

фонетические символы подверглись изменению. В настоящее время в подсистеме безударного вокализма появились две безударные фонемы /i/, /u/, применение которых вызывает особые сложности у студентов языковых вузов. Исследование случаев употребления этих фонем показало, что этот вопрос требует дальнейшей разработки.

ЛИТЕРАТУРА:

1. Апресян Ю.Д., Медникова Э.М., Петрова А.В. Новый большой англо-русский словарь: В 3-х томах // – М.: Русский язык, 2002. – Т. 1. – 832 с.
2. Ярцева В.Н. Большой энциклопедический словарь. Языкознание. – М.: Большая Российская энциклопедия, 1998. – 685 с.
3. Howlett C. The Oxford Russian Dictionary // Oxford University Press. – Oxford, New York, 1995. – 1340 p.
4. International Phonetic Association. Handbook of the International Phonetic Association. – Cambridge Academ, 1999. – 214 p.
5. Jones D. An English Pronouncing Dictionary. – London, Toronto: Dent. – 2nd edition. – 1924. – 426 p.
6. Jones D. An English Pronouncing Dictionary // New York, E.P. Dutton & Company, inc. – 4th edition. – 1937. – 495 p.
7. Jones D. An English Pronouncing Dictionary // London: J.M. Dent & Sons Ltd. – 8th edition. – 1947. – 490 p.
8. Jones D. Everyman's English Pronouncing Dictionary // London: J.M. Dent & Sons Ltd. New York: E.P. Dutton & Co. Inc. – 11th edition. – 1958. – 538 p.
9. Jones D. Everyman's English Pronouncing Dictionary // London, Melbourne and Toronto. J.M. Dent & Sons Ltd. – 14th edition. – 1982. – 560 p.
10. Jones D. English Pronouncing Dictionary // Cambridge University Press. – 15th edition. - 1997. – 560 p.
11. Jones D. English Pronouncing Dictionary // Cambridge University Press. – 16th edition. - 2003. – 606 p.
12. Leontyeva S.F. A Theoretical Course of English Phonetics // Moscow. - Менеджер. - 2002. – 336 p.
13. Lewis J. Windsor A Concise Pronouncing Dictionary of British and American English // London Oxford University Press. – 1972. – 234 p.
14. MacMahon, Michael K. C. Phonetic notation. In P. T. Daniels & W. Bright (Ed.), The world's writing systems // New York: Oxford University Press. - 1996.
15. Pullum G.K. and Ladusaw A.L. Phonetic Symbol Guide // The University of Chicago Press. - 2nd edition. – 1996. – 358 p.
16. Roach P. English Phonetics and Phonology. A Practical Course // Cambridge University Press. - 3d edition. – 2002. – 284 p.
17. Wells J.C. Longman Pronunciation Dictionary // Pearson Education Limited. 1st edition. – 1990. – 804 p.
18. Wells J.C. Longman Pronunciation Dictionary // Pearson Education Limited. 2nd edition. – 2000. – 870 p.

УДК 81'322.4

Рафанов С.В.
МГУ им. М.В. Ломоносова

К ПРОБЛЕМЕ КЛАССИФИКАЦИИ ТЕКСТОВ В МАШИННОМ ПЕРЕВОДЕ*

Аннотация. Классификация текстов, подлежащих автоматизированному переводу, имеет большое теоретическое и практическое значение для повышения эффективности машинного перевода. В статье рассматриваются основные параметры, которые должны учитываться при разработке эффективной классификации. Обосновывается необходимость построения классификации исходя из лексического состава текстов. Отмечается возможность решения проблемы путем детального описания лексических разрядов текстов и необходимость учитывать эту информацию при разработке машинных словарей.

Ключевые слова: машинный перевод, классификация текстов, функциональный стиль, сфера коммуникации, лексический разряд, машинный словарь.

S. Rafanov
TO THE PROBLEM OF TEXT CLASSIFICATION IN MACHINE TRANSLATION

Abstract. Classification of texts liable for automatic translation is of crucial theoretical and practical importance for increasing Machine Translation productivity. The article considers main principles to take into account when efficient classification is being developed. Necessity of making classification on the basis of texts lexis is substantiated. It's also noted that possible solution to the problem is through detailed description of texts' lexical groups and necessity to use this data when machine dictionaries are being designed for machine translation system.

