

На правах рукописи

УДК 05.808.2.002

БАБУШКИНА НАТАЛЬЯ ВЛАДИМИРОВНА

**ИССЛЕДОВАНИЕ РЕЗУЛЬТАТОВ
МАШИННОГО ПЕРЕВОДА ГЕРУНДИЯ:
ОСОБЕННОСТИ АНАЛИЗА
И КРИТЕРИИ РЕДАКТИРОВАНИЯ**

Специальность: 10.02.21 — прикладная и математическая лингвистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата филологических наук

Санкт-Петербург

2007

Работа выполнена на кафедре прикладной лингвистики
государственного образовательного учреждения
высшего профессионального образования
«Российский государственный педагогический университет
им. А. И. Герцена»

Научный руководитель: доктор филологических наук, профессор
БЕЛЯЕВА Лариса Николаевна

Официальные оппоненты: доктор филологических наук, профессор
ФИЛИМОНОВА Ольга Евгеньевна

кандидат филологических наук, доцент
КОВАЛЬ Сергей Аркадьевич

Ведущая организация: Государственное образовательное учреж-
дение высшего профессионального обра-
зования Тольяттинский государственный
университет

Защита состоится «_3_» __апреля_____ в 15.30 час. на засе-
дании диссертационного совета Д 212.199.17 по защите диссертаций на соис-
кание ученой степени доктора наук в Российском государственном педагоги-
ческом университете им. А. И. Герцена по адресу: 191186, г. Санкт-
Петербург, наб. р. Мойки, д. 48, корп. 14, ауд. 314.

С диссертацией можно ознакомиться в фундаментальной библиотеке
Российского государственного педагогического университета им. А. И. Гер-
цена.

Автореферат разослан «___» _____ 2007 г.

Ученый секретарь
диссертационного совета
доктор филологических наук,
профессор

Л. А. Пиотровская

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Современный уровень развития техники, информационных технологий и средств коммуникации обеспечивает доступ к большим объемам многоязычной информации. Ввиду постоянного расширения и усложнения информационных потоков вопрос о скорости и качестве их обработки стоит особенно остро. Обработка «вручную» связана со значительными временными и материальными затратами, поэтому в настоящее время невозможно обойтись без систем, обеспечивающих переработку информации, — лингвистических автоматов различных типов. В реферируемой работе рассматриваются системы, предназначенные для одного из видов обработки текста — перевода: системы в помощь переводчику и системы машинного перевода (МП).

Актуальность исследования. Сегодня существует целый ряд систем в помощь переводчику, которые не выполняют перевод всего текста: электронные словари, системы переводческой памяти, программы обеспечения и поддержки локализации и др. Эти системы чрезвычайно популярны в переводческом сообществе, успешно применяются для облегчения и ускорения работы, однако не решают вопроса автоматического получения перевода — для этих целей предназначены системы МП.

Создание системы МП, осуществляющей перевод, качество которого соответствует профессиональным потребностям специалистов и переводчиков, остается пока недостижимым по ряду причин: технических, кибернетических, лингвистических. Разработки, направленные на улучшение качества работы систем МП, ведутся постоянно: осуществляются попытки слияния систем машинного перевода и систем переводческой памяти, ограничивается тематическая область, расширяются автоматические словари. Тем не менее, переводы, выполняемые современными системами МП, требуют редактирования, при этом для решения сложных случаев редактору необходимо знание и языка перевода, и исходного языка, поскольку анализ исходного текста / предложения приходится проводить заново.

Таким образом, особой проблемой, возникающей при реальном использовании компьютерных систем в практике перевода, является установление принципов и приемов редактирования результатов машинного перевода, выявление диагностических признаков, по которым редактор должен обращаться к исходному тексту и проверять синтаксическую и семантическую корректность переводного текста. Установление таких «контрольных точек» возможно только на основе сравнительно-статистического изучения большого объема текстов на исходном языке и их машинных переводов. Такой подход позволяет выделить сложные в сопоставительном аспекте синтаксические явления и конструкции, модели их перевода, наметить возможные изменения действующих алгоритмов перевода, выявить те проблемные конструкции, на перевод которых необходимо обращать внимание при редактировании.

В силу названных причин изучение возможного упрощения процесса редактирования и создание максимального удобства пользования действующими системами МП является актуальным.

Реферируемая работа посвящена исследованию возможностей упрощения редактирования и улучшения качества машинного перевода на русский язык на примере герундия — неличной формы английского глагола. Дело в том, что как при ручном, так и при машинном переводе этой формы часто возникают ошибки и неточности. Трудности в обработке герундия объясняются его многовалентностью (он сочетает в себе валентности глагола и существительного), широким использованием в различных функциях, наличием в английском языке нескольких омонимов герундия, а также отсутствием в русском языке подобной глагольной формы.

Теоретической основой исследования послужили следующие положения:

1) создание системы высококачественного профессионального машинного перевода, не требующего редактирования, в настоящее время невозможно (Л. Н. Беляева, Р. Г. Пиотровский, W. J. Hutchins);

2) постредактирование текста машинного перевода представляет собой необходимый этап в процессе осуществления профессионального перевода (J. Allen, W. J. Hutchins);

3) герундий, являясь своего рода уникальной многовалентной формой английского языка, сложен для перевода и распознавания в тексте (Л. Г. Верба и Г. В. Верба; К. А. Гузеева и С. И. Костыгина; Е. А. Корнеева; Е. А. Зверева, И. К. Лихачева и Н. С. Щукарева) и создает особые трудности при машинном переводе.

