // *Google Research Blog*. – 2016. – September 27. – URL: <https://research.googleblog.com/2016/09/a-neural-network-for-machine.html> (accessed: 22.03.2019).
15. MT on and for the Web / C. Boitet, H. Blanchon, M. Seligman, V. Bellynck // *Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE-2010)*. – IEEE, 2010. – P. 1–10.

16. Разработка словаря сибирского диалекта русского жестового языка: учебное пособие / А.А. Бертик, М.Г. Гриф, С.В. Елфимова, О.О. Королькова. – Новосибирск: Изд-во НГТУ, 2013. – 63 с.

17. Сокирко А.В. Семантические словари в автоматической обработке текста (по материалам системы ДИАЛИНГ): дис. ... канд. техн. наук: 05.12.17. – М., 2001. – 88 с.

18. Ткачева Т.А., Ахмедышева Н.А. Жестовый язык глухих как особая знаковая коммуникативная система (на материале русского и французского жестовых языков) // Приоритеты и интересы современного общества: материалы Международной научно-практической конференции. – Астрахань, 2010. – С. 361–365.

19. Тузов В.А. Компьютерная семантика русского языка: учебное пособие. – СПб.: СПбГУ, 2003. – 391 с.

20. Гриф М.Г., Мануева Ю.С., Козлов А.Н. Построение системы компьютерного сурдоперевода русского языка // Труды СПИИРАН. – 2014. – Вып. 6 (37). – С. 170–183. – doi: <http://dx.doi.org/10.15622/sp.37.11>.

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

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

23. 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 // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. – 2018. – Т. 16, № 3. – С. 32–44.

24. Воскресенский А.Л. Сопоставительное лексикографическое описание слов русского языка и жестов языка глухих России в словаре RuSLED // Компьютерная лингвистика и интеллектуальные технологии. – 2008. – № 7 (14). – С. 91–96.

25. Мельчук И.А. Русский язык в модели Смысл \Leftrightarrow Текст: учебное пособие. – М.: Языки русской культуры, 1995. – 682 с.

Кагиров Ильдар Амирович – младший научный сотрудник, аспирант лаборатории речевых и мультимодальных интерфейсов Санкт-Петербургского института информатики и автоматизации РАН. Область научных интересов: речевые и мультимодальные интерфейсы. Опубликовано 6 научных работ. E-mail: kagirov@iiias.spb.su

Мануева Юлия Сергеевна – аспирант кафедры автоматизированных систем управления факультета автоматики и вычислительной техники Новосибирского государственного технического университета. Область научных интересов – компьютерный сурдопереводчик. Опубликовано 21 научная работа. E-mail: manueva.js@gmail.com

Kagirov Ildar Amirovich, a junior research worker, a postgraduate student in the laboratory of speech and multimodal interfaces in the St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences. His research interests include speech and multimodal interfaces. He has published 6 research papers. E-mail: kagirov@iiias.spb.su

Manueva Yulia Sergeevna, a postgraduate student at the department of automation and computer engineering in the Novosibirsk State Technical University. His research interests include computer sign language interpretation. She has published 21 research papers. E-mail: manueva.js@gmail.com

DOI: 10.17212/1814-1196-2019-1-41-60

***Development of a predicate model for choosing lexical meanings
when translating into the Russian sign language based
on the V.A. Tuzov's semantic dictionary and the Dialing
text processing systems ****

I.A. KAGIROV^{1,a}, J.S. MANUEVA^{2,b}

¹ St. Petersburg Institute for Informatics and Automation RAS, 39, 14th Line of Vasilyevsky Island, St. Petersburg, 1199178, Russian Federation

² Novosibirsk State Technical University, 20, K. Marx Prospect, Novosibirsk, 630073, Russian Federation