Key words: machine translation, texts classification, functional style, communication sphere, lexical class, machine dictionary.

Основная цель машинного перевода (МП) – качественное улучшение научно-информаци-

* © Рафанов С.В.

онных контактов в обществе и преодоление языковых барьеров посредством аналитико-синтетической обработки документальной информации.

Первоочередной задачей процесса аналитико-синтетического анализа документов в различных сферах информационного обслуживания – архивное дело, канцелярская деятельность, библиотечное дело – является классификация текстов по тем или иным основаниям (содержание, авторство, движение, хранение и т.д.). Классификация документов определяет эффективность дальнейшего их использования и обработки. Для МП данная процедура является определяющей с точки зрения качества получаемого перевода. И решение проблем, связанных с классификацией текстов в МП, является актуальной задачей в процессе совершенствования работы систем автоматизированного перевода

Под классификацией подразумевается разделение предметов какого-либо рода на классы по наиболее существенным признакам, присущим предметам данного рода и отличающим их от предметов других родов. Классификация текстов МП должна основываться на принципах, которые бы наглядно показывали:

- общий, интегральный признак, объединяющий разные классы текстов;
- специфику каждого конкретного класса текстов;
- соотношение классов между собой.

Распределение текстов по классам может основываться не только на наличии или отсутствии какого-либо признака, но и на степени его проявления.

Помимо теоретической составляющей, классификация должна иметь практическое применение, а именно – отвечать задачам совершенствования лингвистического обеспечения с целью дальнейшего расширения использования МП в информационных процессах в обществе.

Многоаспектность и сложность феномена текста в МП не позволяет исходить из какого-либо одного критерия текстовой дифференциации. Необходимо учитывать ряд показателей и охватывать основные признаки текста: структурно-семантические, коммуникативные и информационное наполнение. Каждый из данных критериев теоретически способен стать основанием для соответствующей классификации. Объединенные же вместе, они создают определенные трудности: каждый реальный текст должен теоретически выявить свой собственный, отличный от других признак по каждому из этих оснований. Такая

«идеальная» и непротиворечивая классификация затруднительна, поскольку сходства и различия признаков могут комбинироваться по-разному: например, сходству информационных качеств могут резко противостоять качества коммуникативные и т.п. Выбор критериев осложняется и тем, что один и тот же текст может быть отнесен к различным группам из-за своей собственной многоаспектности: по одному критерию он войдет в одну группу текстов, по другому – в другую. При ориентации на разные критерии можно в первичной дифференциации остановиться на делении «научные и ненаучные тексты»; «художественные и нехудожественные тексты»; «монологический и диалогический тексты»; «моноадресатный и полиадресатный тексты» и т.д. Каждое из этих делений существует, но с точки зрения общей и единой типологии они некорректны: например, художественный текст, с одной стороны, попадет в группу ненаучных, а с другой – одновременно в группы монологических и диалогических. Более того, все они не существенны для целей создания классификации текстов в МП.

На этапе первичной дифференциации проводится выборка текстов по признакам, которые являются исходными для включения того или иного текста в комплекс речевых произведений, подлежащих машинному переводу. Далее, поскольку каждый текст являет собой сложное образование и строгое разграничение по одному показателю было бы неизбежным искажением действительной картины, определяются значимые для МП признаки для каждого транслатологического типа текстов, которые носят описательные характер и могут быть использованы на разных этапах переводческой стратегии при работе с системами МП. Сами же названия транслатологических типов определяются исходя из существующей языковой и переводческой практики.

Первый этап дифференциации связан с традиционным в переводоведении разделением текстов на художественные и нехудожественные. Первая группа текстов принципиально исключается из рассмотрения, так как для перевода таких текстов требуется целый комплекс знаний, навыков и чувств – все то, что объединяется понятием дух языка. На этом этапе следует также добавить ограничивающие критерии – монологичность и письменная форма представления. Дальнейшая дифференциация связана с аспектами семантики и прагматики в МП и местом систем МП в системе общественных коммуникаций. Учитывая требования эквивалентности в МП и прагматичес-

кой адекватности, из рассмотрения исключаются тексты, функционирующие в рамках индивидуально-обиходной и массовой сфер коммуникации. К анализу, таким образом, представлены тексты, функционирующие в рамках специальной коммуникации и в пограничной зоне массовой и специальной сфер.