Объектом исследования является герундий английского языка, его функционирование в научно-техническом тексте и особенности его перевода на русский язык с помощью систем машинного перевода, а **предметом** — способы парсинга (автоматического синтаксического анализа) и модели машинного перевода герундия.

Гипотеза исследования — контрастивный анализ текстов и их машинных переводов позволяет выделить словоформы и конструкции, вызывающие сложности для алгоритмизации при машинном переводе, и установить необходимые изменения алгоритмов перевода, а также определить конструкции, трудности в переводе которых неразрешимы на уровне алгоритмизации и могут быть устранены только редактированием.

В соответствии с этим **целью** диссертационной работы является разработка рекомендаций по редактированию перевода герундия, а также возможных изменений алгоритмов машинного перевода герундия на основе изучения моделей его употребления в современном специальном тексте и моделей его машинного анализа и перевода.

Для достижения указанной цели в диссертации решаются следующие задачи:

1) классификация современных систем в помощь переводчику, изучение общих принципов их работы и возможностей рационального использования систем для оптимизации переводческого процесса;

2) определение особенностей организации систем машинного перевода как систем, производящих лингвистический анализ текста в полностью автоматизированном режиме;

3) исследование герундия, неличной формы английского глагола, как одной из наиболее сложных форм для перевода на русский язык: выявление дистрибуции и валентностей герундия, изучение омонимии герундия и других *ing*-овых форм и категорий английского языка, а также определение моделей функционирования герундия в специальном тексте;

4) изучение реализованных в системе МП моделей машинного анализа и перевода герундиальных употреблений, установление причин возникновения ошибок и неточностей анализа герундия и их влияния на качество перевода;

5) выделение особо сложных для машинного анализа и перевода герундиальных конструкций, определение возможных изменений алгоритмов системы МП, а также критериев редактирования машинных переводов и составление рекомендаций по написанию текстов, предназначенных для перевода с помощью системы МП.

Материалом исследования послужили тексты по сейсмологии, изданные в 1994–2000 гг., общим объемом около 260 000 словоупотреблений. По тематике указанные материалы относятся к специальным научно-техническим текстам и являются примерами реальных современных текстов. Методом сплошной выборки из указанных текстов выделялись употребления *ing*-овых омонимов: самого герундия (около 1350 употреблений), а также других *ing*-овых форм и категорий, оказавшихся проблемными для МП из-за сложности разрешения омонимии с герундием. Все выделенные примеры (на уровне предложения) в пакетном режиме были переведены системой МП. Основной системой МП для проведения анализа послужила система SILOD-Windows в виде библиотеки прикладных программ WORD+, разработанная в лаборатории машинного перевода РГПУ им. А. И. Герцена.

Методы исследования. Собранный языковой материал был проанализирован с помощью методов синтаксического, семантического, статистического и контрастивного анализа, а также методов компьютерного моделирования.

Научная новизна исследования заключается в том, что:

1) создано комплексное описание функционирования герундия в современном научно-техническом тексте;

2) впервые разработана типология герундиальных конструкций, представляющих сложность для систем МП;

3) впервые систематизированы ошибки и неточности, возникающие при машинном переводе герундия;

4) разработаны рекомендации по их устранению и редактированию.

На защиту выносятся следующие положения:

1. Контрастивный анализ исходных текстов и их машинных переводов, осуществляемый на разных уровнях реализации, является инструментом исследования дистрибуции лингвистических явлений и особенностей их автоматического анализа и перевода.

2. Герундий как многофункциональная и многозначная форма английского глагола представляет особую сложность для машинного анализа на всех уровнях и, соответственно, для машинного перевода.

3. В тексте машинного перевода следует различать ошибки и неточности. Неточности связаны со стилистической некорректностью переводного предложения, они не препятствуют его пониманию, но требуют редактирования при создании текста в издательских целях. Ошибки возникают в случае неправильного определения грамматического статуса конкретных форм, препятствуют пониманию текста, и для их редактирования необходимо производить анализ исходного предложения заново.

4. Ошибки и неточности возникают на всех уровнях машинного анализа текста. Способы установления и исправления ошибок индивидуальны для каждого уровня. На основе выявления основных ошибок и неточностей в представительной выборке текстов могут вноситься изменения в действующие алгоритмы системы МП и словарные статьи автоматического словаря.

5. Установление критериев редактирования машинных переводов и составление рекомендаций редактору позволяет ускорить и упростить процесс редактирования.

Теоретическая значимость исследования заключается в том, что:

1) комплексное описание герундия, включающее его теоретическое исследование и рассмотрение функционирования в современном специальном тексте, способствует дальнейшему развитию алгоритмов парсинга для машинного перевода;

2) установлена необходимость и целесообразность анализа особенностей перевода на различных уровнях реализации системы машинного перевода, что способствует дальнейшему развитию теории машинного перевода и вносит вклад в переводоведение;

3) определена целесообразность интеграции систем машинного перевода и систем переводческой памяти в аспекте упрощения редактирования и уменьшения его объема;

4) предложенная методика сопоставительного исследования текстов на основе применения системы машинного перевода вносит вклад в методологию современных контрастивных исследований.

Практическая значимость исследования состоит в разработке рекомендаций по вопросу о методах установления и устранения ошибок при ре-

дактировании МП, а также комплекса изменений в действующие алгоритмы МП и рекомендаций по написанию текстов, предназначенных для последующего машинного перевода. Результаты исследования могут быть использованы для улучшения качества работы действующих систем машинного перевода с английского языка на русский, в подготовке редакторов текстов МП, а также при обучении переводу, поскольку многие случаи, вызывающие сложности при МП, также сложны и для переводчиков.

