DOI: 10.15514/ISPRAS-2023-35(6)-22

Решение проблемы многоязычия в международном научно-техническом информационном пространстве

[1] *К.К. Колин*, ORCID: 0000-0002-8187-9314 <kolinkk@mail.ru>
[1,2] *А.А. Хорошилов*, ORCID: 0000-0003-4885-3232 <khoroshilov@mail.ru>
[2,3] *А.В. Кан,* ORCID: 0000–0001–9410–406X <kanav@nrczh.ru>
[1] *Ю.В. Никитин,* ORCID: 0000-0002-7641-0247 <yuri.v.nikitin@gmail.com>

[1] *Федеральный исследовательский центр "Информатика и управление" РАН, 119333, Россия, г. Москва, Вавилова, д.44, кор.2.*
[2] *Национальный исследовательский центр «Институт имени Н.Е. Жуковского», 125319, Россия, г. Москва, ул. Викторенко, д.7.*
[3] *Московский авиационный институт (национальный исследовательский университет), 125993, Россия, г. Москва, Волоколамское шоссе, д. 4.*

Аннотация. Решение проблемы многоязычия в международном научно-техническом информационном пространстве связано с технологиями машинного перевода (МП). Процесс перевода в рамках концепции фразеологического концептуального перевода текстов можно представить, как процесс передачи смыслового содержания исходного текста, средствами выходного языка. В рамках этой концепции перевод текстов обеспечивается понятийным анализом исходного текста и трансформацией его смыслового содержания на целевой язык. Этот подход базируется на понимании закономерностей функционирования естественных языков и теоретических преставлений о смысловой структуре текстов. Основой подхода являются технологии автоматизированного формирования тематических словарных баз, адекватно отражающих понятийный состав различных тематических областей.

Ключевые слова: проблемы многоязычия в международном научно-техническом информационном пространстве; машинный перевод; естественный язык; машинные грамматики; семантико-синтаксический и концептуальный анализ текстов; фразеологический концептуальный перевод текстов.

Для цитирования: Колин К.К., Хорошилов А.А., Кан А.В., Никитин Ю.В. Решение проблемы многоязычия в международном научно-техническом информационном пространстве. Труды ИСП РАН, том 35, вып. 6, 2023 г., стр. 337–346. DOI: 10.15514/ISPRAS–2023–35(6)–22.

Solving the Multilingualism Problem in the International Scientific and Technical Information Space

[1] *K.K. Kolin*, ORCID: : 0000-0002-8187-9314 <kolinkk@mail.ru>
[1,2] *A.A. Khoroshilov*, ORCID: 0000-0003-4885-3232 <khoroshilov@mail.ru>
[2,3] *A.V. Kan*, ORCID: 0000–0001–9410–406X <kanav@nrczh.ru>
[1] *Yu.V. Nikitin*, ORCID: 0000-0002-7641-0247 <yuri.v.nikitin@gmail.com>

[1] *Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44 building 2, Vavilova st., Moscow, 119333, Russia.*
[2] *Analytical Department of the National Research Center "Zhukovsky Institute", 7, Viktorenko st., Moscow, 125319, Russia.*
[3] *Moscow Aviation Institute (National Research University), 4, Volokolamskoye Shosse, Moscow, 125993, Russia.*

Abstract. The solution to the problem of multilingualism in the international scientific and technical information space is connected with machine translation (MP) technologies. The translation process within the framework of the concept of phraseological conceptual translation of texts can be represented as the process of transmitting the semantic content of the source text by means of the output language. Within the framework of this concept, the translation of texts is provided by the conceptual analysis of the source text and the transformation of its semantic content into the target language. This approach is based on understanding the laws of the functioning of natural languages and theoretical ideas about the semantic structure of texts. The basis of the approach is the technology of automated formation of thematic dictionary bases that adequately reflect the conceptual composition of various thematic areas.

Keywords: problems of multilingualism in the international scientific and technical information space, machine translation, natural language, machine grammars, semantic-syntactic and conceptual analysis of texts, phraseological conceptual translation of texts.

For citation: Kolin K.K., Khoroshilov A.A., Kan A.V., Nikitin Yu.V. Solving the multilingualism problem in the international scientific and technical information space. Trudy ISP RAN/Proc. ISP RAS, vol. 35, issue 6, 2023. pp. 337-246 (in Russian). DOI: 10.15514/ISPRAS-2023-35(6)-22.

