

当代信息空间中的语言资源

Л.Н. 别利亚耶娃

(俄罗斯赫尔岑国立师范大学语文系, 圣彼得堡 191186)

提 要: 本文主要讨论在术语、词典和翻译工作中语用信息手段和语言工具的问题, 共包括 5 个部分。论文第 1 部分指出翻译与信息技术、语言资源的关系, 认为自动化操作平台能提高翻译工作的效率, 并使各种资源和工作得以协调; 第 2 部分重点论述自动化操作平台的组成及其在语言信息处理过程中的作用; 第 3 部分阐述机器翻译系统如何服务语言信息处理; 第 4 部分介绍认知术语学中知识表征的工具, 如叙词表、词汇语义网络等, 重点讨论知识本体及其与其他表征手段的区别; 第 5 部分论述运用数据库服务术语工作的问题。

关键词: 语言技术; 自动工作平台; 术语抽取; 应用词典学

中图分类号: H083

文献标识码: A

В условиях постоянно возрастающей роли различных видов коммуникации и актуализации новых информационных технологий усиливается необходимость дифференциации видов информации и учета различных путей и способов ее извлечения, обработки, оценки и использования. Необходимость использования многоязычной информации на всех этапах ее существования требует применения информационных технологий (ИТ) для создания систем ее поиска, генерации и поддержки, для локализации данных и программного обеспечения. Информационные технологии в области естественного языка (лингвистические технологии), реализующие алгоритмы автоматической переработки текста, являются необходимым условием решения множества актуальных задач, связанных с извлечением и систематизацией знаний. В этом аспекте применение лингвистических технологий в современном прикладном терминоведении и техническом переводе требует согласованности и взаимного обогащения.

1. Перевод, информационные технологии, лингвистические ресурсы

В наше время потребность в точном и быстром переводе с одного языка на другой стала абсолютно ясной. Но, к сожалению, несмотря на многие годы страстных предварительных обсуждений, научно-техническое сообщество в целом и переводчики в особенности абсолютно не готовы перерабатывать огромные потоки информации на различных языках ни технически, ни психологически. В то же время уже давно стало очевидно, что только использование компьютера при обработке текстов и потоков документа на различных языках может позволить специалисту искать, переводить и использовать необходимую информацию.

Использование лингвистических технологий давно стало элементом профессиональной работы переводчика, лексикографа и терминоведа, а для специалиста — средством извлечения знаний из иноязычного текста. Грамотное использование ресурсов лингвистических технологий: электронных баз данных и знаний, систем машинного перевода (МП), тезаурусов, онтологий, систем проверки орфографии, систем доступа к информации по различным сетям передачи данных — вошло в реальный обиход специалистов в различных областях знаний [Hutchins 1998].

Современный специалист работает сегодня в высокотехнологичной среде, часто в команде (особенно в различных производственных и научных организациях, фирмах и корпорациях), привлекая различные электронные ресурсы и обмениваясь информацией с коллегами. Следовательно, необходима разработка не только структуры такой высокотехнологичной информационной среды, но и средств реализации конкретных функций систем, необходимых для ее поддержки и использования. В частности, различным специалистам, использующим информационные технологии для решения задачи извлечения информации и знаний, необходимо специализированное автоматизированное рабочее место (АРМ), которое можно использовать как комплекс поддержки работы профессионального переводчика, как систему работы с терминологией и как систему обучения [Hutchins 1998; Демич, Костыркин, Шаляпина 2010; Google Translator Toolkit].

В условиях современной высокотехнологичной среды на базе электронных ресурсов должны решаться задачи, возникающие на разных этапах и уровнях исследования, производства и обучения:

— автоматический поиск, извлечение и обогащение информации и знаний, получаемых из различных многоязычных источников, в том числе мультимедийных, и источников, связанных с коммуникацией различных участников;

— межъязыковое или многоязычное извлечение, презентация и распространение информации;

— автоматическое обнаружение и “отслеживание” новой фактографической информации, извлекаемой из неструктурированных мультимедийных данных и поликодовых текстов;

— автоматический анализ источников знаний разметки знаний и обеспечения доступа к ним;

— поддержка вопросно-ответного взаимодействия человека и компьютера для извлечения знаний из источников различной природы, структуры и состава;

— поддержание дистанционного обучения в системах открытого образования;

— создание интеллектуальных средств поддержки автоматизированного ведения библиографической работы, анализа и понимания документов;

— моделирование знаний, потребностей и намерений различных типов пользователей на основе анализа их запросов к различным системам, созданных ими продуктов и взаимодействия с компьютером;

— обеспечение возможности устного диалога с компьютером и поддержки анализа и порождения звучащей речи.

В соответствии с этим набором задач можно выделить несколько видов ресурсов, которые целесообразно включать в любое АРМ в качестве базы информации и знаний, к ним относятся:

— база полнотекстовых корпусов, включающих учебные и научные тексты, извлекаемые из фондов библиотек, электронных в том числе, а также корпуса проблемно-ориентированных текстов, структурированных по предметным областям и языкам для специальных целей;

— база оперативно получаемых и пополняемых переводов текстов, выполняемых конкретным переводчиком или переводческим коллективом;

— многофункциональная и многоязычная база переводных словарей, также оперативно поддерживаемая и обновляемая, жестко ориентированная на разные области знаний и языки для специальных целей;

— база электронных учебников и учебных пособий, образовательных программ и учебно-методических комплектов к ним;

— база средств и систем анализа и синтеза текстов, включающих программные комплексы для автоматического перевода, реферирования, извлечения терминологии и т.д.;

— комплекс систем обучения и тестирования, необходимых для реализации различных образовательных программ, методов контроля и видов самостоятельной работы.

Функции АРМ должны соответствовать потребностям технического переводчика, лексикографа и терминолога, основным особенностям их деятельности и выполняемым операциям по преобразованию текста на исходном языке в текст на языке перевода. Сегодня именно деятельность технического переводчика оказывается базой для оперативного извлечения и дальнейшего анализа информации. Особенно важными при этом являются характеристики скорости выполнения и высокого качества перевода, поскольку перевод, выполненный поздно или с использованием произвольно выбранных терминов, может привести к последствиям если не трагическим, то критическим, особенно в областях, связанных с реальным риском для жизни и здоровья (в медицине, сейсмозащите, атомной энергетике и т.д.). Реализация АРМ позволит организовать эту деятельность максимально эффективно и обеспечить возможность кооперации ресурсов и работ в рамках информационной среды.

2. Автоматизированное рабочее место в структуре информационной среды

Выбранные в каждом конкретном случае функции АРМ определяют наполнение системы поддержки от “простых” терминологических ресурсов (доступа к всевозможным одноязычным и многоязычным учебным пособиям, обучающим системам, словарям и глоссариям, находящимся как в памяти компьютера, так и в сети) до более сложных систем. К таким системам и ресурсам, как уже говорилось выше, относятся системы поиска и обработки информации, машинного перевода, электронные учебники и учебно-методические комплекты, системы специализированных грамматик для различных языков для специальных целей и т.п. Кроме того, АРМ должен включать системы представления данных (инструментальные средства) для возможного издания подготовленной информации.

Грамотное использование автоматизированного рабочего места, включающего комплекс резидентных словарей, тезаурусов, систем проверки орфографии, систем доступа к информации по различным сетям передачи данных, должно войти в обиход не только переводчика, лексикографа, терминолога, но и специалиста в различных (не обязательно гуманитарных) областях знаний [Hutchins 2001]. Собственное АРМ следует научиться компоновать из различных систем обработки информации, подбирать и формировать

системы автоматических и автоматизированных словарей в соответствии со своими запросами и сферой деятельности, точно знать их ограничения и возможности, знать, какие лексикографические источники отсутствуют в электронном формате. Кроме того, необходимо хорошо представлять себе ресурс Интернет и знать, что из него можно извлечь.