Рекомендации по использованию результатов исследования. Выделенные в тексте сложные модели употребления герундия могут быть использованы при обучении переводчиков, а также редакторов текстов. Рекомендации по устранению ошибок могут найти применение в работе не только с системой SILOD-Windows, но и с другими системами перевода с английского языка на русский.

Кроме того, проведенный обзор современных систем МП может быть использован в качестве справочного материала при выборе системы для использования в условиях профессиональной переводческой службы или для личного использования. Эта же справочная информация может найти применение в курсах обучения машинному переводу.

Описание герундия может быть использовано в курсе обучения теоретической и практической грамматике английского языка, а массив примеров употребления и перевода герундия — при усовершенствовании систем машинного перевода специальных текстов, в частности, их лексической базы.

Апробация работы. Основные положения диссертации обсуждались на международной конференции «Прикладная лингвистика без границ» (Санкт-Петербург, март 2004 г.), международной конференции «Прикладная лингвистика в науке и образовании» (Санкт-Петербург, март 2006 г.), межвузовской научно-практической конференции «Язык и межкультурная коммуникация» (Санкт-Петербург, май 2004 г.), а также на заседаниях кафедры прикладной лингвистики РГПУ им. А. И. Герцена. По теме диссертации опубликованы 5 работ общим объемом 1 п. л.

Объем и структура работы. Диссертация состоит из введения, трех глав, заключения, списков принятых сокращений, использованной литературы (229 наименований, из них 156 — на английском языке), словарей (2), справочных изданий (7) и 7 приложений. Основные научные результаты работы представлены в 25 таблицах. Общий объем работы — 242 стр., из них 176 стр. — основное содержание и 66 стр. — приложения.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обоснован выбор темы исследования и ее актуальность, определены теоретическая основа, цель, задачи, объект, предмет, гипотеза, материал и методы исследования, раскрыта его научная новизна, сформулированы положения, выносимые на защиту, обоснована теоретическая и

практическая значимость, даны рекомендации по использованию результатов исследования, отражена апробация работы.

В **первой главе** диссертации «Современные технологии в помощь переводчику» излагаются основные предпосылки исследования. Изучается классификация современных компьютерных систем в помощь переводчику, рассматриваются системы МП, производящие собственно лингвистический анализ текста. Изучаются проблемы оценки качества МП и рассматриваются различные подходы к оценке качества, определяется роль редактирования в процессе машинного перевода. Кроме того, приводится обзор современных компьютерных систем в помощь переводчику.

На протяжении последних двадцати лет были разработаны основы типологии компьютерных систем в помощь переводчику, которая принимается практически всеми разработчиками и исследователями (А. Н. Баранов, Л. Н. Беляева, W. J. Hutchins, H. Somers, S. Nirenburg): системы классифицируются по количеству языков, степени автоматизации и области применения. Однако нельзя говорить о том, что вопрос установления типологии этих систем решен окончательно, поскольку она начала складываться тогда, когда создание системы МП, способной полностью заменить человека-переводчика и дающей высококачественный перевод, считалось основной целью разработок в области автоматизации перевода. Учитывая тот факт, что создание таких систем МП на сегодняшний день остается пока недостижимым (Л. Н. Беляева, Р. Г. Пиотровский, D. Arnold) и переводчик всегда участвует в создании окончательного текста перевода, представляется целесообразным вместо классификации по степени автоматизации использовать следующий критерий классификации систем: проведение (или не проведение) системой лингвистического анализа текста.

По количеству языков системы делятся на бинарные и многоязычные, по области применения — на домашние, профессиональные, системы для перевода интернет-страниц и электронной почты, для перевода звучащей речи и т. п. (Compendium of Translation Software).

В зависимости от того, производит ли система лингвистический анализ текста на исходном языке и его преобразование на язык перевода (при этом необязательно высококачественное), компьютерные системы в помощь переводчику разделяются на:

- вспомогательные системы поддержки перевода,
- системы МП (с возможностью редактирования перевода).

В состав вспомогательных систем поддержки перевода включаются системы, программы и утилиты, направленные на ускорение и упрощение процесса перевода, но лингвистический анализ и собственно перевод текста не выполняющие. К системам такого типа относятся:

- 1) электронные словари;
- 2) системы создания и ведения терминологических баз данных;
- 3) системы переводческой памяти;
- 4) средства обеспечения и поддержки локализации.

Первые два типа систем предназначены для работы с терминологией. Электронный словарь представляет собой уже готовую базу лексических единиц, снабженную механизмом поиска, а системы создания и ведения терминологических баз данных дают возможность пользователю или группе пользователей создавать специализированные базы терминов для определенного проекта или для текстов определенной тематики.

Системы переводческой памяти предназначены для создания, пополнения и ведения двуязычных баз данных (БД) ранее переведенных текстов и осуществления поиска в этих БД. Система переводческой памяти анализирует новые тексты на предмет выявления в них ранее переведенных фрагментов (чаще предложений), хранящихся в базе данных, чтобы предоставить переводчику уже готовый перевод предложения, части предложения или целого отрывка, если он уже был однажды переведен и хранится в БД. Такие системы позволяют избавить переводчика от повторной рутинной работы, а также обеспечить преемственность используемой терминологии и синтаксических конструкций при исправлении документов и переиздании их в новой редакции.

Системы обеспечения и поддержки локализации помогают в языковой и культурной адаптации программного продукта (а также некоторых интернет-систем) для иноязычных пользователей, т. е. для работающих на иностранных языках, отличных от национального языка, для которого изначально писалась программа.