1. Введение

Современное международное информационное пространство быстро развивается в результате все более широкого распространения компьютерных телекоммуникаций, которые сегодня охватывают все страны мира и становятся неотъемлемой частью их культуры, научно-технологической и социально-экономической деятельности [1]. При этом особую важность приобретает научно-техническая информация, которая содержит сведения о новых достижениях в области науки и технологий, здравоохранения, организации общественного производства, а также о методах противодействия новым вызовам и угрозам XXI века.

Серьезная лингвистическая проблема использования такой информации специалистами различных стран состоит в том, что она, как правило, содержит большое количество специальных терминов, требующих адекватного перевода. А этого современные средства перевода текстов в необходимой степени еще не обеспечивают. Поэтому проблема повышения качества перевода текстов научно-технической информации и является той актуальной и стратегически важной проблемой, без решения которой эффективное использование передовых достижений научно-технического прогресса и международное научно-техническое сотрудничество практически невозможно.

Необходимо отметить, что попытки решения этой проблемы предпринимались неоднократно, начиная с середины минувшего века, когда появились средства вычислительной техники, и продолжаются до сих пор. Однако полученные в них результаты еще нельзя признать удовлетворительными. Наглядным примером здесь может служить современное состояние этой проблемы в странах Европейского экономического союза, для

которых систему высококачественного автоматизированного перевода текстов создать пока еще не удалось [2].

Аналогичная проблема существует и в странах Евразийского экономического союза, а также в странах, которые являются членами БРИКС, ШОС и СНГ. Причем, здесь она осложняется еще и существенным различием алфавитов, на которых представлена текстовая информация. Так, например, в Китае используются иероглифы, в Индии – слоговое письмо, а в других странах этих новых объединений государств – латиница и кириллица. Все это привело к созданию так называемого языкового барьера при доступе к разноязычной информации. В настоящее время преодоление этого барьера осуществляется с помощью систем машинного перевода.

В настоящей работе показано состояние этой проблемы в России и возможности использования уже полученных результатов в интересах развития многоязычного международного пространства научно-технической информации.

2. Современное состояние проблемы машинного перевода текстов

Анализ подходов к решению проблемы машинного перевода, приведенный в работе [3] показал, что в настоящее все большую популярность завоевывают системы машинного перевода, базирующиеся на нейросетевых подходах (Neural Machine Translation, NMT) [4-6]. Так, не выдержав конкуренции с современными системами NMT, с IT-рынка ушли традиционные системы на основе правил (Rule-based Machine Translation, RBMT) [7,8,9], а также системы статистического перевода (Statistical based machine translation, SMT)[10]. Действительно, NMT системы, примером которых может служить портальный сервис перевода Google Translate, показывают удовлетворительные результаты при переводе текстов по некоторым тематическим областям в ряде направлений перевода. В основу этих систем положена модель глубокого обучения, представляющая собой нейросеть, состоящую из ряда слоев, на каждом из которых производится извлечение большого числа признаков из текстовых данных, необходимых для реализации процесса перевода аналогичных текстовых ситуаций.

Широкому применению этих технологий способствовало наличие большого числа свободно распространяемых программных средств, обеспечивающих обучения моделей перевода, а также возможности их реализации и встраивания в любые технологические процессы. Для обучения модели необходимы только исходные данные, представляющие собой тексты, фрагментированные на предложения на исходном языке, и переводы этих предложений на целевом языке.

Обычно для создания NMT-переводчика достаточно только «дообучить» одну из доступных предобученных моделей и использовать программные средства, обеспечивающие реализацию процесса перевода. При этом полностью исключаются трудоемкие работы по создания языковых моделей и технологий трансформации текстов с одного ЕЯ на другой. Но при этом всегда нужно было принимать во внимание, что обученная модель реализует только *те возможности, которые были заложены в нее в процессе обучения*.

Казалось бы, замечательная технология – без значительных затрат решает все проблемы перевода с любых языков по любой тематике. Но, как оказалось при ближайшем рассмотрении, эти технологии, изначально при их разработке, потребовали огромные финансовые и интеллектуальные ресурсы, посильные лишь таким крупным транснациональным компаниям, как Google, IBM и др. Поэтому необходимо понимать, что, если возникнет необходимость создать «с нуля» новое направление перевода по новой тематической области, то, даже при наличии соответствующего программного обеспечения, потребуется специализированное дорогостоящее оборудование и огромные массивы размеченных двуязычных параллельных текстов. В случае, если не окажется в наличии одного из этих компонентов, то этой технологией невозможно будет воспользоваться.