Дальнейшие выборка и разграничение должны учитывать стилевую принадлежность текстов. Здесь мы ограничимся стилями официально-деловым и научно-техническим. Следует отметить, что в современной языковой практике каждый из этих стилей представлен не только в сфере специальной коммуникации. Деятельность человека в научной сфере, особенно в сфере разработки и проектирования новых технических устройств, обусловила появление речевой деятельности, направленной на обеспечение обслуживания и эксплуатацию данных устройств не только специалистами. В обществе функционируют огромные массивы текстов, рассчитанные на широкий круг читателей (признак массовой коммуникации) и, в то же время, предназначенные для передачи информации, относящейся к разнообразным техническим операциям (признак специальной коммуникации), – например, бытовая инструкция.

Таким образом, классификационная схема текстов может быть представлена в виде таблицы 1.

По степени общности можно выделить 2 группы: 1) фасеты 1 и 2; 2) фасеты 3 и 4. Для каждой из них должны быть определены и описаны внутригрупповые интегрирующие и дифференцирующие признаки, релевантные с точки зрения МП. Интегрирующие признаки в каждой из групп определяются общей структурой текста, особенностями синтаксиса и коммуникативной задачей. Дифференцирующие признаки связаны с особенностями лексического состава текстов и их предметной направленностью. Данные признаки связаны с двумя уровнями машинного анализа любого текста – автоматический синтаксический \ морфологический анализ и автоматический лексический анализ.

Архитектура современных систем МП устроена так, что алгоритм обработки входного текста, применительно к морфологии и синтаксису, не варьируется в зависимости от типа текста. Процедура синтаксического анализа текста бытовой инструкции, например, не имеет принципиальных отличий от процедуры анализа собственно научного текста. Определенные процедурные вариации могут быть связаны лишь с переводом отдельных грамматических форм и они, как правило, задаются на этапе предпереводческой подготовки в виде алгоритмов: перевод модальных глаголов безличными конструкциями, перевод императивных форм инфинитивом, перевод местоимения *И* местоимением *Мы* и т.д.

На этапе зарождения МП в основе большинства теорий лежало требование перевода слов и словосочетаний. Именно этот уровень текстовой организации является отправной точкой всех алгоритмов анализа и синтеза. В связи с этим лексический уровень организации текстов рассматривается как основной критерий их дифференциации. В данном случае дифференцирование текстов имеет два вектора: качественное дифференцирование и количественное. Первое предполагает разграничение текстов по жанровой и содержательной направленности, второе – характеристика каждого класса с точки зрения разрядного состава лексических единиц и их количественного соотношения.

Опираясь на классификацию лексических разрядов, описанную Ю.В. Рождественским [Рождественский Ю.В., Волков А.А., Марчук Ю.Н., 1988, 116], в группах текстов, представленных в табл. 1, можно выделить следующие категории слов:

- 1) слова общелитературного языка, выступающие в традиционных для литературного языка значениях.
- 2) слова общелитературного языка, которые в научно-техническом тексте имеют особые значения (*являться, иметь, изучать, реакция, плечо* и пр.);
- 3) слова, свойственные специализирован-

Таблица 1

Стиль	Сфера коммуникации	
	Сфера специальной коммуникации	Пограничная зона массовой и специальной сфер
Научный и научно-технический	фасет 1	фасет 2
Официально-деловой	фасет 3	фасет 4

ным текстам (научно-техническим, юридическим) и лишь иногда употребляемые за пределами соответствующих им коммуникативных ситуациях (*обеспечение, исследовать, анализировать, рампа, комбинировать, инфракрасный*);

4) терминологические фразеологические выражения (*распространение радиоволн, синтез компонентов, закрыть поддувало, языковедческие вопросы*);

5) терминологическая лексика конкретной сферы деятельности (*маржа, форсунка, шестерня, ритуфель, бугшприт*);

6) слова общелитературного языка, не свойственные языку специальных текстов, но употребляющиеся в них, поскольку их содержание становится предметом научного рассмотрения (*ледник, сокол* и пр.);

7) схемы, символы, математические формулы и т.п.