^a kagirov@iiias.spb.su ^b manueva.js@gmail.com

Abstract

Existing systems of semantic analysis of the Russian language are considered: multipurpose linguistic processor "ETAP-3", ABBYY Compreno technology, Texterra technology, statistical and neural network approaches in Yandex Translator and Google Translator. Based on the analysis of the advantages and disadvantages of these systems, it was concluded that it was necessary to develop a semantic analysis system for improving the quality of interpretation and introducing interpretation from the sounding Russian language into the Russian sign language into the system called "Surdofon". The purpose of this article is to describe the developed predicate model of semantic analysis of the Sign Language Interpretation computer system based on V.A. Tuzov's semantic dictionary. To determine the correct meaning of a polysemantic word and a homonym in a sentence the "Dialing" system, V.A. Tuzov's dictionary and a list of gestures of the Russian sign language were used. A model for choosing the meaning of a polysemantic word and a homonym based on the logic of first-order predicates is considered. When performing a semantic analysis of the sentence, the meaning of trueMeaning of the word is considered correct if the predicate P (word, trueMeaning) has the meaning "True". To improve the interpretation quality and resolution of situations in which there is no possibility of unambiguous determining the meaning of a polysemantic word, a statistical algorithm is proposed based on of the Russian language corpus. The criterion for evaluating the correctness of the obtained meaning of a polysemantic word is found. A comparative analysis of the developed module with the existing Yandex Translator and Google Translator systems was made. It is concluded that the developed semantic analysis system gives a more accurate interpretation in the presence of lexical ambiguity. The developed predicate model taking into account the statistical algorithm as compared with the V.A. Tuzov semantic dictionary makes it possible to solve the problem of choosing lexical values in the case when the conditions for two or more alternative meanings are true. In this situation, the best combination of lexical meanings is determined by the criterion proximity to the Russian language corpus.

Keywords: Russian sign language, computer sign interpretation, semantic analysis, predicate model, homonyms, statistical analysis

REFERENCES

1. Veale T., Conway A. Cross modal comprehension in ZARDOZ an English to sign-language translation system. *Proceedings of the Seventh International Workshop on Natural Language Generation*, Kennebunkport, Maine, 1994, pp. 249–252.
2. Andre E. The generation of multimedia presentations. *Handbook of Natural Language Processing*. New York, Marcel Dekker, 2000, ch. 12, pp. 305–328.

* Received 05 September 2018.

3. Wakefield M. *VisiCAST Milestone: final report no. IST-1999-10500*. Information Societies Technology, 10 December 2002. 97 p.
4. *Mnogotselvoi lingvicheskii protsessor ETAP-3* [ETAP-3 multipurpose linguistic processor]. Available at: <http://iitp.ru/ru/researchlabs/922.htm> (accessed 22.03.2019).
5. *ABBYY Intelligent Search SDK*. (In Russian). Available at: <https://www.abbyy.com/ru-ru/isearch/comprono/> (accessed 22.03.2019).
6. *TEXTERRA. Tekhnologiya avtomaticheskogo postroeniya ontologii semanticheskogo analiza teksta* [TEXTERRA. Technology for automatic ontology construction and semantic text analysis]. Available at: <http://www.ispras.ru/technologies/texterra/> (accessed 22.03.2019).
7. *Yandex Perevodchik. O servise* (Yandex Translator about the service) <https://translate.yandex.ru/about> (accessed 22.03.2019).
8. *Yandex. Tekhnologii. Mashinnyi perevod* [Yandex technology. Machine translation]. Available at: <https://yandex.ru/company/technologies/translation/> (accessed 22.03.2019).
9. *About Google Translate*. Available at: <https://translate.google.com/intl/en/about/> (accessed 22.03.2019).
10. Baldwin R. Google introduces Google Translate Chrome Extension for inline translations of text. *The Next Web*, October 16, 2014. Available at: <https://thenextweb.com/google/2014/10/16/google-introduces-google-translate-chrome-extension-inline-translations-text/> (accessed 22.03.2019).
11. Franz Och, Ph.D., *expert in machine learning and machine translation*: press release. La Jolla, CA, Human Longevity, 2014. 2 p.
12. Och F. *Statistical machine translation: foundations and recent advances*: tutorial at MT Summit 2005 Phuket, Thailand, September 12, pp. 1–4.
13. Tanner A. Google seeks world of instant translations. *Thomson Reuters*, 2007, March 28. Available at: <https://www.reuters.com/article/us-google-translate/google-seeks-world-of-instant-translations-idUSN1921881520070328> (accessed 22.03.2019).
14. Le Q., Schuster M. A neural network for machine translation, at production scale. *Google Research Blog*, 2016, September 27. Available at: <https://research.googleblog.com/2016/09/a-neural-network-for-machine.html> (accessed 22.03.2019).
15. Boitet C., Blanchon H., Seligman M., Bellynck V. MT on and for the Web. *Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE-2010)*, IEEE, 2010, pp. 1–10.
16. Bertik A.A., Grif M.G., Elfimova S.V., Korolkova O.O. *Razrabotka slovyarja sibirskogo dialekta russkogo zhestovogo yazyka* [Development of the dictionary of the Siberian dialect of Russian sign language]. Novosibirsk, NSTU Publ., 2013. 63 p.
17. Sokirko A.V. *Semanticheskie slovari v avtomaticheskoi obrabotke teksta (po materialam sistemy DIALING)*. Diss. kand. tekhn. nauk [Semantic dictionary in automated text processing. PhD eng. sci. diss.]. Moscow, 2000. 88 p.
18. Tkacheva T.A., Akhmedysheva N.A. [The sign language of the deaf as a special sign communication system (based on Russian and French sign languages)]. *Prioritety i interesy sovremenogo obshchestva: materialy Mezhdunarodnoi nauchno-prakticheskoi konferentsii* [Materials of the International scientific and practical conference "Priorities and Interests modern society". Astrakhan, 2010, pp. 361–365. (In Russian).
19. Tuzov V.A. *Komp'yuternaya semantika russkogo yazyka* [Computer semantic of Russian language]. St. Petersburg, St. Petersburg University Publ., 2003. 391 p.
20. Grif M.G., Manueva Yu.S., Kozlov A.N. Postroenie sistemy komp'yuternogo surdoperevoda russkogo yazyka [Computer sign language interpretation system development of Russian language]. *Trudy SPIIRAN – SPIIRAS Proceedings*, 2014, iss. 6 (37), pp. 170–183. doi: <http://dx.doi.org/10.15622/sp.37.11>.
21. Grif M.G., Timofeeva M.K. Problema avtomatizatsii surdoperevoda s pozitsii prikladnoi lingvistiki [Sign language machine translation as a problem of applied linguistics]. *Sibirskii filologicheskii zhurnal – Siberian Philological Journal*, 2012, no. 1, pp. 211–219.
22. Grif M.G. *Metody i tekhnologii komp'yuternogo surdoperevoda* [Methods and techniques of computer sign language]. Novosibirsk, NSTU Publ., 2012. 71 p.