Другими словами, технологии NMT базируются на принципе *перевода последовательности слов на основе аналогии с ранее выполненными переводами*, которые в явном виде не предполагает реализацию основополагающего принципа профессионального перевода: *передачу смыслового содержания исходного текста средствами другого языка*.

А как быть с учетом ряда таких сложных явлений ЕЯ, как вариативность представления смысла в текстах, явления синонимии, гипонимии в ЕЯ? Необходим также учет явления пресуппозиции и ряда других аномальных ситуаций в языке и речи, которые нужно учитывать при переводе текстов с одного ЕЯ на другой.

Таким образом, как не хотелось бы сэкономить на качественном переводе и обойтись только дешевыми технологиями NMT, без трудозатратных работ по созданию полноценных языковых моделей и адекватных двуязычных фразеологических и терминологических словарей, качественного терминологически обоснованного перевода научно-технических текстов получить невозможно.

Однако, многих из выше перечисленных проблем удалось избежать в технологии *фразеологического машинного перевода*, которая, по сути, является дальнейшим развитием традиционного лингвистического перевода, основанного на правилах, но полностью переосмыслена и дополнена рядом гибких языковых моделей, которые базируются на принципе лингвистической аналогии с возможностью значительного расширения их признакового пространства, необходимого для решения задач смыслового анализа текстов.

3. Концепция фразеологического машинного перевода

Впервые возможность получения высококачественного перевода научно-технических текстов предложил и обосновал военный ученый профессор Г.Г. Белоногов[1] в 1975 году в рамках ***концепции фразеологического машинного перевода*** (Phraseological Machine Translation, PMT) [11-12]. В основу этой концепции было положено понимание того факта, что в процессе перевода необходимо выявить смысловое содержание исходного текста и передать его фразеологическими конструкциями того языка, на который переводится текст.

В рамках этой концепции, реализовывалась возможность моделирования деятельности человека-переводчика при переводе текстов. Она включает процедуры восприятия смыслового содержание переводимого текста (анализ смыслового содержания текста) и «пересказывания» содержания этого текста с помощью той понятийной терминологической системы, к тематике которой относился переводимый текст. При этом осуществляется преобразование понятийной структуры исходного текста в понятийную структуру на целевом языке и выполняется порождение осмысленного терминологически и грамматически связанного текста.

В концепции PMT в качестве основных единиц смысла используются *фразеологические словосочетания, выражающие понятия*. Ведь именно понятия являются теми элементарными мыслительными образами, используя которые можно строить более сложные мыслительные образы, соответствующие переводимому тексту. Поэтому перспективные системы МП должны переводить *не слова и их последовательности, а мыслительные образы, выраженные словами и словосочетаниями* [11-12].

Проф. Г.Г. Белоногов в концепции PMT связал воедино иерархию смысловых единиц текста в формализованную многослойную двуязычную модель текста. По его мнению, иерархия понятийной системы текста составляет основу сложного мыслительного образа всего текста. А фразеологические понятия являются теми базовыми «строительными блоками», на основе

[1] Проф. Белоногов Г.Г. – известный советский и российский ученый в области информатики, компьютерной лингвистики и машинного перевода, доктор технических наук, профессор, академик Международной академии информационных процессов и технологий, один из основоположников отечественной информатики, признанный как у нас в стране, так и за рубежом.

которых формируются смысловые единицы более высоких уровней – предложения, сверхфразовые единства, входящие с состав текста [11-12].

При этом формальным инвариантом смысловой структуры предложения является его предикатно-актантная структура (ПАС). Ее компонентами служат понятия-предикаты (признаки и отношения) и понятия-актанты, выступающие в роли описываемых объектов. Использование в процессе машинного перевода моделей ПАС предложений в качестве смысловых единиц обеспечивает возможность адекватной передачи смыслового содержания исходного текста терминологическими и фразеологическими конструкциями на целевом языке.

В работах [11-12] показано, что, согласно этой концепции, «*...система РМТ должна включать в свой состав многоязычную понятийную базу, содержащую переводные эквиваленты часто встречающихся терминологических словосочетаний, фрагментов фраз, служебных конструкций и отдельных слов, средства анализа смысловой структуры исходного текста, средства формирования понятийной структуры на целевом (языке перевода) и средства порождения (генерации) текста на целевом языке*».