Выделенные лексические разряды традиционно объединяются в три обобщающих группы: терминологическая лексика, общенаучная лексика, общелитературная лексика. Тесное взаимодействие этих трех групп лексических единиц не только в рамках одного текста, но и в рамках языка в целом определяет основные трудности при работе с программами МП. Как показывает анализ переводов, выполненных с помощью систем МП, основной процент переводческих ошибок встречается именно в сфере лексики – самом динамичном уровне языка. Ошибки, как правило, связаны с неправильным переводом терминов и общенаучной лексики, причем количество последних значительно преобладает [Борисова Л.И. 2005, 5-7]. Появление таких ошибок обусловлено применением единых принципов обработки разных текстов в процессе автоматизированного перевода. В связи с этим разработка классификации текстов должна учитывать лексическое разнообразие объектов перевода.

Практика работы с системами МП показывает, что качество перевода терминов в них довольно высокое. Основные трудности здесь могут быть связаны с анализом следующих групп терминов:

- терминов-омонимов, относящихся к разным терминосистемам, например *mouse* – *мышь* (в биологии); *мышка* (деталь компьютера);

- терминов-синонимов, близких по значению и встречающихся в одной терминосистеме; их часто приходится переводить одним термином, например *reservoir* и *storage pond* – *водохранилище*;

- терминов, встречающихся в одной тер-

миносистеме, но имеющих разные значения, зависящие от контекста; например, слово *state* в политической терминологии США может значить как «государство», так и «штат».

Однако большинство этих проблем снимается на этапе создания словаря, и на практике мы уже работаем с готовым продуктом. Отличие терминов от других видов лексики заключается в их относительной стабильности и специализированности терминосистем, что позволяет использовать узкоспециализированные машинные словари по определенной тематике – финансовый, медицинский, юридический, строительный, военный, морской и т.п. – значительно повышающие качество выходного текста (организация машинных словарей, в том числе и специализированных, рассматривается нами в 3 главе данного исследования). Слова-термины имеют идентичные принципы перевода независимо от типа текста, поэтому данный лексический класс, на наш взгляд не может служить достаточным основанием для дальнейшей дифференциации текстов, кроме как по содержательному-предметному признаку, что не является актуальным для целей МП.

Иначе дело обстоит с так называемой общенаучной лексикой. Несмотря на составляющий корень данного термина элемент «-научный», данная группа лексики, помимо собственно научно-технических текстов, активно используется в большинстве текстов официально-делового стиля. Необходимость вычленения данной лексической группы и введения термина «общенаучный» применительно к лексике была обусловлена общественно-языковой практикой. По мнению Л.И. Борисовой, общенаучный слой лексики играет важную роль в структурной и смысловой организации текстов специальной коммуникации. «Общенаучные слова являются средством обозначения интердисциплинарных понятий, а также связующим и организующим звеном лексического состава научно-технических текстов» [Борисова Л.И. 2005, 6].

Комплекс проблем, возникших в связи с использованием и переводом общенаучной лексики, Ю.В. Рождественский объясняет ее быстрым и хаотичным развитием, которое не находит отражения в современной лексикографии. Его слова о том, что в словарном хозяйстве «пока нет места специальному сбору, анализу и описанию общенаучной лексики» [Рождественский Ю.В. 1984, 22], справедливо описывают современную ситуацию. Опыт практического применения современных систем МП показывает, что

генеральные словари (содержащие общеупотребительную лексику) и специализированные терминологические словари (предназначенные для расширения словарной базы системы и настройки на определенную предметную область) не содержат переводных эквивалентов, необходимых для передачи значений многих общенаучных английских слов на русский язык. Традиционные двуязычные словари не отвечают реальным потребностям МП, так как они создаются на базе текстов, преимущественно, художественной литературы и публицистики и отражают общелитературную семантику этих слов.