23. Grif M.G., Korolkova O.O., Manueva Yu.S. Razrabotka algoritmicheskogo i programmnogo obespecheniya vybora znacheniya mnogoznachnogo slova i omonima v sisteme komp'yuternogo surdoperevoda russkogo yazyka na osnove semanticheskoi modeli [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 Novosibirskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya – Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2018, vol. 16, no. 3, pp. 32–44.

24. Voskresenskii A.L. Sopostavitel'noe leksikograficheskoe opisanie slov russkogo yazyka i zhestov yazyka glukhikh Rossii v slovare RuSLED [Comparative lexicographic description of Russian words and gestures of Russian sign language the deaf language in RuSLED dictionary]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii – Computational linguistics and intellectual technologies*, 2008, no. 7 (14), pp. 91–96.

25. Mel'chuk I.A. *Russkii yazyk v modeli Smysl <=> Tekst* [Russian language in the model Meaning <=> Text]. Moscow, Yazyki russkoi kul'tury Publ., 1995. 682 p.

Для цитирования:

Кагиров И.А., Мануева Ю.С. Разработка предикатной модели выбора лексических значений при переводе на русский жестовый язык на основе семантического словаря В.А. Тузова и системы автоматической обработки текста «Диалинг» // Научный вестник НГТУ. – 2019. – № 1 (74). – С. 41–60. – DOI: 10.17212/1814-1196-2019-1-41-60.

For citation:

Kagirov I.A., Manueva Yu.S. Razrabotka predikatnoi modeli vybora leksicheskikh znachenii pri perevode na russkii zhestovyi yazyk na osnove semanticheskogo slovary V.A. Tuzova i sistemy avtomaticheskoi obrabotki teksta "Dialing" [Development of the predicate model for choosing lexical meanings when translating into Russian sign language based on V.A. Tuzov's semantic dictionary and text processing systems "Dialing"]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2019, no. 1 (74), pp. 41–60. DOI: 10.17212/1814-1196-2019-1-41-60.