Предметная ориентированность информационных ресурсов, составляющих часть информационной среды, определяет целесообразность их иерархизации в случае создания специализированной рабочей или образовательной среды. При такой иерархизации терминальные узлы (словарные базы или корпуса текстов) соответствуют областям знаний, релевантным для конкретного пользователя. В случае включения комплекса АРМ в единое образовательное или производственное пространство иерархическая структура информационных и технологических ресурсов предполагает, что на верхних уровнях (уровнях общего доступа) должны быть максимально универсальные ресурсы, использование которых не связано со специализацией исследования и/или обучения. Эта же иерархическая структура должна реализоваться для каждого конкретного (терминального) ресурса.

Соответственно, например, при единой, выбранной в качестве рабочей системе машинного перевода выделяются конкретные системы автоматических словарей, соответствующие областям знаний. Так, например, в предметной области «Филология» общим словарем верхнего уровня является словарь общефилологических терминов; на следующем уровне выделяются подобласти лингвистики, литературоведения и образовательных технологий в филологии, для каждой из которых должен создаваться общий словарь и система предметно-ориентированных словарей.

Формирование лингвистических ресурсов представляет собой сложную задачу, которая, к сожалению, часто рассматривается как чисто техническая, предполагающая решение собственно технических проблем объединения уже имеющихся терминологических и текстовых ресурсов. Следует учитывать, что организация АРМ в области переработки текстов, создание и использование систем автоматизации этой переработки предполагает предварительное моделирование речемыслительной деятельности, реализуемой в АРМ. Дело в том, что любая модель речемыслительной деятельности, учитывает особенности потенциально возможного речевого акта и отражение в его описании ментальной деятельности человека. Именно изучение особенностей протекания этой деятельности составляет ракурс изучения языка такими направлениями современного языкознания как когнитивная лингвистика, психолингвистика и нейролингвистика. Когнитивная лингвистика является системообразующим центром наук когнитивного цикла, позволяющим рассматривать языковые данные как проявление деятельности познания, при этом языковые данные и объекты и явления, ими номинируемые, рассматриваются как результаты процессов категоризации и концептуализации [Кубрякова 2010: 117]. Исследование особенностей семантики различных категорий лексических единиц оказывается в ряде случаев единственным источником информации о том, как протекает в мозгу процесс когниции. Соответственно, рассматривая потенциал АРМ и его лингвистические ресурсы нельзя не учитывать современные возможности моделирования значения терминов.

Когнитивно-дискурсивная парадигма анализа действительно плодотворна при исследованиях текста и дискурса, результатом которых должно быть создание

компьютерных моделей терминологии и языков для специальных целей. Исследуемые и разрабатываемые в рамках когнитивистики ментальные процедуры категоризации и модели памяти человека представляют собой гипотетический конструкт, реальная верификация которого возможна только на основе нейролингвистических данных, сложно «добываемых» и идентифицируемых.

Направление исследований, зародившееся в рамках когнитивистики и, уже, искусственного интеллекта и выделившееся в самостоятельную дисциплину, связанную с вопросами извлечения, структурирования, формирования, обработки и приобретения знаний, называется сегодня инженерией знаний (*Knowledge Engineering*). Как правило, под извлечением знаний понимают процедуру взаимодействия специалиста с источником знаний (человеком-экспертом, текстовыми документами разного состава, природы и сложности, мультимедийной информацией и т.д.), в результате которой выявляется имплицитная информация текстов и других источников, а также структура предметной области [Гаврилова, Червинская 1992].

Если обратиться к истории типологических представлений, то становление логического этапа, связанного с именами Платона и Аристотеля, определяется принципиальным различием **мнения** и **знания**, при этом знание может относиться только к непреходящим свойствам, а все свойства, меняющиеся со временем, не могут быть предметом знания [Чебанов, Мартыненко 2008: 330]. Сами знания принято делить на: знания процедурные (“знаю, как”) и знания декларативные (знаю, “что”).

В основе формирования системы знаний лежит умение извлекать, анализировать и обрабатывать информацию из различных источников. Появление вычислительной техники, возможности, которые она предоставляет исследователю и хранителю текстовых и материальных объектов, кардинально изменили ситуацию во многих областях, основой деятельности которых является накопление и обработка информации. Поэтому сегодня одним из аспектов культурного и научного взаимодействия является создание национальных фондов знаний в различных областях деятельности человека, фондов, дающих возможность оперативно получать информацию о тех единицах, для хранения которых создан фонд, пополнять и менять эту информацию по мере необходимости.

Без соответствующих средств анализа информации, содержащейся в текстах различного состава, природы и сложности, обнаружение полезного знания, скрытого в огромных объемах необработанных данных, представляет сложнейшую задачу. Экспоненциальный рост текстовых и мультимедийных данных, разнообразный характер как самих данных, так и задач их анализа, сложность анализа смешанных структурированных данных и текстов являются именно теми факторами, которые превращают извлечение информации и формирование знаний в сложную задачу.

Лингвистическое исследование при таком подходе предполагает проведение предварительной лексикографической работы для отбора и описания терминологии на разных языках, осуществление гармонизации этих описаний и согласование терминологических систем разных языков. Подобная система в рамках АРМ должна создаваться как терминологический банк данных, включая не только переводы терминов на различные языки, но и их толкования в авторитетных изданиях.

Соответственно, АРМ переводчика, лексикографа и терминолога реализуется как комплекс средств, в структуру которого должны войти следующие программные и лингвистические модули:

1) лингвистический автомат как комплекс средств автоматизированной переработки текста (выполняющий функции компрессии и поиска информации, перевода и предварительной лексической обработки текстов и т.д.). Эта часть комплекса может непосредственно использоваться в процессе перевода, а также при обучении переводу, редактированию, аннотированию текстов, письменной практике и т.д.

2) база полнотекстовых данных, предполагающая хранение, модификацию и поиск текстов произведений научной и технической литературы на разных языках с формированием массивов параллельных и псевдопараллельных текстов. Эта часть комплекса может непосредственно использоваться для анализа конкретных лингвистических фактов. Кроме того, подобная база является важным источником сведений для создания словарей разного состава и назначения;

3) база терминологических данных, предполагающая хранение, структуризацию, составление тезауруса терминов и поиск терминов в текстах, написанных на языках для специальных целей. Эта база может быть использована в учебных целях для самостоятельной работы студентов при обучении переводу;

4) база референтных и автоматических словарей, словарей на машинных носителях, учебных словарей, объединенных в единый комплекс, позволяющий в рамках АРМ использовать любую накопленную словарную информацию в учебных и научных целях;

5) база специализированных лингвистических программных средств, обеспечивающих получение словарей разных видов. Работа с лингвистическими программами позволит овладеть соответствующими методами, переводчики получат возможности исследования больших массивов текстов.

6) база средств обработки мультимедийной информации, поддерживающей обучение различным аспектам межкультурной коммуникации и речевого поведения.