На сегодняшний день какие-либо обобщающие работы, содержащие комплексный анализ общенаучной лексики, отсутствуют. Можно говорить только о частных исследованиях, рассматривающих самые общие вопросы функционирования общенаучных слов в языке. Это работы Ю.В. Рождественского, М.М. Глушко, Н.А. Шубовой, Н.М. Разинкиной и некоторых других исследователей. Одна из главных задач, решаемых в этих исследованиях, – стратификация лексики текстов специальной коммуникации (преимущественно – научно-технической). Здесь можно отметить два основных подхода: 1) классификация лексического состава на две группы – термины и слова-нетермины (последние имеют различные названия в разных работах: нетерминологическая лексика, общая научно-техническая лексика и т.п.); 2) разделение лексического состава на три группы – общеупотребительные слова, общенаучные слова и термины. Сопоставление списков слов-нетерминов, представленных исследователями, показывает, что они совпадают: одни и те же слова называются общенаучной или общеупотребительной лексикой. Сопоставление слов общей лексики в словаре Э.Г. Жигачевой [Жигачева Э.Г. 1978] со списками слов-нетерминов, проведенное Л.И. Борисовой [Борисова Л.И. 2005], показывает, что общая лексика включает слова, которые могут употребляться в их общенаучном и общеупотребительном значениях.

В связи с этим наиболее логичным представляется первый подход к стратификации лексики специальных текстов – деление слов на термины и слова-нетермины. При этом последняя группа включает в себя 2 подкласса слов. В первый подкласс входят слова относительно узкой семантики, имеющие общенаучное значение, независимо от сферы использования, но употребляемые преимущественно в рамках сферы специальной коммуникации. Этот подкласс составляет ядерную

часть (каркас, основу) слов-нетерминов. В его состав входят лексические единицы, обозначающие общенаучные междисциплинарные понятия, – наиболее важные для сферы специальной коммуникации полнозначные слова, непосредственно описывающие объекты исследований, разработок и их использование. Они образуют основу всего слоя лексики специальной коммуникации. Как правило, это отобранные кругом специалистов слова, привычные для них, позволяющие им не задумываться над способом выражения мысли, обозначающие основные общенаучные понятия и понятия общенаучного обихода (*application, development, implement, extension, generate, equipment, analysis, function, support, requirement* и т.п.). Такие слова имеют синонимичные лексические единицы, используемые коммуникантами вне сферы специальной коммуникации. Например: *application – use* (использование, применение), *development – working out \ improvement* (разработка, улучшение), *implement – carry out \ fulfill* (выполнять, осуществлять), *extension – spreading \ enlargement \ widening* (расширение, распространение), *generate – produce \ take* (создавать, делать), *contender – opponent \ enemy* (оппонент, противник) и т.д. Особенностью этого лексического пласта текстов является возможность свести все вероятные переводные эквиваленты к относительно небольшому объему значений – семантическому ядру. Если говорить о традиционном переводе, осуществляемом человеком, то, например, для слова *generate* с учетом лексических и синтаксических трансформаций всего предложения можно выделить 5 переводных эквивалентов: *получать, создавать, обусловить, выдвигать (требование), формировать*. Однако если речь идет об автоматизированном переводе, то в целях улучшения качества выходного текста в словарную статью можно включить один или два эквивалента, являющихся семантическим ядром, – *создавать*. Например:

Главная задача в данном случае – сохранить основной смысл высказывания на выходе. На наш взгляд, такой подход значительно облегчит редактирование (если оно понадобится). Определение переводного эквивалента (эквивалентов) должно учитывать: частотность его употребления; соответствующие семантические изменения, происходящие в связи с развитием языка, используемого в конкретных сферах деятельности (учет особенностей стиля и его нормативных требований как в языке оригинала, так и в языке перевода).

Текст оригинала	Перевод, выполненный вручную	Автоматизированный перевод
<p><u>Generate (создавать)</u> 1) We note that $W=0$ cannot be <i>generated</i> in the transducers. 2) Complex filters can be <i>generated</i> as holographic filters in the usual way</p>	<p>1) Следует отметить, что преобразователи не позволяют <i>получить</i> составляющую с $W=0$ 2) Сложные фильтры можно <i>формировать</i> в виде голограмм обычным способом</p>	<p>1) Мы отмечаем, что $W=0$ не может быть <i>создана</i> в преобразователях 2) Сложные фильтры могут быть <i>созданы</i> в виде голографических фильтров обычным способом</p>
<p><u>Develop (развивать \ разрабатывать):</u> 1) How writers in nonprofit organizations <i>develop</i> grant proposals 2) We need to <i>develop</i> our HR to increase global efficiency</p>	<p>1) Как в некоммерческих организациях <i>создаются</i> предложения на выполнение гранта 2) С целью повышения глобальной эффективности мы должны <i>совершенствовать</i> нашу кадровую службу</p>	<p>1) Как авторы в некоммерческих организациях <i>разрабатывают (развивают)</i> предложения на выполнение гранта 2) Мы должны <i>разрабатывать (развить)</i> свой HR-отдел, чтобы увеличить глобальную эффективность</p>