В процессе перевода текстов система использует хранящиеся в этой базе переводные эквиваленты в следующем порядке: сначала для очередного предложения исходного текста делается попытка перевести его как целостную фразеологическую единицу. Далее, в случае неудачи – входящие в его состав наиболее длинные синтаксические конструкция, при их отсутствии – более короткими словосочетаниями, и, наконец, осуществляется пословный перевод тех фрагментов предложения, которые не удалось перевести первыми тремя способами. Фрагменты выходного текста, полученные всеми рассмотренными способами, должны грамматически согласовываться друг с другом (с помощью процедур морфологического и синтаксического синтеза) [11-12].

Нужно также отметить, что проф. Белоногов еще рубеже 50-60 гг. прошлого века разработал уникальную машинную грамматику русского языка, ориентированную на широкое применение принципов лингвистической аналогии при реализации , жесткое соответствие между формой представления слов и их грамматической информацией позволило создать на этой основе новые классы – «*…классы слов, имеющие одинаковые наборы грамматических признаков, соответствующие их формам представления в сходных контекстных окружениях...*»[13].

Идея создания новых классов слов, ориентированных на схожесть грамматических признаков слов и схожесть их синтаксических функций в предложении, была впервые предложена в работе [14] для разрешения грамматической омонимии английских слов. Для решения этой задачи и ряда аналогичных задач, была разработана универсальная центроидно-контекстная языковая модель (ЦКЯМ), основная идея которой заключается *в возможности однозначного выявления свойств модели по ее контекстному окружению*.

Каждая языковая модель использует определенное признаковое пространство, в пределах которого и разрешается конкретная языковая ситуация. В частности, для разрешения грамматической омонимии английских слов использовался набор грамматических характеристик английских слов в конкретных текстовых ситуациях. Решением каждой ситуации было однозначное определение грамматических характеристик слов в конкретном контексте, а также и в значительном числе аналогичных контекстов.

В синтаксической модели текстов на основе использования механизма обобщенных синтагм была разработана иерархия синтаксических конструкций предложений, обеспечивающая при синтаксическом анализе возможность адекватного построения синтаксических структур предложений любой сложности. Эта модель решает задачу построения синтаксической структуры предложения в рамках следующего утверждения: «*...представление синтаксической структуры текстов в виде последовательности контактно расположенных двухбайтовых индексов обобщенных синтагм, обладающих*

грамматическими свойствами конкретных слов-эталонов, позволяет фиксировать грамматические и синтаксические свойства различных отрезков реальных текстов, а также дает возможность в ряде задач распознавать аналогичные по заданным свойствам отрезки текстов...» [15].

Широкое применение принципа лингвистической аналогии и использование динамических последовательно контекстных и центроидно-контекстных языковых моделей (ЦКЯМ), базирующихся на значительном признаковом пространстве, обеспечило возможность однозначного разрешения сложных текстовых ситуаций на различных этапах анализа текста.

4. Основные модели системы РМТ

Функциональные *модули перевода* в процессе многоступенчатой обработки текстов исходного текстов формируют ряд моделей исходного текста – модели его трансформации и модели порожденного текста на целевом языке.

Морфологическая модель слов исходного текста представляет собой (так же, как и в модели NTM) вектор характеристик слов фиксированной длины. Состав этих характеристик имеет различную природу и четкое распределение между грамматическими и семантическими признаками слов. В модели также обозначены их словообразующие и формообразующие характеристики, а также определен набор формальных характеристик, отображающих смысл слов и характеристики их конкретных форм [14].

Семантико-синтаксическая модель предложений исходного текста представляет собой двухмерную матрицу, в которой позициям токенов предложения поставлены в соответствие расширенные векторы их характеристик. На основе этой матрицы реализован ряд синтаксических моделей, а также осуществляется с помощью центроидно-контекстных моделей извлечение и классификация синтаксических конструкций предложения, определяется их роль в предложении, выявляются именные и глагольные словосочетания и устанавливается их структура.

На завершающем этапе производится выявление смыслового каркаса предложения и установление связей между его элементами. Результатом обработки предложений является формирование нескольких логически связанных формальных представлений: в виде бинарных отношений токенов предложения (дерева зависимостей), в виде смыслового каркаса («скелета») предложения, в виде предикатно-актантной структуры. [15].