Системы машинного перевода — единственный тип систем, производящих собственно лингвистический анализ текста; входными единицами являются полные предложения на одном естественном языке, а выходными — соответствующие полные предложения на другом языке.

Как упоминалось выше, такие системы не создают перевод, равный по качеству профессиональному, но с их помощью можно значительно ускорить и удешевить сам процесс перевода, если подходить к вопросу оценки качества и адекватности перевода с точки зрения задач и требований заказчика или пользователя. Дело в том, что для извлечения общей информации из текста специалисту достаточно даже грубого МП, а для создания текста издательского качества такой перевод обязательно потребует редактирования (FEMPI, W. J. Hutchins). При этом редактирование должно занимать меньше времени, чем ручной перевод, иначе использование системы МП становится невыгодным. Следовательно, редактор должен быть подготовлен к работе с результатами МП так, чтобы редактирование из процесса быстрого исправления ошибок не превратилось в переделывание перевода.

Одной из составляющих подготовки редакторов для работы с системами МП должно быть изучение случаев, сложных для алгоритмизации, и наиболее типичных ошибок машинного перевода. Установление и описание сложных для МП единиц и конструкций не только укажет редактору, на что необходимо обращать внимание в тексте перевода, но и позволит определить те изменения, которые нужно внести в алгоритмы перевода и словарные ста-

тью автоматических словарей (АС), поскольку именно неточность или недостаточность информации в АС часто приводит к ошибкам в переводе.

Наибольшие сложности для анализа и перевода вызывают словоформы и конструкции, для распознавания которых человек-переводчик обращается к общей теме текста и контексту словоформы, в то время как система МП анализирует лишь ближайший грамматический контекст и работает по предложениям (т. е. рассматривает текст как набор не связанных между собой предложений), поэтому результаты анализа (парсинга) одного предложения не распространяются на последующие и приходится повторять анализ уже разобранных ранее конструкций. К таким «проблемам для МП» относятся длинные сложные и неполные предложения, обороты, именные группы, омонимы и т. д. Особые сложности вызывает перевод *ing*-овых омонимов.

Следует также отметить, что наибольшего полезного эффекта от систем МП можно добиться при использовании их в комбинации с несколькими системами в помощь переводчику. Такой комплекс нескольких переводческих программ и утилит, интегрированных в одну систему, образует автоматизированное рабочее место переводчика (W. J. Hutchins).

Во **второй главе** «Герундий в английском языке: образование, функции, дистрибуция» рассматривается глагольная форма герундия как одна из наиболее сложных форм для машинного анализа и перевода. Изучаются морфологические и синтаксические свойства герундия, непосредственно влияющие на результат его машинного анализа. Описываются результаты исследования выборки современных научно-технических текстов на предмет определения моделей функционирования герундия в текстах такого типа. Рассматривается омонимия герундия и причастия как наиболее сложный тип омонимии *ing*-овых форм для машинного разрешения.

Герундий — наиболее своеобразная неличная форма в системе английского глагола. Он сочетает в себе свойства существительного и глагола и передает комплексные отношения, во многих случаях выражаемые в других языках придаточным предложением. В русском языке нет глагольной формы, соответствующей герундию.

Герундий сложен для машинного перевода по ряду причин.

Во-первых, он обладает широкой дистрибуцией. Благодаря наличию именных (способность вводиться предлогом, определяться притяжательным местоимением и существительным в притяжательном падеже) и глагольных (способность принимать беспредложное дополнение) валентностей герундий используется в большом наборе конструкций, при этом в одном и том же синтаксическом окружении он может выполнять различные синтаксические функции. Поэтому однозначное определение роли герундия в предложении и синтаксических связей с другими членами предложения может быть для системы МП затруднительным.

Во-вторых, в английском языке существует целый ряд частей речи, образующихся с помощью морфемы *-ing*: отглагольное существительное, прилагательное, предлог, союз, а также глагольные формы: причастие I, инфини-

нитив в форме Continuous и герундий. Человек разрешает подобную омонимию на основе своего тезауруса, собственных фоновых знаний о внешнем мире и предметной области, обращаясь не только к ближайшему контексту словоформы, но и к общему контексту, рассматривая текст в целом. Система МП производит грамматический анализ лишь в рамках предложения, базируясь на ближайшем окружении каждой словоформы. В такой ситуации разрешение омонимии часто является проблематичным.

В-третьих, отсутствие герундия в русском языке и многообразие конструкций, в которых герундий употребляется, требуют разработки индивидуальных моделей перевода для конкретных употреблений, поскольку универсального способа перевода герундия на русский язык не существует.

Суммируя все вышесказанное, можно утверждать, что герундий представляет особые сложности для:

1) машинного анализа из-за сложности определения статуса герундия (снятия омонимии с другими *ing*-овыми формами), а также определения связей между герундием и его окружением;

2) машинного синтеза из-за отсутствия соответствующей глагольной формы в русском языке.

В результате исследования установлено, что не все *ing*-формы одинаково трудны для машинного разрешения омонимии с герундием. Наиболее сложным является снятие омонимии герундия и причастия, так как их формы совпадают в парадигмах всех глаголов (в специальном тексте используются только формы Indefinite Active), они имеют схожую дистрибуцию, а также чаще других *ing*-форм используются в тексте.