Важной составляющей АРМ являются автоматические и автоматизированные словари и терминологические базы данных. В принципе, использование информационных технологий в лексикографической работе предполагает ряд возможностей для

1) поддержания создания и ведения словарей:

— решения задач отбора лексических единиц (ЛЕ), их лексикографирования, извлечения информации о лексической единице из проблемно-ориентированного массива текстов;

— создания, редактирования и коррекции оригинал-макета словаря, его последующего издания;

— создания и ведения словников на основе отбора ЛЕ из баз терминологических данных по заданному критерию или набору критериев;

— создания и ведения терминологических баз данных и онтологий.

2) поддержания работы переводчика при обращении к словарям разного типа и на разных носителях:

— извлечения информации из различных лексикографических источников (автоматических, автоматизированных, резидентных словарей);

— исследования лексического состава и динамики лексического спектра конкретного языка/ подъязыка.

Необходимость использовать весь потенциал информационных технологий на этапах создания и ведения современных лексикографических систем определяется сегодня потребностями научного и технического сообщества, задачами исследования терминологии, развитием множества (более 300) [Кияк 2009] языков для профессиональных (специальных) целей, а также потребностями современной лексикографии.

Профессиональная предметно-ориентированная лексикография предполагает введение в структуру АРМ специального технологического комплекса для создания и поддержки словарей и словарных баз данных. Подобный комплекс позволит интегрировать выполнение различных задач, ориентированных на выделение терминов из текстов, получение статистических данных о встречаемости терминов в обрабатываемых текстах, просмотр конкорданса, просмотр терминологических сочетаний по заданным параметрам контекстного окна, автоматическое пополнение словаря, построение онтологий. Эти возможности реализованы, в частности, в автоматизированной лексикографической среде Alex+, с помощью которой анализируется русскоязычная терминология корпусной лингвистики [Митрофанова, Захаров 2009].

Однако анализ современного набора переводных словарей, издаваемых в нашей стране и/или включенных в различные автоматизированные словарные системы, позволяет фиксировать их отставание от современного развития науки и техники, несоответствие основным направлениям развития отраслей знаний. Это связано не только с естественным отставанием лексикографии, связанным с необходимостью переработки больших массивов современной информации, а прежде всего с традиционным подходом к созданию словарей с опорой на уже опубликованные источники, а уже затем на результат анализа текстов.

Использование информационных технологий при таком преимущественно компилятивном подходе ничего не меняет по сути, но только уменьшает трудоемкость работы при сравнении и объединении словарных источников, а также при редактировании сформированного массива. Следовательно, оперативное создание словарей, отражающих картину лексического состава отдельных отраслей знаний, фиксирующих нормативную (рекомендуемую) и реально встречающуюся лексику, представляет собой особую задачу.

На предварительном этапе настройки АРМ должен быть осуществлен выбор основных лингвистических ресурсов: электронных словарей, энциклопедических баз данных, корпусов текстов и средств работы с ними.

Этап поддержки ресурсов АРМ необходим для постоянного ведения системы (добавления и/или уточнения ее состава, введения новых «пользовательских» ресурсов, пополнения корпуса параллельных текстов и т.п.) и настройки ее на задачи конкретного переводчика.

3. АРМ технического переводчика в структуре информационной системы

Структура профессиональной переводческой деятельности за последние 20 лет изменилась кардинально. Так, например, по данным Всероссийского центра переводов научно-технической литературы и документации [Пушнов и др. 2009] в общей структуре выполняемых переводов резко сократился заказ на научно-технические переводы (с 96,2% в 1985 году до 38,2% в 2005). В то же время заказ на бизнес-переводы (договоры, контракты, финансовые и юридические документы) за тот же период вырос с нуля до 22,1%, и большой объем (31,8% в 2005 году) составили социально-бытовые переводы, ранее не заказываемые.

Кроме того, изменилась структура заказываемых Центру научно-технических переводов: основную часть составляют технические и технологические материалы и документация (33% в 2005 году), а заказ переводов научных и научно-технических текстов и литературы с 2000 года сократился до нуля.

Изменение структуры переводов связано, безусловно, с социальными причинами (сокращением возможности финансирования подобной деятельности со стороны организаций), но отсутствие заказов на перевод научной литературы является показателем того, что научное сообщество предпочитает самостоятельно переводить и анализировать необходимую литературу, пользуясь различными средствами информационных и лингвистических технологий, активно используя системы машинного перевода.

Во всем мире использование систем машинного перевода, несмотря на все их слабые стороны, давно стало элементом профессиональной работы переводчика. Системы машинного перевода можно рассматривать как модели переводческого процесса. В модели переводческого процесса в рамках коммуникативно-функционального подхода в этот процесс вовлекаются: анализ коммуникативной ситуации, при которой переводчик последовательно выступает в функциях адресата и адресанта, анализ прагматики автора и его установки, моделирование возможной интерпретации, учет уровня и компетенции адресата, анализ типа и функции текста и т.д. [Bell 1991; Сдобников 2009: 24-25; Новиков 2007: 18; Синявская-Суйковска 2009: 41].

Можно утверждать, что при автоматическом анализе текста выявляется именно предметная ситуация, описываемая текстом, то есть реализуется ситуативно-денотативная модель перевода [Ревзин, Розенцвейг 1964], собственно коммуникативная ситуация и уровень реципиента в этом случае должны учитываться при постредактировании [Гавриленко 2011].

Таким образом, автоматизированное рабочее место переводчика, как правило, включает систему машинного перевода с настроенными пользовательскими словарями, средства переводческой памяти, онлайн-словари, доступные практические средства автоматизации извлечения терминологии, средства проверки текстов на соответствие требованиям контролируемого языка [Steinberger 1998: 5], средства формирования моделируемых текстов определенной структуры [Rychtuckyj 2002: 208]. В переводческих системах, создаваемых в больших производственных компаниях, используется описание специально создаваемого контролируемого языка и средства работы с ним.

Контролируемые языки, активно используемые в настоящее время там, где требуется оперативная переработка больших массивов информации на разных языках, являются языками институциональными, относятся к классу искусственных языков. Контролируемые языки (*controlled languages*) используются для подготовки текстовых материалов определенной жесткой структуры, а также для улучшения результатов перевода (в основном машинного), поскольку принятые в них ограничения словаря и грамматики позволяют не только повысить эквивалентность результатов ручного перевода и эффективность систем МП, но и облегчают выравнивание в корпусах параллельных текстов. Контролируемый язык (КЯ) должен повышать понятность текста, особенно в ситуации научно-технической коммуникации [Muegge 2009]. Использование контролируемого языка при создании текста или при его предварительном редактировании в системе подготовки и

перевода информации является сегодня распространенной практикой в сфере перевода и анализа терминологии.

Поскольку в основе создания любого контролируемого языка лежит исследование терминологии конкретного языка для специальных целей, выявление рекомендуемых и нерекондуемых терминов из массивов параллельных текстов, контролируемые языки можно рассматривать как особый вид языков для специальных целей.

Сегодня различными фирмами, для которых перевод описаний собственной продукции является обязательной частью производства, создаются собственные КЯ с различным сводом правил, при этом единственным общим правилом, принятым во всех доступных для анализа КЯ, является предпочтение коротких предложений (при длине предложения в 12 слов машинный перевод является оптимальным). При этом контролируемый язык может использоваться для написания моделированных текстов – инструкций по эксплуатации, документации на готовую продукцию, сопроводительной документации, рекламных текстов и т.п., для локализации программного обеспечения и технических описаний [Калмыков, Макарова 2010:73-74]. КЯ является важной частью автоматизированного рабочего места специалиста в любой области знаний и важной компонентой систем машинного перевода.