К этому подклассу также можно отнести слова-нетермины, являющиеся строевыми элементами языка специальной коммуникации. К таковым относятся преимущественно служебные лексические элементы – предлоги, союзы, артикли, вспомогательные глаголы, местоимения и т.п. Они образуют лексический фундамент языка в целом, и их значение не сильно варьируется в зависимости от сферы употребления.

Второй подкласс образуют слова общелитературного языка, значения которых варьируются в зависимости от сферы употребления, т.е. слова, реализующие, в зависимости от тематической направленности текста, его содержания, адресата, назначения и жанровой принадлежности, либо общелитературное значение, либо общенаучное. Это самые разнообразные лексические средства, которые, наряду с ядерной лексикой, представленной выше, участвуют в описании и характеристике различных процессов и объектов специальной деятельности. Данный слой представлен различными частями речи (преимущественно знаменательными). Эти слова являются неотъемлемой составляющей специальной речевой деятельности, однако не определяют ее так, как ядерная лексика. Данная группа представляет собой более открытую систему по сравнению с ядерной лексикой и является наименее таксономичной. Открытость проявляется не только в постоянном пополнении количественного состава, но в изменении семантического наполнения.

Специфика слов общелитературного языка в текстах специальной коммуникации обусловлена рядом факторов. Это, прежде всего, – изменение предметной соотнесенности. Наряду с характеристикой непосредственно человека и его ежедневного быта в сферах неспециальной коммуникации, в рамках специальной коммуника-

ции эти слова стали характеризовать предметы, процессы и явления, относящиеся к конкретной сфере деятельности, что привело к появлению новых значений и обособлению их от значений, присущих общелитературному языку. Выделение новых значений связано с изменением функционально-стилистической сферы употребления слова – жесткая структура специальных текстов, их нормы и требования, смена сочетаемости, изменение синтаксической функции в предложении – все это отражается на семантике слов общелитературного языка в специальной коммуникации. Так, например, слово *exciting* имеет общелитературные значения: 1) *волнующий*, 2) *захватывающий*, 3) *восхитительный* 4) *увлекательный* (*an exciting story, an exciting life* и т.п.); однако его употребление в сфере специальной коммуникации привело к обособлению таких значений, как: 1) *перспективный* 2) *важный* 3) *широкий* 4) *интересный* (*exciting development, exciting result, exciting aspect* и т.п.).

Во-первых, к этому подклассу можно отнести слова, которые в текстах специальной коммуникации редко реализуют свои общелитературные значения и стремятся к значительной специализации. Так, например, слово *chore* имеет общелитературное значение «рутинная работа», однако в научно-техническом тексте имеет регулярные переводные эквиваленты *задача, работа, функция*; или английское слово *advantageous* имеет общелитературные значения *благоприятный \ выгодный \ полезный*, но в текстах специальной коммуникации реализуются значения *целесообразный \ перспективный*.

Во-вторых, слова с очень широкой семантикой типа *feature, associate, support, predict, unique, available, handle, beauty, birth, delicate, headache, infancy, niche, horizon, striking, due to, in the con-*

text of, in terms of и т.п. Такие слова имеют много переводных эквивалентов, причем наряду с устойчивыми эквивалентами реализуется широкий спектр нерегулярных эквивалентов. Так, например, Л.И. Борисова отмечает около 30 новых устойчивых эквивалентов для глагола *handle*, отсутствующих в БАРСе, у существительного *contribution* – 14, у прилагательного *meaningful* – 35 эквивалентов [Борисова Л.И. 2005].

В-третьих, стилистически окрашенные лексические средства. Эта группа слов также реализует широкую семантику и включает как регулярные средства для выражения субъективной авторской оценки (*dramatic, dramatically, extremely, excellent, exciting, fascinating, remarkably, remarkable, smartly* и т.п.), так и индивидуальные лексические единицы, такие, как образное сравнение, метонимия, метафора.