Предлог, а также единичные употребления отрицательного местоимения *no* в препозиции к герундию являются единственным индикатором его отличия от причастия (притяжательное местоимение и существительное в притяжательном падеже в препозиции к герундию в исследованной выборке не зафиксированы), но если предлог отделен от герундия несколькими словами, то система МП может не установить связи между ним и герундием. В остальном в специальном тексте дистрибуция герундия и причастия совпадает.

По данным исследования особые сложности при машинном снятии омонимии вызывают следующие модели употребления герундия и причастия (см. табл. 1). Под моделью употребления понимается минимальная дистрибуция *ing*-овой формы.

В таблице 1 использованы следующие обозначения:

- Ving — *ing*-овая форма,
- Ving(n) — герундий (субстантивное употребление *ing*-овой формы),
- Ving(a) — причастие (атрибутивное употребление *ing*-овой формы),
- N — существительное или местоимение в общем падеже, другая именная категория или именная группа,
- V — глагол-сказуемое,
- p — предлог,

- N_n — порядковый номер элемента N в модели, без учета *ing*-овой формы,
- $()$ — присутствие единицы не обязательно,
- $[]$ — скобки, ограничивающие сегмент модели, элементы которого связаны между собой, а весь сегмент, в свою очередь, связан с остальными элементами модели,
- \leftarrow, \rightarrow — распределение зависимостей внутри модели (группы, сегмента и т. п.), стрелка направлена от главного слова к зависимому.

Таблица 1. Модели употребления герундия и причастия в специальном тексте, вызывающие трудности при машинном разрешении омонимии

Общая дистрибуция	Функция и модели употребления герундия	Омонимичные герундиальным модели употребления и функции причастия
$(N_1)+Ving+(N_2)$	Подлежащее $(N_1)\leftarrow Ving(n)\rightarrow(N_2)$	Обстоятельство (причастный оборот в начале предложения) $(N_1)\leftarrow Ving(a)\rightarrow(N_2)$ Препозитивное определение $Ving(a)\leftarrow N$
$be+(N_1)+Ving+(N_2)$	Составное именное сказуемое $[be+(N_1)\leftarrow Ving(n)]\rightarrow(N_2)$	Форма Continuous $[be+Ving(a)]\rightarrow(N)$ Препозитивное определение $Ving(a)\leftarrow N$
$V+Ving+(N)$	Составное глагольное сказуемое $[V+Ving(n)]\rightarrow(N)$	Обстоятельство $Ving(a)\rightarrow(N)$ Препозитивное определение $Ving(a)\leftarrow N$
$V+(N_1)+Ving+(N_2)$	Беспредложное дополнение $V\rightarrow[(N_1)\leftarrow Ving(n)\rightarrow(N_2)]$	Постпозитивное определение $N_1\rightarrow Ving(a)\rightarrow(N_2)$ Препозитивное определение $Ving(a)\leftarrow N$ Обстоятельство $Ving(a)\rightarrow(N)$
$p+(N_1)+Ving+(N_2)$	Предложное дополнение Обстоятельство Постпозитивное определение $p+[(N_1)\leftarrow Ving(n)\rightarrow(N_2)]$	Постпозитивное определение $N_1\rightarrow Ving(a)\rightarrow(N_2)$ Препозитивное определение $Ving(a)\leftarrow N$ Обстоятельство $Ving(a)\rightarrow(N)$
$Ving+N$	Препозитивное определение $Ving(n)\leftarrow N$	Препозитивное определение $Ving(a)\leftarrow N$ Постпозитивное определение (только при наличии N_2) $N_1\rightarrow Ving(a)\rightarrow N_2$ Обстоятельство (только при наличии N) $Ving(a)\rightarrow N$

Как показывают данные таблицы 1, каждой модели употребления герундия в конкретной синтаксической функции соответствует несколько моделей употребления причастия в различных синтаксических функциях, что вызывает проблемы разрешения омонимии при машинном анализе.

Кроме того, таблица 1 иллюстрирует сходство моделей употребления герундия в различных синтаксических функциях, что также может привести к ошибкам в анализе, в частности — к неправильному определению связей между герундием и его окружением. Например, поскольку герундий может употребляться в функции препозитивного определения, а также принимать беспредложное дополнение, могут возникать сложности при машинном установлении связей в группах типа Ving(n)+N. Герундий может быть главным элементом с беспредложным дополнением (Ving(n)→N) или зависимым — в функции препозитивного определения (Ving(n)←N). При этом формальных показателей роли герундия нет. Таким образом, необходимо изучение реально встречающихся в тексте подобных случаев, сложных для МП, чтобы составить рекомендации по исправлению ошибок в их переводе.

Третья глава «Машинный перевод герундия» посвящена моделям машинного анализа и перевода герундия. Исследуется процедура машинного анализа герундия на всех уровнях преобразования: лексико-морфологическом, уровне групп, функциональных сегментов и предложения. Определяются ошибки и неточности. Предложены рекомендации по внесению изменений в действующие модели парсинга и в автоматический словарь системы, а также рекомендации по редактированию текстов МП и советы авторам текстов, предназначенных для машинного перевода.

Ввиду различного подхода к решению задач редактирования текстов МП в реферируемой диссертации предложено различать неточности и ошибки: неточности связаны лишь со стилистической некорректностью переводного предложения и могут быть отредактированы без обращения к исходному тексту, поскольку в целом перевод понятен; ошибки препятствуют пониманию текста, для их редактирования зачастую необходимо заново анализировать исходное предложение или его часть.

В системе МП анализ выполняется на следующих уровнях: лексико-морфологическом, уровне групп, функциональных сегментов, предложения. Возникновение ошибок возможно на всех уровнях анализа исходного текста, что приводит к неверному преобразованию (трансферу) и, соответственно, к неправильному переводу.