Системы машинного перевода (МП) текстов могут использоваться как часть автоматизированного рабочего места, обеспечивая при этом получение рабочего варианта перевода, жестко ориентированного на конкретную предметную область, задачи пользователя и тип документации. Кроме того, такая система может помочь специалисту, не знающему иностранный язык, очень быстро и с небольшими затратами получить приблизительный (грубый) перевод текстов в интересующей его области знаний, перевод, достаточный для понимания информации, передаваемой текстом на иностранном языке.

Использование системы МП дает переводчику выигрыш в продуктивности работы за счет меньшего времени на перевод; выигрыш в качестве за счет согласованности и корректности перевода терминологии; выигрыш в удобстве организации работ за счет возможности контроля над переводом потоков текстов и возможности разделения работы между несколькими переводчиками.

В то же время ограничения любой системы МП определяются следующими особенностями реализации процесса преобразования текстов: локальностью перевода (переводом по предложениям), особенностями работы с неопознанными словами (геоназваниями и именами собственными, фирменными знаками и редкими словами), линейностью распознавания оборотов.

Учет этих ограничений и особенностей структурной и семантической организации текста позволяет использовать систему перевода в практической деятельности. Деятельность переводчика с использованием АРМ можно условно разделить на 2 этапа: основной (собственно перевод) и вспомогательный (подготовка к переводу, настройка и модификация собственного лингвистического ресурса). Вспомогательный этап также можно разделить на 2 части (предварительную и поддерживающую), поскольку АРМ должно быть сформировано и настроено на потребности конкретного переводчика и/или специалиста.

Выбирая систему машинного перевода, необходимо учитывать не только уровень ее лингвистического обеспечения (оно никогда не будет соответствовать потребностям

конкретного переводчика, тем языкам для специальных целей, с текстами на которых ему приходится сталкиваться в повседневной работе), но и удобство интерфейса и всей системы для него лично. Практическая работа переводчика с использованием автоматизированного рабочего места или с отдельными сервисами из спектра информационных технологий предусматривает: подготовку исходного текста (массива текстов) к переводу—ручное редактирование текста; настройку системы машинного перевода или комплекса используемых словарей на конкретную предметную область или язык для специальных целей; собственно перевод текста, выполняемый в любом из возможных режимов: машинный перевод всего текста, ручной перевод с использованием автоматических и автоматизированных словарей, гибридный перевод с использованием систем переводческой памяти; редактирование результатов перевода—ручное постредактирование переводов, требующее подключение не только дополнительных словарных баз, но и других источников информации, включая справочники и энциклопедии; ведение собственного (пользовательского) словаря, фиксирующего результаты работы с машинными переводами и определяющего настройку системы МП на задачи конкретного переводчика.

При включении процесса перевода в систему обработки информации первый этап выявления номинируемых объектов и анализ способов их номинации является задачей терминоведа, который выделяет термины в структуре текста и определяет их перевод в рамках конкретного языка для специальных целей. Вопрос о фиксации выделенных номинаций в словарной базе конкретного АРМ решается терминоведам и переводчиком в зависимости от степени распространенности соответствующей лексической единицы.

В процессе перевода в режиме реального времени текст может предварительно обрабатываться, при этом:

- 1) получение частотного словаря и миниконкорданса по конкретному тексту помогает выявить основную терминологию и установить ее контекст,
- 2) предварительное редактирование текста позволяет снять его стилистические несообразности, устранить сверхдлинные предложения и т.д.
- 3) использование системы машинного перевода, выбранной и настроенной на необходимую предметную область, дает вариант перевода, который требует анализа и постредактирования.

Собственно процесс перевода может в этом случае осуществляться либо в стандартном двухоконном интерфейсе Windows, когда исходный текст и перевод представлены в двух параллельных окнах, что позволяет редактировать и заново переводить оба текста, либо в режиме работы только с текстом перевода. Опыт практической работы с системами МП позволяет рекомендовать именно первый вариант как наиболее удобный.

Постредактирование результатов МП и получение окончательного варианта перевода текста требует обращения к словарным и энциклопедическим базам данных, выбранным переводчиком, а также к заранее выбранным корпусам текстов. При решении вопроса о выборе перевода конкретной терминологической единицы необходимо привлечение миниконкорданса. В результате работы на этапе собственно перевода должен формироваться пользовательский словарь, характеризующий терминологические особенности конкретного текста. Этот словарь на этапе поддержки АРМ добавляется в его лингвистические ресурсы.

Таким образом, после завершения перевода конкретного текста должна происходить перенастройка лингвистических ресурсов: пополняться корпус параллельных текстов за счет исходного текста и его перевода, формироваться и/или пополняться пользовательский словарь, включающий терминологию, выявленную и проверенную переводчиком, пополняться база словарей.

Учет описанных выше ограничений, методов работы с системой МП и особенностей структурной и семантической организации текста позволяет использовать систему перевода в практической деятельности. Более того, использование результатов машинного перевода оказывается важным ресурсом терминологической работы, источником для формирования словарей и глоссариев. При этом именно текст оказывается в центре особого внимания.

Дело в том, что при современном антропоцентрическом подходе к лингвистическому исследованию можно установить ряд базовых подходов к тому, что понимается под текстом как источником информации о человеке—авторе, адресанте и о той коммуникативной ситуации, в которой он создается, воспринимается и/или передается [Чернявская 1999]. Язык научных текстов может рассматриваться как особый функциональный язык, разновидность общелитературного языка, которому придается большая автономность—язык для профессиональных (специальных) целей. В отличие от большинства искусственных систем переработки, хранения и передачи информации язык представляет собой открытую динамическую неравновесную метасистему [Пиотровский 2006].

Текст можно рассматривать как результат решения задачи передачи информации и источник (отправную точку) ее извлечения. Соответственно, одним из основных компонентов информационной составляющей специального текста [Герд 2011], т.е. той информации, которая может быть извлечена из текста реципиентом, является его денотативный компонент, связанный с особенностями номинации описываемых объектов [Кручинина 2009: 48]. При включении процесса перевода в систему обработки информации первый этап выявления номинируемых объектов и анализа способов их номинации является задачей терминолога, который выделяет термины в структуре текста и определяет их перевод в рамках конкретного языка для специальных целей.

4. Моделирование и концептуализация в терминологическом исследовании: когнитивная карта vs. онтология

В ситуации перевода возможность извлечения информации из научного текста определяется для читателя корректностью передачи терминов—имен объектов. При этом адекватность восприятия текста на лексическом уровне определяется насыщенностью текста именными единицами, степенью компрессии и/или развернутостью номинации объектов. Если рассматривать это в синергетическом аспекте, то можно предположить, что при «перенасыщенности» текста перевода слишком длинными или чрезмерно свернутыми лексическими комплексами, текст разрушается и его восприятие затруднено или даже невозможно.

Учет сложности восприятия подобных лексических образований и их преобразование в более простые комплексы позволяет создавать адекватный текст перевода, то есть от хаоса нагромождения терминологических единиц переходить к новому состоянию, отражающему структуру денотативного пространства текста и терминосистему конкретного языка в том числе. Кроме того, в реальном тексте могут встретиться номинации, характерные для

конкретной организации, их значения как правило не фиксируются отраслевыми словарями [Кит 2010], представляют особую проблему для адекватного перевода и должны включаться в лексикографическую базу АРМ.