Автоматизированная обработка именно этого подкласса слов-нетерминов вызывает наибольшие трудности. В отличие от первого подкласса, для таких слов нельзя определить один или два переводных эквивалента, определяющих их семантическое ядро, из которых на этапе постредактирования можно выбрать наиболее подходящий. Так, например, глагол *handle* только в рамках специальных текстов может иметь более 15 эквивалентов (например: *управлять, оперировать чем-л., манипулировать, перемещать, перекладывать, транспортировать, обрабатывать, работать с чем-л., обращаться с чем-л.* и т.д.). Весь комплекс значений здесь можно разграничить на две категории – общенаучные значения и общелитературные значения, определяемые спецификой текста. Так, например, для слово *favourable* реализует общелитературные значения *благоприятный, благоприятствующий, подходящий, годный, удобный (favourable balance – активный/благоприятный платежный баланс, favourable circumstance – благоприятные обстоятельства, favourable event – благоприятствующее событие, favourable outcome – подходящий результат, favourable terms – благоприятные условия)*; однако в научно-технических текстах ему соответствуют переводные эквиваленты *хороший, удовлетворительный, выгодный, высокий, удачный (favorable ballistic coefficients – хорошие аэродинамические показатели, favorable strength – высокая прочность, favorable result – удачный результат)*. Проблема перевода этой группы лексики должна решаться путем использования контекстологических словарей.

Наличие подобного разнообразия лексических групп обуславливает необходимость рассмотр-

ения каждого типа текстов с точки зрения качественного и количественного функционирования основных лексических классов. Рассмотрение лексического состава в качестве дифференциального признака обусловлено проблемами практического использования современных систем МП: слова и словосочетания являются главным объектом автоматизированного перевода; большинство ошибок связано именно с неправильным подбором переводных эквивалентов, что вызвано некорректной стратификацией лексики в словарях систем МП. Современные коммерческие программы МП предполагают самое общее деление лексического фонда текстов, что отражается на принципах организации словарей: как правило, предлагается только два вида – специализированный (содержит словарные статьи по терминам) и генеральный словарь (содержащий общеупотребительную лексику, представленную общелитературными переводными эквивалентами). При таком подходе на практике не учитывается более половины реализуемых значений слов-нетерминов, что ведет к увеличению затрачиваемого времени на редактирование конечного продукта, а порой – к отказу от использования СМП.

Разработка классификация текстов на основе ситуативной и лексической дифференциации определяет дальнейшие направления исследований в данной области. Помимо чисто теоретической значимости, она имеет и прикладное значение. Данная классификация определяет: 1) организацию и структуру машинных словарей, их взаимодействие; 2) принципы редактирования конечного продукта.

ЛИТЕРАТУРА:

1. Борисова Л.И. Лексические особенности англо-русского научно-технического перевода. Теория и практика перевода.: Уч. пос. – М.: НВИ-ТЕЗАУРУС, 2005. – 216 с.
2. Жигачева Э.Г. Системная организация общей лексики / На материале английских текстов по водному транспорту: Дис. ... канд. филол. наук. – Л., 1978. – Т. 1 – 136с.; Т. 2. – 272 с.
3. Марчук Ю.Н. Компьютерная лингвистика: учебное пособие. – М.: АСТ: Восток – Запад, 2007. – 317 с.
4. Нариньяни А.С. Проблема понимания ЕЯ-запросов к базам данных решена. – В кн.: DIALOG '95. Труды Международного семинара по компьютерной лингвистике и ее приложениям. – Казань, 1995. – С. 206-215.
5. Рождественский Ю.А., Сычев О.А. Общенаучная лексика в автоматизированном переводе / Междунар. форум по информатике и документации. – 1984. – Т. 9 – № 2. – С. 22-25.
6. Рождественский Ю.В., Волков А.А., Марчук Ю.Н. Введение в прикладную филологию: Языковая семиотика – М.: МГУ, 1988. – 116 с.
7. Штиндлова Й. Обратные словари // Автоматизация в лингвистике. – Л.: Наука, 1966. – С. 84-88.