Ошибки лексико-морфологического уровня наиболее редки в переводе герундия и наиболее просты для исправления, поскольку во многих случаях достаточно изменить информацию в АС. К таким изменениям относится пополнение АС неопознанными ранее словами. Кроме того, иногда возникают ошибки в переводе членов омонимического ряда, один из которых используется чаще, чем другие. Например, словоформа *following* определяется системой МП только как прилагательное, хотя в приведенном ниже примере является герундием:

*Rothman solved the prevalent problem of "cycle skips" in linearized methods of residual statics computation of noisy data by **following** the Monte Carlo optimization technique of Metropolis et al. ... —*

‘Rothman решил распространенную проблему «рикошеты цикла» в линейризованный методы остаточного вычисления statics шумных данных путем следующего техники оптимизации в Монте Карло Метрополис и др. ...’.

Для таких случаев какие-либо изменения в алгоритме для улучшения перевода сделать трудно, поскольку невозможно выделить однозначные формальные показатели прилагательного и герундия: у них схожая дистрибуция (предлог в препозиции, существительное в постпозиции и т. п.). Так как герундий *following* используется в тексте значительно реже омонимичного прилагательного, то введение дополнительного «герундиального» перевода будет лишь загромождать текст. Зная о возможной ошибке, редактор может исправить перевод герундия *following*, не обращаясь к исходному предложению, поскольку его значение понятно.

Самое большое количество ошибок парсинга приходится на уровень групп: в английском языке нет развернутой падежной системы и категории грамматического рода, возможна конверсионная омонимия, сильна беспредложная связь и т. п.; иными словами, набор формальных показателей, на которые система МП опирается при анализе групп, сравнительно невелик. Основные сложности возникают при определении границ групп, семантических связей между их элементами и, при наличии омонимов, — грамматического статуса элементов. Анализ выборки показал, что ошибки в определении статуса герундия возникают при переводе простых именных групп, неточности чаще встречаются при переводе групп с предложным управлением.

Под группами с предложным управлением в диссертации понимаются группы, в которых предлог относится к *ing*-форме и находится непосредственно перед ней. В исследованной выборке наиболее часто герундий употребляется с предлогами *by*, *for*, *in* и *of*. В таких группах система МП практически всегда верно определяет статус герундия: поскольку основные трудности вызывает снятие омонимии типа герундий / причастие, при переводе предложных конструкций предлог «работает» как индикатор «не-причастия». Связи внутри группы также устанавливаются системой МП верно, поскольку зависимость между элементами, как правило, линейная: следующий элемент зависит непосредственно от предыдущего, что является наиболее простой схемой для машинного анализа.

В сочетании с некоторыми предлогами (в выборке: *behind*, *into*, *upon* и *on*) перевод герундия может быть неточен и требует редактирования, например:

*The motivation behind **resorting** to the inversion scheme to be dealt with in this paper comes from the absence of even a one-dimensional local velocity model... —*

‘Мотивация за обращением к схеме инверсии, которую нужно рассмотреть в этом докладе исходит из отсутствия даже одномерная местная скоростная модель...’.

Существительное *мотивация* в русском языке не принимает зависимых слов с предлогом *за*, а управляет существительным в родительном падеже без предлога. Поскольку перевод понятен, для редактирования не требуется обращения к исходному предложению.

Под простыми именными герундиальными группами (ИГГ) в диссертации понимаются именные группы, состоящие из двух и более простых элементов, в которых *ing*-форма не управляется предлогом, стоящим непосредственно перед ней, и связи между членами группы не оформлены предлогами. Такие группы различаются по количеству элементов в группе и по позиции, занимаемой герундием. В текстах наиболее распространены двух- и трехкомпонентные группы, группы с пятью и более составляющими встречаются очень редко. При этом с увеличением длины группы возрастает процент ошибочных переводов, так как структура зависимостей между элементами становится более сложной.

В большинстве случаев ядром простой ИГГ является ее последний элемент — это чаще всего не *ing*-форма (за исключением двухкомпонентных групп); герундий употребляется непосредственно слева от ядра, т. е. в предпоследней позиции в группе, а зависимости в группе направлены справа налево: (N)←Ving(n)←n (n — один элемент простой именной группы). Ошибки в переводе чаще всего заключаются в том, что герундий определяется системой МП как причастие; элемент, стоящий непосредственно перед герундием, — как имя, управляющее причастием, и, соответственно, связи в группе — как направленные слева направо: N→Ving(a)→n. В русском языке такой перевод реализуется причастным оборотом:

The aim of our depth imaging methodology is to determine the geometry and the location of fault planes... —

‘Цель нашей глубины, отображающей методологию состоит в том, чтобы определить геометрию и расположение плоскостей сброса...’.

Система МП правильно производит анализ лишь некоторых двухкомпонентных групп, а также тех трехкомпонентных групп, где перед герундием стоит словоформа, которая не может принимать причастный оборот, как *painstaking* в следующем примере:

Setting out each receiver, however, required a painstaking leveling process... —

‘Выставляющий каждый приемник, однако, необходимый тщательный процесс нивелирования...’.

Улучшить перевод простых ИГГ посредством изменения алгоритма затруднительно. Ошибки можно исправлять только в процессе редактирования, причем потребуются повторный анализ исходного предложения, поскольку в большинстве случаев по переводу невозможно определить границы проблемной группы и связи между ее элементами. Устойчивые сочетания, образу-

ющие группу или ее часть, следует вводить в АС как иконические обороты, чтобы избежать ошибок в их переводе.