Так, например, техническая терминология Компании Форд представляет собой сложную задачу при выборе эквивалентов перевода для хорошо определенных терминов, используемых в специальном контролируемом языке компании [Rychtyskyj 2002:210-212]. В этом языке существует большое количество терминов, которые описывают процессы автомобилестроения и комплектующие, которые используются только в рамках Компании Форд по производству автомобилей. Эти термины включают акронимы, аббревиатуры, геоадреса, связанные с местоположением Форда, и другие термины, которые не могут быть переведены теми, кто не знаком с корпоративным языком. Обнаруживается, что многие из терминов не были понятны всем сотрудникам компании, поскольку они могут использоваться только в рамках одного подразделения завода. Все эти термины должны быть опознаны, переведены вручную и только после этого введены в словари и глоссарии компании. Кроме того, проблемы связаны с теми словарными статьями (см., например, *shotgun* в значении *механизированное приспособление для развинчивания труб*, а не *дробовик*), которые используются неофициально для описания инструментальных средств и оборудования завода. Переводные отраслевые бумажные словари оказываются в таких случаях бессильны. Соответственно, в АРМ компании Форд разработаны специальные средства, позволяющие вводить в пользовательские словари используемой системы машинного перевода корпоративную терминологию [Rychtyskyj 2002: 213].

Сопоставительный анализ структуры именных групп разных языков позволяет сделать вывод о различиях в принципах номинирования сложных объектов и степени отражения особенностей этих внеязыковых объектов при расчлененной (многокомпонентной) номинации. Вопрос о том, являются ли эти различия культурологическими, когнитивными или определяются типом языка, остается дискуссионным.

Одной из сложных проблем работы со знаниями является выбор адекватной модели их представления. В настоящее время разработано множество подходов к представлению знаний, в основе которых используются такие модели представления знаний, как формально-логические, продукционные, семантические сети, фреймы, онтологии. Эти модели представления знаний имеют ограниченные области решаемых задач в силу присущих им свойств и ограничений.

Моделирование динамических систем, к которым можно отнести и сам язык, и речемыслительную деятельность в целом, и терминологические системы языков для специальных целей, представляет собой сложную задачу, требующую решения задач категоризации, а также применения качественных и количественных (вероятностных) оценок создаваемых моделей. Кроме того, моделирование требует проверки адекватности разработанных моделей при условии использования при их создании субъективного опыта исследователя. В когнитивных науках структура субъективного опыта человека изучается в эмпирическом аспекте на основе моделирования, эксперимента и проверки [Баксанский 2006: 94].

Утверждение, что содержание каждой языковой единицы в системе языка представлено определенной структурой знания, вызывает вопрос о том, какова эта структура. Из данных нейролингвистики и записей энграмм памяти следует, что один и тот

же когнитивный объект оказывается компонентом сразу нескольких ассоциативных множеств — и по оси сенсорных модальностей, и по осям разного рода парадигматических и синтагматических связей [Черниговская 2008; 2010]. Вопрос о том, имеем ли мы дело с работой долговременной памяти или с постоянной перестройкой нейронных сетей в памяти ассоциативной, пока остается открытым.

Сегодня можно утверждать как некоторую аксиому, что объективное исследование терминологии конкретной предметной области требует предварительного моделирования этой предметной области для того, чтобы установить систему ее понятий и связей между ними. При этом принято различать **терминопole** как системное образование плана содержания, представляющее собой организованную совокупность специальных понятий и связей, и **терминосистему** как совокупность языковых средств, терминопole отражающих.

Установление и описание терминосистемы конкретной предметной области или языка для специальных целей предполагает установление не только набора терминов, но и понятий, что требует гармонизации терминологии. Процесс гармонизации включает сравнение систем понятий, что предполагает установление отношений между понятиями, фиксацию количества понятий и глубины структуры, устранение дублирования и т.п., что приводит к построению новой гармонизированной системы понятий.

В основе решения этой задачи лежит структурирование терминопole на базе выявления основных узлов (подтем) и связей между ними: гиперо-гипонимических, меронимических, ассоциативных и т.п. По отношению к этой структуре и происходит фиксация терминосистемы конкретного подъязыка или терминосистем разных языков. Наличие зафиксированной в виде графа структуры терминосистемы позволяет соотносить терминологические единицы разных языков с одним и тем же узлом терминопole и, соответственно, между собой.

Наиболее традиционным подходом к изучению соотношения терминопoleй и отражающих их терминосистем является структурное моделирование с помощью тезауруса. При этом под тезаурусом понимается упорядоченное множество понятий предметной области вместе с заданными на этом множестве отношениями. Материалом для тезауруса служит синхронный срез лексики, который определяется на основе изучения современного состояния моделируемого терминопoleа и отражающей его термносистемы, извлекаемой из словарного и текстового материала на одном или нескольких языках. При этом значение термина определяется его позицией в соответствующей модели и структурой отношений разных типов, которыми термин связан с другими. Как правило, отношения, фиксируемые в тезаурусе, являются таксономическими. В идеале эта структура должна отражать соответствующие отношения между элементами терминопoleа (понятиями) соответствующей предметной области. Двумерная структура сети схематично задается деревом.

Особый интерес к созданию и исследованию когнитивных моделей терминологии проявляется в конкретных разработках, в частности, в лексико-семантической системе WordNet, разрабатываемой лабораторией когнитивистики Принстонского университета. Эта система может рассматриваться как часть единой семантической сети и фрагмент онтологии, формируемой как словарь понятий. Глобальная задача построения онтологии представляет собой задачу создания модели мира, необходимой для смысловой переработки текста. Онтология должна задавать понятия для представления значений слова

в лексиконе, связи этих понятий с номинирующими единицами и концептами, иными словами, онтология представляет собой (или, скорее, должна представлять) модель предметной области и ее манифестации в соответствующей терминологии.

Термин **онтология** для описания процесса концептуализации прежде всего относится к проблеме создания открытого для общего и многократного использования информационного ресурса, формируемого как словарь понятий (термин в этом значении был введен Томом Грубером [Gruber 1993]). При современном подходе онтология представляет собой базу знаний, хранящую информацию о понятиях, существующих в мире или предметной области, их свойствах, и о том, как они связаны друг с другом. Тем самым фокусом описания в онтологии являются отношения между понятиями и соответствующими объектами. Опыт использования онтологий показывает, что они являются и мощным инструментом создания/выявления систем понятий для различных предметных областей [ТКЕ 2010].

Онтология отличается от тезауруса тем, что содержит только не зависящую от языка информацию и множество семантических отношений, кроме того, она содержит таксономические отношения. Она должна задавать понятия для представления значений слова в лексиконе [Biebow, Szulman 1999]. Онтологию можно рассматривать как пространственное представление семантической сети — сферу, в узлах внешней сетчатой поверхности которой располагаются терминологические единицы разных языков. Эти термины связаны с узлами-понятиями, располагающимися внутри (на следующем уровне) сферы. Узлы-понятия, в свою очередь, могут быть связаны с концептами, отражающими факты внешнего мира и/или нашими представлениями о них и хранящимися на более глубоких уровнях (слоях) сферы.

Альтернативный подход к моделированию терминологии предлагается в некоторых современных публикациях [Массалина, Новодранова 2009: 17-18, 48; Новодранова 2010: Пахалюк 2009], в которых предлагается использовать когнитивные карты для моделирования терминологии и, более того, науки. Следует отметить, что идея моделирования значения термина как единицы терминосистемы в виде нейронной сети появилась в рамках когнитивной лингвистики, активно вовлекающей в свой арсенал методы и термины когнитивистики как некоторой обобщающей науки. При этом нейрон (узел сети) рассматривается как модель понятия (концепта), характеризуемого вероятностью активации в конкретной ситуации (как реального мира, так и речевой), а динамические каузальные (причинно-следственные) связи между концептами (дуги) характеризуются положительными или отрицательными вероятностями [Aguilar 2002: 403]. Такую схему представления семантики и принято называть когнитивной картой.