Модель концептуального анализа исходного текста в качестве основной задачи ставит выявление текстовой системы понятий и преобразование ее в унифицированное формализованное представление. Также в рамках этой модели устанавливаются парадигматические и синтагматические связи между понятиями. На конечном этапе анализа цифровые характеристики всей иерархии моделей исходного текста преобразуются в многослойную цифровую модель метаданных [12].

Сформированная *модель метаданных исходного текста* обеспечивает возможность передачи информации об анализируемом исходном тексте в трансформационную модель. В *трансформационной модели* основным процессом является механизм вероятностного соотнесения понятийной модели исходного текста с понятийной моделью целевого языка. В процессе этого соотнесения учитываются принадлежность понятий к типу двуязычного словаря и степень покрытия синтаксической структуры предложений исходного текста. При этом предполагается, что большая длина текстовых представлений понятий обеспечивает более точный перевод. Обязательным условие смысловой связанности переведенного текста является включение переводных соответствий в единое понятийное пространство. Использование многофакторной вероятностной модели в процессе соотнесения понятий исходного языка с понятиями целевого языка обеспечивает более точный терминологический перевод текста [12].

Основой этой модели является многомиллионный комплекс двуязычных словарей наименований понятий, охватывающий широкий спектр тематических областей. Этот комплекс словарей включает четыре уровня словарей: политематический, тематический пользовательский и словарь ТМ (Translation Memory – память переводчика). Перевод осуществляется в соответствии с приоритетами словарей в следующем порядке: словарь ТМ, пользовательский словарь, тематический словарь и, наконец, политематический словарь.

Словарь ТМ подключается по желанию пользователя. Приоритет словарей обеспечивает возможность пользователю влиять на процесс перевода, путем оперативного формирования словарных статей или установления приоритетов переводных соответствий, имеющихся в словарной базе.

Завершает процесс перевода *модель порождения текста* на целевом языке. В этой модели производятся необходимые локальные перестановки слов внутри словосочетаний и глобальные перестановки словосочетаний в пределах предложения, а также осуществляется грамматическая трансформация форм слов предложения в соответствие с грамматическим строем целевого языка. Этот комплекс трансформаций текстового представления позволяет сформировать терминологически и грамматически связанный переведенный на целевой язык исходный текст.

5. Основные модели системы РМТ

Основой любой системы перевода является его словарная база, каком бы виде она не была бы представлена. Качество перевода обеспечивается наличием в ней наиболее часто встречающихся представлений текстовых ситуаций и возможностью оперативного пополнения отсутствующих в словаре необходимых представлений ситуаций. Текстовые представления ситуаций в словарных базах представлены различными текстовыми конструкция: отдельными словами, именными и глагольными словосочетаниями, словосочетаниями в контекстных окружениях и целыми предложениями. Система в процессе перевода должна извлекать из словарной базы наиболее адекватные словарные конструкции и располагать их в той последовательности, которая будет соответствовать наиболее точной передачи смыслового содержания текста.

Процесс извлечения наиболее адекватных словарных конструкций производится в модели трансфера – в соответствии с принципами, заложенными в трансформационной модели. Но для полноценного функционирования этого модуля должна быть подготовлена соответствующая двуязычная словарная база.

В концепции ФМТ большое внимание уделено разработке *автоматизированных технологий создания двуязычных словарей*. Основным требованием к этим технологиям является возможность адекватного отображения понятийного состава предметной области в словарной базе системы ФМТ. Другими словами, в тематической словарной базе должен содержаться основной понятийный и терминологический состав отрасли, контекстная система отношений между наименованиями понятий, а также необходимый набор фразеологических конструкций, обеспечивающий связывание переводных соответствий в грамматически согласованный переведенный текст.

В настоящее время авторами разработаны основные технологии создания политематической, тематической и пользовательской словарных баз. Наиболее важной из этих технологий является *технология создания тематической словарной базы*. Исходными данными для ее создания является репрезентативный корпус текстов на языке оригинала по заданной тематике. В монографии [12] детально описан технологический процесс создания такой словарной базы.

Необходимо отметить, что процесс извлечения фразеологического и терминологического понятийного состава тематики производится полностью автоматически точными и предиктивными методами концептуального анализа текстов. В рамках этого процесса

предусмотрены механизмы оценки необходимых трудозатрат для получения такой словарной базы, которая обеспечит требуемое качество перевода отраслевых документов.