На уровне анализа функциональных сегментов герундий определяется в следующих функциях: подлежащего, части сказуемого (глагольного и именного) и беспредложного дополнения. В составе указанных сегментов герундий в основном употребляется следующим образом: герундиальное подлежащее — в абсолютном начале предложения, герундий в составе сказуемого — непосредственно после вспомогательного глагола, беспредложное дополнение — непосредственно после сказуемого в активном залоге и в препозиции по отношению к существительному / именной группе с артиклем. Такие употребления хорошо переводятся системой МП. Следующий пример иллюстрирует перевод беспредложного дополнения:

*The physics of rock stability may allow **increasing** the minimum magnitude to 6.0* —

‘Физика устойчивости породы может позволять увеличение минимальной величины 6.0’.

Если модели употребления герундия в составе функциональных сегментов отличны от описанных выше моделей, то в переводе возможны ошибки. Например, герундиальное подлежащее после вводного оборота иногда переводится неверно, так как система МП может не распознать этот оборот как иконический:

*In the present case, **depending** only on the average error criterion seems to be sufficient* —

‘В присутствующем случае [корпусе], зависящем только в среднем критерий ошибки как кажется будет достаточным’.

В данном случае в АС следует ввести оборот *in the present case*.

В целом, редкие ошибки в переводе функциональных сегментов могут быть исправлены при редактировании, в некоторых случаях даже без повторного анализа исходного текста.

Сложности анализа и перевода герундия на уровне предложения в обработанном массиве вызывали только однородные члены. В большинстве случаев система МП не распознает связи между ними, независимо от того, образованы ли они герундием или другими словоформами: система МП анализирует не связь между однородными членами, а группы, образующие каждый из членов ряда, отдельно, что приводит к ошибкам в переводе простых именных групп. Так, например, статус второго герундиального подлежащего в следующем случае определен неверно:

***Picking** fault locations and **interpreting** the core of a fault-propagation fold are problematic because of large lateral velocity changes* —

‘Выбор обнаружений неисправности и при интерпретации активную зону распространения складки разлома являются проблематичными из-за больших боковых скоростных изменений’.

Редактирование перевода беспредложных однородных членов всегда требует повторного анализа исходного предложения, поскольку в переводе

часто трудно выделить структуру самого предложения, однородные члены и структуру группы каждого из них.

По результатам проведенного исследования сделан вывод, что наибольшее число ошибок машинного анализа приходится на уровень групп и уровень предложения. В переводе герундия в составе простых именных групп и однородных членов ошибки значительно преобладают над неточностями и верным переводом, следовательно, для редактирования практически всегда необходимо заново производить анализ исходного предложения. Редактору, знакомому с типичной структурой сложных для перевода групп, распределением зависимостей в группе и наиболее распространенными ошибками анализа и перевода, будет легче выделить в тексте перевода и оригинала проблематичную группу и отредактировать ее.

По результатам исследования переводов текстовой выборки можно также рекомендовать авторам текстов, предназначенных для последующего машинного перевода, по возможности не использовать простые именные группы с числом элементов больше трех и развернутые предложения с несколькими беспредложными герундиальными однородными членами.

В приложениях рассматривается перевод сложных форм герундия (Приложение 1), разрешение омонимии герундия и отглагольного существительного (Приложение 2), а также приводятся сводные таблицы примеров машинного перевода герундия на разных уровнях анализа (Приложения 3–7).

ЗАКЛЮЧЕНИЕ

Проведенное в работе сопоставительное исследование реальных предложений и их переводов и анализ реализованных в системе машинного перевода моделей анализа и перевода герундия с английского языка на русский позволили сделать следующие **выводы**.

1. Использование систем машинного перевода, а также компьютерных систем в помощь переводчику давно стало элементом профессиональной работы переводчика. Компьютерные системы в помощь переводчику предназначены для облегчения переводческой работы: они предоставляют доступ к различного рода справочной информации, систематизируют наработанный переводческий «опыт» и т. п., в то время как системы машинного перевода, несмотря на все их слабые стороны, рассчитаны на получение полностью автоматического перевода. Основная задача любой системы МП заключается в том, чтобы переводить документы более быстро и более дешево, чем человек-переводчик. При этом система МП рассматривается как практический инструмент, который должен прежде всего обеспечить удобство работы переводчика и специалиста в конкретной области знаний, не владеющего языком, на котором написан текст, или знающего его недостаточно хорошо. Таким образом, при обучении переводчиков и пользователей работе с системами МП необходимо рассматривать различные подходы к оценке и использо-

ванию результатов МП и вырабатывать у переводчиков и пользователей особые навыки работы с компьютерными текстами.

2. Герундий, неличная форма глагола английского языка, представляет собой сложную задачу для машинной обработки. Он сложен для машинного анализа ввиду своей широкой дистрибуции, а также потому, что в английском языке существует целый ряд омонимов герундия, оканчивающихся морфемой *-ing*. Кроме того, он сложен и для машинного синтеза, так как в русском языке нет аналогичной герундию глагольной формы. Ввиду указанной специфики герундия он часто переводится системой МП ошибочно.

3. Учитывая особый подход к оценке качества МП, в машинном переводе следует различать ошибки и неточности.

Неточности связаны со стилистической некорректностью переводного предложения, при этом статус самой *ing*-овой словоформы и ее связи с окружающими словоформами определяются правильно. Неточности не препятствуют пониманию текста, но требуют редактирования для создания текста в издательских целях. Редактирование может быть произведено быстро, без обращения к исходному тексту.