При моделировании терминополья нечеткость значений лексических единиц [Пиотровский 2006] приводит к необходимости провести операцию сгущения значения, позволяющую определить то конкретное значение, которое будет приписано соответствующей единице в модели. Если мы рассматриваем подход к такому моделированию как «ориентированный на концепты» [Уралев 2008], то каждому концепту (узлу сети или когнитивной карты) должно быть приписано одно лексическое значение, в случае онтологии это значит, что внутреннему узлу-концепту соответствует столько терминов на разных языках, сколько у него есть значений. Каждому концепту при этом приписывается уникальный идентификатор, а терминам — идентификатор языка и

идентификатор концепта. Поскольку значения лексических единиц (и терминов в том числе) описываются в терминах теории нечетких множеств, то для моделирования соответствующих понятий используются так называемые нечеткие когнитивные карты [Kosko 1986].

Формально нечеткая когнитивная карта представляет собой нечеткий знакоориентированный граф с самовставлениями (не дерево!), который моделирует миры как совокупности концептов и каузальных отношений между ними [Aguilar 2002: 404]. Концепты репрезентируются узлами графа, взаимосвязи между концептами—ребрами. Нечеткость проявляется в том, что узлами таких карт являются концепты, которые могут быть представлены как нечеткие множества, а каузальные отношения между такими концептами представляют собой нечеткие импликации, условные вероятности и т.п. Ориентированные ребра между узлами-концептами измеряют, насколько один из этих концептов связан с другим причинно-следственными отношениями, насколько они каузируют друг друга. Соответственно, отношения каузации в такой системе связей можно описать как положительные или отрицательные в интервале $[-1, 1]$. Взятые вместе, узлы и связи подобной когнитивной карты позволяют оценить состояние моделируемой системы. Когнитивные карты широко применяются для структуризации знаний эксперта, для выявления согласованного мнения группы экспертов или анализа различий в таких мнениях, а также для подготовки управленческих решений [Авдеева, Коврига 2010].

В определенном смысле нечеткая когнитивная карта может рассматриваться как обобщенный способ представления динамического фрейма, позволяя при корректно вычисляемых отношениях между концептами моделировать ситуации внешнего мира [Aguilar 2002:407]. В то же время, когнитивная карта представляет собой, скорее, многофакторную модель деятельности, в которой каскады вероятностных оценок взаимодействия концептов в результате развития ситуации позволяют прогнозировать события и предлагать возможные способы нивелирования отрицательных последствий. Моделирование же в области конкретной науки (предметной области) и, более того, в сфере языка для специальных целей не позволяет ни ввести такие осмысленные вероятностные оценки, ни экспериментально проверить их адекватность. Можно полностью согласиться с опасностью подмены анализа языковых фактов рассуждениями о фактах действительности [Рахилина 2010: 229]. Поэтому использование термина когнитивной карты и соответствующего ей аппарата анализа и описания при исследовании терминологии представляется непродуктивным, поскольку нет объективных методов установления каузальных зависимостей между единицами как терминосистемы, так и терминопоя.

Материалом для тезауруса или онтологии служит синхронный срез лексики, который определяется, как правило, современным состоянием моделируемого терминопоя и отражающей его терминосистемы, которая извлекается из словарного и текстового материала на одном или нескольких языках. В то же время мы привыкли постулировать постоянное развитие различных научных и технических сфер деятельности человека, уделяя недостаточное внимание отражению этого развития в языке. Если вслед за Гийомом считать, что язык представляет собой систему систем или диахронию синхроний, то необходим именно диахронный подход к моделированию терминосистемы для того, чтобы установить границы возможных синхронических срезов, причины и особенности перехода от одного состояния к другому.

При таком подходе на уровне, предшествующем созданию текста, возникает антиномия стабильности—изменчивости, заключающаяся в том, что терминосферы представляют собой образования целостные и устойчивые, иначе их изменчивость была бы препятствием для коммуникации. С другой стороны, изменение предметной сферы влечет за собой изменение в терминосфере, иначе коммуникация также была бы нарушена. Поскольку терминосистемы основных европейских языков развиваются в тесном контакте, определяемом теперь еще и понятием «Европейской деревни», то их контрастивное исследование позволяет понять, чем определяется устойчивость терминосистемы конкретного языка и те условия, при которых она (терминосистема) изменяется, не теряя своей коммуникативной функции.

Особенностью терминологических систем английского языка, описание которых необходимо для создания корректных переводных словарей, является использование этого языка как глобального. Большая часть научных текстов, попадающих в поле зрения переводчиков, написаны именно на глобальном английском языке. При этом текст можно рассматривать как результат решения задачи передачи информации и источник (отправную точку) ее извлечения. Соответственно, содержание научного текста, то есть та часть его смысла, которая является универсальной и может быть извлечена при минимальном совпадении тезаурусов автора и получателя, определяется в основном информацией об объектах, лингвистически описываемых именами существительными и именными словосочетаниями.

5. Терминологические ресурсы современных систем и корпусные технологии

Современная работа терминолога в рамках переводческого центра меняется, поскольку именно терминолог идентифицирует в тексте термины, проверяет фиксацию их и их переводов в базе данных и ведет всю лексикографическую работу по установлению перевода и описанию нового термина. Только после этого "размеченный" и/или обработанный системой машинного перевода текст передается переводчику [Towards Consolidation 2006].

Поддержание лексикографической системы, используемой переводчиком, в актуальном состоянии требует постоянного "отслеживания" новых терминов, для чего могут, в частности, использоваться методы собственно лингвистического и лингвостатистического анализа и метрики для выделения терминологических словосочетаний из текста. Автоматическое извлечение терминов (как универбов, так и многокомпонентных лексических единиц—коллокаций) основано на предварительном выравнивании текстов на разных языках, идентификации терминологических единиц в текстах на одном языке и дальнейшем установлении их переводных эквивалентов или, скорее, кандидатов в возможные переводные эквиваленты. Хотя утверждается, что подобная задача хорошо решена для разных языковых пар в случае анализа параллельных текстов, ее решение еще требует исследований в случае сопоставления языков с различной графикой. Для выделения коллокаций разработано более 80 различных метрик, которые оценивают информацию о сочетаемостных предпочтениях лексических единиц (*unithood*), их синтагматической близости, информацию о степени терминологичности словосочетаний (*termhood*), информацию о характерных особенностях (*salience*) словосочетания для конкретного корпуса тестов или терминологии языка для специальных целей [Hertog et al. 2010].

Методы, используемые для оценки синтагматичности, т.е. устойчивости фиксации элементов словосочетания, основываются на оценках частоты словосочетаний в текстах конкретной предметной области. Для оценки терминологичности словосочетания необходимо сравнение его частотных характеристик в разных предметных областях, т.е. требуется контрастивный подход.

Применение соответствующих метрик требует создания корпуса текстов и его подготовки для автоматизации терминологического анализа. Идеальным источником материала являются корпуса параллельных текстов, построенные на основе материалов узкой предметной области (статей, монографий, материалов конференций и их переводов на другой язык). Такой корпус должен быть выровнен по предложениям, что позволяет выявлять и анализировать термины и их переводы, оценивать стандартизованность и единство переводов, распространенность конкретных вариантов. Однако получение такого параллельного корпуса не всегда возможно. Одним из вариантов создания материала для последующего лексикографического анализа является формирование особых корпусов текстов, включающих параллельное представление исходных текстов, их машинных переводов и отредактированных переводов, согласованных с экспертами в конкретной области знаний. Важно отметить, что качество и потенциал такого корпуса в большой степени зависит от сотрудничества с экспертами при отборе исходного материала и редактировании переводов.