Политематическая словарная база формируется путем слияния всех имеющихся тематических словарных баз по строго установленной технологии, учитывающий приоритеты смежных тематик по отношению к заданной тематике. Эта технология позволяет обеспечить формирование единой словарной базы, в которой исключены все дублирующие переводные эквиваленты, а оставшиеся эквиваленты выстроены в соответствии с их приоритетами. В каждой сессии перевода выполняется автоматическое формирование конфигурации виртуальной словарной базы, позволяющей произвести однократный поиск переводных соответствий одновременно как в тематической, так и политематической словарных базах.

Пользовательская словарная база обеспечивает возможность получения высококачественного перевода в режиме диалогового общения пользователя с системой РМТ. Этот режим позволяет пользователю в процессе перевода формировать свой высокоприоритетный пользовательский словарь, в который он может включать любые текстовые конструкции, практически любой длины.

По усмотрению пользователя, часто встречающиеся предложения с их переводами могут им помещаться в *словарь ТМ*. Пользователь также имеет уникальную возможность оперативной коррекции политематических и тематических словарей. И, наконец, пользователю предоставлена возможность оперативной коррекции языковых моделей исходных и целевых ЕЯ.

Необходимо отметить, что отличительной особенностью словарей РМТ является простая структура словарных статей, входами в которые могут быть любые текстовые фрагменты, а в качестве их переводных соответствий - эквивалентные по смыслу переводные соответствия, специфичные для данной предметной области. Причем, в словарях РМТ полностью отсутствует какая-либо сопутствующая грамматическая или семантическая информация, а наименования понятий и их переводные эквиваленты на целевом языке могут быть представлены в любой грамматической форме.

Все эти особенности технологий и структур словарей позволяют в короткие сроки создавать словарные базы больших объемов. Так, например, к настоящему времени создано более 150 таких тематических словарей по различным направлениям развития научно-технического прогресса: авиация, космонавтика, вычислительная техника, ядерная энергетика и т.п. Их общий объем составляет более 7.5 млн. словарных статей.

6. Заключение

В современную эпоху создания многополярного мироустройства особое значение приобретает международное научно-технологическое сотрудничество с дружественными странами - членами БРИКС, ШОС и СНГ. Возникающий при этом языковый барьер можно преодолеть путем разработки современных гибридных технологий машинного перевода, базирующихся как традиционных технологиях машинного перевода, так и на технологиях NMT. Такие гибридные технологии мы применили при разработке программного комплекса «Умный текстовый процессор» в АО «НПК «ВТ и СС», который был удостоен Национальной премии в области информационных технологий «Приоритет: Цифра-2023» в номинации «Искусственный интеллект».

Список литературы / References)

[1]. Колин К. К., Урсул А. Д. Информация и культура. Введение в информационную культурологию. М.: Изд-во Стратегические приоритеты, 2015. – 300 с.

[2]. Колин К. К., Хорошилов А. А. Проблема многоязычия в информационном общества и интеллектуальные переводческие технологии // Информационное общества, 2012, № 1. С. 56-61.