Использование корпуса параллельных текстов позволяет не только максимально автоматизировать отбор терминологических словосочетаний, но также служит для обогащения набора словарных статей за счет выбора свободных словосочетаний, используемых в исходных текстах, что чрезвычайно важно для тех, кто переводит на язык, не являющийся родным; уточнения употребительности конкретных словосочетаний в текстах определенной предметной области; верификации значений лексических единиц, зафиксированных в двуязычных словарях, особенно в том, что касается идиом и терминологических выражений; выделения устойчивых терминологических словосочетаний, которые целесообразно вводить в словарь конкретной отрасли знаний.

На основе полнотекстовых баз параллельных выровненных текстов возможно выделение устойчивых пар слов типа “исходное слово—перевод”, однако применение статистических процедур, как правило, допускает соответствие слов, но не словосочетаний. Это достаточно жесткое ограничение для выбора потенциальных компонентов словаря может быть уменьшено, если параллельные тексты предварительно проходят процедуру парсинга или подвергаются ручной разметке, что позволяет соотносить не отдельные слова, а фрагменты предложения.

Итак, современный подход к созданию лингвистических ресурсов предполагает формирование и использование корпуса реальных текстов, который может рассматриваться как база данных для решения не только исследовательских, но и практических лексикографических задач. Корпусы письменных текстов, как правило, включают сами тексты, а также их разметку с точки зрения формата и предложений по результатам парсинга, позволяющего установить принадлежность лексических единиц к конкретным частям речи. Эти тексты могут служить для создания конкордансов, словарей слов и словосочетаний в случае одноязычного корпуса, а также для создания многоязычных лексиконов и многоязычных конкордансов в случае корпуса параллельных массивов.

При создании переводного словаря на основе корпуса текстов необходимо определить принципы формирования выборочной совокупности для создания исследовательского параллельного корпуса текстов и ее необходимый и достаточный объем; установить требования к разметке текстов в корпусе и получить базовую лингвистическую информацию; определить необходимые для работы средства информационных технологий, включая системы машинного перевода, переводческой памяти, средства выравнивания параллельных текстов, редакторы разметки (тэггирования), средства формального извлечения терминов из текста и т.д.

Средства информационных технологий могут быть ориентированы на выбор и обработку терминов и/или понятий, на работу с конкретной языковой парой, многоязычными или одноязычными ресурсами. Подобные средства целесообразно включать в автоматизированное рабочее место лексикографа, на основе которого можно осуществлять запись, обработку, сохранение и использование различных лингвистических и лексикографических данных.

Создание исследовательского корпуса текстов для решения лексикографических задач основывается на следующих этапах формирования выборочной совокупности:

- установление подобластей предметной области (структуры языка для специальных целей) на основе выбранной классификации знаний;
- отбор текстов разных жанров и сопоставимого объема;
- разметка текстов, включающая фиксацию коллокаций.

В корпусной лингвистике принято различать корпуса параллельных текстов (*parallel corpora*) и корпуса псевдопараллельных текстов (*comparable corpora*). Их различие, кроме всего, связано с принципами отбора текстов. В случае псевдопараллельных текстов их отбор может осуществляться на основе достаточно ясных критериев. При создании корпуса параллельных текстов для последующего терминологического анализа и извлечения пар типа термин—перевод важным условием “успешности”, адекватности созданного ресурса является качество выбранных переводов. При создании подобного корпуса для других целей качество перевода не столь важно.

Соответственно, проблема оценки качества перевода оказывается кардинальной для формирования корпуса параллельных текстов, критериями такой оценки могут быть последовательность использования номинаций в тексте перевода, соблюдение переводчиком норм языка перевода, сохранение логической структуры исходного текста и, в целом, экспертная оценка качества перевода.

Рассмотрение корпуса параллельных и/или псевдопараллельных текстов в качестве лексикографической базы предполагает необходимость его дополнения корпусом машинных переводов текстов, что позволяет явным образом выделить те лексические единицы, которые должны быть введены в словарь или перевод которых требует модификации.

Заключение

Описанные выше особенности работы переводчика, терминолога и лексикографа, а также проблемы моделирования терминологических систем и процесса перевода определяют необходимость создания лингвистических и программных сервисов, которые должны [<http://www.w3.org/WAI/PF/usage/languageUsageAndAccess.html>]:

—облегчить взаимодействие между людьми, поддерживая с помощью компьютера их совместную работу несмотря на различия в естественном языке;

—адаптировать содержание текста, что существенно улучшает доступ к информации для людей с различными нарушениями чтения или для тех, кто нуждаются в переводе;

—сделать многозначное содержание мультимедийных форматов Интернета доступным для автоматической обработки, и интерпретируемым системами, основанными на семантическом анализе.

В настоящее время разрабатывается огромное количество разнообразных программных продуктов, которые отличаются используемыми платформами, на которых они реализованы, и специфическими особенностями классов решаемых задач. Эта ситуация требует создания особых программных средств—промежуточных программных средств (middleware), которые позволили бы транслировать нужные функции уже имеющихся программных средств и систем обработки данных [Wichadakul, Nahrstedt 2002].

Все кратко описанные выше системы и виды информации могут храниться как непосредственно в базе данных АРМ, так и извлекаться из системы Интернет, в последнем случае в АРМ должна храниться специализированная информация о соответствующих сайтах. Возможность использования ресурсов сети Интернет оказывается решающей для создания полноценного АРМ, поскольку в этой сети созданы возможности для широкого обмена многоязычной и мультимедийной информацией.

Список литературы

- [1]Авдеева З. К., Коврига С. В. Эвристический метод концептуальной структуризации знаний при формализации слабоструктурированных ситуаций на основе когнитивных карт // Управление большими системами. 2010. Выпуск 31, с. 6-34
- [2]Баксанский О.Е. «Когнитивная карта» и реальность // Эпистемиология и философия науки, 2006, Т. VII, № 1, с. 93-96
- [3]Гавриленко, Н.Н. Формирование умений информационно-справочного поиска при переводе профессионально-ориентированных текстов. // Вестник МГЛУ, 2011, вып. 12 (618), с. 81-89
- [4]Гаврилова Т.А., Червинская К.Р. Извлечение и структурирование знаний для экспертных систем. Москва, 1992.
- [5]Герд А.С. Введение в изучение языков для специальных целей. СПб: СПбГУ. РИО Филологический факультет, 2011, 59 с.
- [6]Демич К.И., Костыркин А.В., Шаляпина З.М. РАМЕЯ/и – автоматизированное рабочее место лингвиста для иероглифических языков: общая организация // Вестник РГГУ, 2010, № 9, с. 117-148
- [7]Калмыков В.А., Макарова Г.И. Локализация программного обеспечения как переводческая проблема // Вестник Университета Российской академии образования. Москва: Университет Российской академии образования, 2010, № 1, С. 72-75
- [8]Кит М.С. О стратегии построения высокоэффективных сетевых словарей (на базе разработки словаря LexSite) // Вестник РГГУ, 2010, №9, с. 149-160
- [9]Кияк Т.Р. Лингвистика профессиональных языков и терминоведение // Терминология и знание. Материалы I международного симпозиума. Москва: Институт русского языка им. В.В. Виноградова РАН, 2009, с.21-27
- [10]Кручинина А.А. Информационная составляющая специального текста и ее реализации в переводе. // Вестник НГЛУ. Нижний Новгород: изд-во НГЛУ им. Н.А.Добролюбова, 2009, Вып. 6, С. 47-55

- [11]Кубрякова Е.С. Роль когнитивной лингвистики в науках когнитивного цикла // Когнитивные науки: проблемы и перспективы: Материалы российско-французского семинара. Москва: РГГУ, 2010, с. 113-119
- [12]Массалина И.П., Новодранова В.Ф. Дискурсивные маркеры в английском языке военно-морского дела. Калининград: Изд-во КГТУ, 2009, 278 с.
- [13]Митрофанова О.А., Захаров В.П. Автоматизированный анализ терминологии в русскоязычном корпусе текстов по корпусной лингвистике // Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции «Диалог–2009». Москва, 2009, С. 321–328
- [14]Новиков А.И. Текст и его смысловые доминанты. Москва: Институт языкознания РАН, 2007, 224 с.
- [15]Новодранова В.Ф. Десять лет когнитивному терминоведению // Терминология и знание. Материалы II Международного симпозиума. Москва: Издательский дом «Азбуковник», 2010, с. 142-151
- [16]Пахалюк В.Г. Когнитивная карта науки о рекламе // Вопросы когнитивной лингвистики, 2009, № 3 (020), с. 30-34
- [17]Пиотровский Р. Г. Лингвистическая синергетика: исходные положения, первые результаты, перспективы. СПб.: Филфак СПбГУ, 2006, 159 с.
- [18]Пушнов, И.А., Семенов, А.Л., Убин, И.И. Перевод: проблемы и решения. Москва: ВЦП, 2009, 98 с.
- [19]Рахилина Е.В. О цельности концептов и границах категоризации // Когнитивные науки: проблемы и перспективы: Материалы российско-французского семинара. Москва: РГГУ, 2010, с. 223-230
- [20]Ревзин И.И., Розенцвейг В.Ю. Основы общего и машинного перевода. Москва: Высшая школа, 1964, 243 с.
- [21]Сдобников В.В. Коммуникативно-функциональный подход к переводу в российском переводоведении. // Вестник МГУ. Сер. 19. Лингвистика и межкультурная коммуникация. Москва: МГУ, 2009, №2, с. 19-30
- [22]Синявская-Суйковска Т.В. Цель перевода, тип текста и текстовые категории как основные переменные модели перевода // Вестник Томского университета, 2009, №4 (8), с.38-46
- [23]Уралев Г.В. Концептуальная спецификация некоторых задач информатики // Известия Южного федерального университета. 2008, Том 86, № 9, с. 144-151
- [24]Чебанов С.В., Мартыненко Г.Я. Из истории типологических представлений // Структурная и прикладная лингвистика, вып. 7. СПб, 2008, с. 328-390
- [25]Черниговская Т. В. От коммуникационных сигналов к языку и мышлению человека: эволюция или революция // Российский физиологический журнал им. И. М. Сеченова. 2008, Том 94, №9, с. 1017-1028
- [26]Черниговская Т.В. Мозг и язык: врожденные модули или обучающаяся сеть // Вестник РАН, 2010, Том 80, № 5, с. 461-465
- [27]Чернявская В.Е. Интертекстуальное взаимодействие как основа научной коммуникации. СПб: Изд-во СПбГУЭФ, 1999, 209 с.
- [28]Aguilar J. Adaptive Random Fuzzy Cognitive Maps // IBERAMIA. Berlin, 2002, pp. 402-410
- [29]Bell, R.T. Translation and Translating. London: Longman, 1991, 298 p.
- [30]Biebow, B., Szulman, S. 'TERMINAE : a method and a tool to build a domain ontology', in Benjamins, V.R., Fensel, D. and Gomez-Perez, A. (eds) Proceedings of International Workshop on Ontological Engineering on the Global Information Infrastructure, Dagstuhl Castle, Germany, 1999, pp 25-30
- [31]Gruber T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing // Int. Journal Human-Computer Studies, 1993, Vol. 43, pp. 907-928
- [32]Hertog de D., Heylen K., Speelman D., Kockaert H. A Variational Linguistic Approach to Term Extraction // TKE 2010: Presenting terminology and knowledge engineering resources online: models and challenges. Dublin: Dublin City University, Ireland, 2010, pp. 226-249
- [33]<http://www.w3.org/WAI/PF/usage/languageUsageAndAccess.html>

- [34]Hutchins J. Machine Translation and Human Translation : in Competition are in Complementation? / J. Hutchins // Machine Translation : Theory & Practice. New Delhi, 2001, pp. 5–20
- [35]Hutchins J. The origin of the translator's workstation // Machine Translation, 1998, vol. 13, no 4, pp. 287-307
- [36]Kosko B. Fuzzy Cognitive Maps // Int. Journal of Man-Machine Studies, Vol.24, 1986, pp. 65-75
- [37]Muegge U. Controlled language - does my company need it? // электронный ресурс www.tekom.de/artikel/artikel_2756.html. 2009
- [38]Rychtyckyj N. An Assessment of Machine Translation for Vehicle Assembly Process Planning at Ford Motor Company// S. D. Richardson (ed):AMTA 2002, Lecture Notes in Computer science, Vol. 2499, Berlin Heidelberg: Springer-Verlag, 2002, pp. 207-215
- [39]Steinberger R. Language Engineering Technologies and their use for TF-UCLAF. A Report on JRC's Institutional Support Activities // электронный ресурс http://langtech.jrc.it/Documents/Report-98_Steinberger_LangTech4OLAF.pdf
- [40]TKE 2010: Presenting terminology and knowledge engineering resources online: models and challenges. Dublin: Dublin City University, Ireland, 2010, 102 p.
- [41]Towards Consolidation of European Terminology Resources. Experience Recommendations from Euro TermBank Project / Edited by: Signe Rirdance, Andrejs Vasiljevs. Riga: Tilde, 2006, 123 p
- [42]Wichadakul D, Nahrstedt K. A Translation System for Enabling Flexible and Efficient Deployment of QoS-Aware Applications in Ubiquitous Environments // Lecture Notes in Computer Science. Volume 2370. Berlin Heidelberg: Springer-Verlag, 2002, pp. 210-221

Linguistic Resources in Modern Information Space

Belyaeva L.

(the Faculty of Philology the Herzen State Pedagogical University of Russia, St. Petersburg, 191186)

Abstract: The paper considers a set of problems for terminological and translation work on the base of modern information and linguistic tools: machine translation systems, automated and automatic dictionaries and databases that are problem- and subject-oriented. Principles of modeling terminological items and systems are discussed in the framework of modern cognitive science. Potentiality of using parallel and comparable text corpora for solution of practical aims in establishing collocations, proper translations of new terms, controlling the correctness of new texts translations are under consideration.

Key words: linguistic technologies; automated workstation; term extraction; applied lexicography

作者简介: 别利亚耶娃 Л.Н., 女, 俄罗斯赫尔岑国立师范大学语文系教授、博士, 应用语言学教研室主任, 发表论文 200 余篇, 代表性的专著有: 《现代人文技术中的信息自动化》(2007)、《翻译理论与实践》(2007)、《自动词典编纂: 人文技术》(2011)。主要研究方向: 词典学、翻译学、术语学。

收稿日期: 2012-09-03

[责任编辑: 叶其松]