[3]. Искусственный интеллект в технологиях машинного перевода / Колин К.К., Хорошилов Ал-др. А., Никитин Ю.В., Пшеничный С.И., Хорошилов Алексей А. // Социальные новации и социальные науки. – Москва: ИНИОН РАН, 2021. – № 2.
[4]. Johnson M., Schuster M., Le Q. V., et al. Google's multilingual neural machine translation system: Enabling zeroshot translation // T. Assoc. Computational Linguistics, 2017. Vol. 5.
[5]. Хобсон Л., Ханнес Х., Ховард К. Обработка естекственного языка в действие. Изд. Питер, 2020.
[6]. Ганегедара Т. Обработка естественного языка с TensorFlow. М. ДМК Пресс, 2020.
[7]. Мельчук И. А. Опыт теории лингвистических моделей «Смысл <=> текст». «Наука», Москва, 1974 г.
[8]. Кулагина О. С. Исследования по машинному переводу. «Наука», Москва, 1979.
[9]. Пиотровский Р. Г. Новые горизонты машинного перевода. Сб. «Научно-техническая информация», сер. 2, № 1, ВИНИТИ, 2002.
[10]. Белоногов Г. Г., Хорошилов Ал-др А., Хорошилов Ал-сей А., Козачук М. В., Рыжова Е. Ю., Гуськова Л. Ю. Каким быть машинному переводу в XXI веке // Перевод: традиции и современные технологии. – М.: ВЦП, 2002.
[11]. Белоногов Г. Г., Калинин Ю. П., Хорошилов Ал-др А., Хорошилов Ал-ей А. Системы фразеологического машинного перевода. Изд. Русский мир, Москва, 2007.
[12]. Хорошилов Ал-др А., Кан А.В., Хорошилов А.А. Фразеологический машинный перевод. – М.: Изд-во «Директ-Медиа», 2019.
[13]. Аблов И.В., Козичев В.Н., Ширманов А.В., Хорошилов А.А., Хорошилов А.А. Средства машинной грамматики русского языка (по Г.Г. Белоногову) // Сб. "Научно-техническая информация", Серия 2, № 6, ВИНИТИ, 2018.
[14]. Белоногов Г. Г., Калинин Ю. П., Хорошилов А. А. Компьютерная лингвистика и перспективные информационные технологии. Теория и практика построения систем автоматической обработки текстовой информации. Изд. Русский мир, Москва, 2004.
[15]. Кан А. В., Ревина В. Д., Руснак В. И., Хорошилов Александр А., Хорошилов Алексей А. Автоматическое формирование синтаксической модели языка для задач машинного перевода и информационного поиска. Сб. «Научно-техническая информация», Сер. 2, № 12, ВИНИТИ, 2018.
[16]. Захаров В. Н., Никитин Ю. В., Хорошилов Александр А., Хорошилов Алексей А. Технологии создания новых направлений перевода для системы МетаФраз (на примере казахско-русского перевода). Сб. «Научно-техническая информация», Сер. 2, № 9, ВИНИТИ, 2017.
[17]. Хорошилов А.А., Никитин Ю.В., Лазарев А.С. и др. Программный комплекс "Умный текстовый процессор" (ПК УТП). Свидетельство о регистрации программы для ЭВМ RU 2023611992, 26.01.2023. Заявка № 2022683941 от 06.12.2022.

Информация об авторах / Information about authors

Константин Константинович КОЛИН – Доктор технических наук, профессор, главный научный сотрудник Федерального исследовательского центра "Информатика и управление" Российской академии наук. Сфера научных интересов: философия информации, философские и социальные проблемы информатики.

Konstantin Konstantinovich KOLIN – Dr. Sci. (Tech.), Professor, Chief Researcher of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences. Research interests: philosophy of information, philosophical and social problems of computer science.

Александр Алексеевич ХОРОШИЛОВ – Доктор технических наук, профессор Московского авиационного института, ведущий научный сотрудник Федерального исследовательского центра "Информатика и управление" Российской академии наук, старший научный сотрудник 27-го Центрального научно-исследовательского института Министерства обороны России. Сфера научных интересов: системный анализ, машинный перевод и искусственный интеллект.

Alexanser Alexeevich KHOROSHILOV – Dr. Sci. (Tech.), Professor of the Moscow Aviation Institute (National Research University), Lead Researcher of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Senior Researcher of the 27

Central Research Institute of the Ministry of Defense of the Russian Federation. Research interests: system analysis, machine translation and artificial intelligence.

Анна Владимировна КАН – кандидат технических наук, доцент Московского авиационного института, начальник аналитического отдела Научно-исследовательского центра “Институт имени Н. Е. Жуковского”. Сфера научных интересов: системный анализ, имитационное моделирование и искусственный интеллект.

Anna Vladimirovna KAN – Cand. Sci. (Tech.), Associate Professor of the Moscow Aviation Institute, Head of the Analytical Department of the National Research Center “Zhukovsky Institute”. Research interests: system analysis, simulation and artificial intelligence.

Юрий Викторович НИКИТИН – научный сотрудник Федерального исследовательского центра “Информатика и управление” Российской академии наук, руководитель группы разработчиков Научно-Промышленной компании “Высокие технологии и стратегические системы”. Сфера научных интересов: компьютерная лингвистика, технологии автоматической обработки и семантического анализа текстов.

Yuri Viktorovich NIKITIN – Researcher of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Development Team Leader of the Scientific and Industrial Company “High Technologies and Strategic Systems”. Research interests: computational linguistics, technologies of automatic processing and semantic analysis of texts.