Ошибки возникают в случае неправильного определения грамматического статуса *ing*-овой формы или ее связей с окружающими словоформами. Они препятствуют пониманию текста, и для их редактирования зачастую необходимо производить анализ исходного предложения заново.

4. Ошибки и неточности возможны на всех уровнях машинного анализа текста: к таким уровням относятся лексико-морфологический уровень, уровни групп, функциональных сегментов и предложения. Способы установления и исправления ошибок индивидуальны для каждого уровня.

5. На лексико-морфологическом уровне и уровне функциональных сегментов система МП достаточно хорошо распознает и переводит герундий. Ошибки, допускаемые на лексико-морфологическом уровне, наиболее редки и просты в исправлении: работа с ними сводится в большинстве случаев к изменению информации в автоматическом словаре. В переводе функциональных сегментов встречаются редкие ошибки, в основном — в переводе беспредложного дополнения, которые следует исправлять в процессе редактирования, но в целом правильный перевод преобладает над ошибочным. Таким образом, редактирование перевода на этих уровнях не представляет серьезных сложностей для редактора.

6. Наибольшее количество употреблений герундия анализируется системой МП на уровне групп. На этом уровне целесообразно разделять группы с предложным управлением и простые именные группы, поскольку модели их анализа и перевода значительно различаются.

7. Группы с предложным управлением переводятся системой МП практически безошибочно, иногда встречаются неточности в переводе, которые можно исправлять на этапе постредактирования по мере необходимости.

Некоторые редкие ошибки, связанные с переводом самих предлогов, можно разрешить путем введения соответствующей информации в АС. Таким образом, перевод групп с предложным управлением требует минимальной редакторской правки.

8. Простые именные группы переводятся системой МП в большинстве случаев ошибочно, причем с увеличением длины группы растет процент ошибочных переводов. Редактирование перевода таких групп требует повторного анализа исходного предложения, поскольку в большинстве случаев по переводному предложению невозможно определить границы проблемной группы и связи между ее элементами. Проведенное исследование показало, что система МП использует ограниченное число моделей анализа развернутых простых герундиальных групп, поэтому редактору, знакомому с этими моделями, будет легче выделить в тексте перевода проблематичную группу с целью последующего ее редактирования.

В целом, по результатам исследования машинного анализа на уровне групп предложены следующие рекомендации авторам текстов, предназначенных для МП: следует избегать использования развернутых простых именных групп и, по возможности, оформлять связь между словоформами с помощью предлогов.

9. При анализе герундия на уровне предложения сложности вызывают некоторые однородные герундиальные члены: качество перевода зависит от функции однородных членов, их ближайшего окружения, словоформы, ими управляющей (если однородными не являются подлежащие или сказуемые), и т. п. Ошибки в переводе однородных членов можно исправить только в процессе редактирования, при этом предложения зачастую настолько развернуты, а употребления — индивидуальны, что трудно разработать какие-либо рекомендации редактору, поэтому приходится заново производить анализ предложения.

10. Если система машинного перевода интегрирована с системой переводческой памяти, то целесообразно заносить наиболее сложные случаи в базу данных. В целом, для профессионального переводчика максимально удобно автоматизированное рабочее место переводчика, в котором сочетались бы система МП, система переводческой памяти, электронные словари и, по мере необходимости, другие системы в помощь переводчику.

С целью дальнейшего улучшения перевода *ing*-овых форм возможно дополнительное исследование интерференции реализованных в системе моделей перевода герундия на перевод причастия, в настоящей работе такой задачи не ставилось. Такое исследование позволит наметить дополнительные изменения алгоритмов перевода *ing*-овых форм и разработать более детальные критерии редактирования.

По теме диссертации опубликованы следующие работы.

1. Бабушкина Н. В. Лингвистические и методические аспекты изучения герундия // Тенденции развития языкового и литературного образования в школе и вузе : материалы междунар. науч.-практ. конф., 22—23 апреля 1998 г. / Рос. Федерация, Рос. Акад. образования, Сев.-Зап. отд-ние, Рос. гос. пед. ун-т им. А. И. Герцена ; [редкол.: Т. Г. Рамзаева (отв. ред.) и др.]. — СПб. : Сударыня, 1998. — С. 234—236 (0,15 п. л.).

2. Бабушкина Н. В. Проблемы лексико-грамматического анализа при переводе с английского языка на русский с помощью систем машинного перевода // Проблемы прикладной лингвистики : сб. материалов Всерос. семинара : [в 2 ч.] / о-во «Знание» России, Приволжский Дом знаний, Пенз. гос. пед. ун-т ; [под ред. А. П. Тимониной]. — Пенза : [б. и.], 2000. — Ч. 1. — С. 19—21 (0,1 п. л.).

3. Бабушкина Н. В. Проблемы снятия омонимии *ing*-овых форм английского языка при переводе на русский с помощью систем МП // Вопросы теории и практики перевода : Всерос. конф., февраль 2003 : сб. материалов / Пенз. гос. пед. ун-т им. В. Г. Белинского, о-во «Знание» России, Приволжский Дом знаний ; [под ред. А. П. Тимониной]. — Пенза : [б. и.], 2003. — С. 10—12 (0,1 п. л.).

4. Бабушкина Н. В. Машинный перевод и проблема постредактирования: к вопросу о редактировании машинного перевода *ing*-овых форм английского языка // Прикладная лингвистика : новый век / Рос. гос. пед. ун-т им. А. И. Герцена, Филол. фак., Каф. приклад. лингвистики. — СПб. : ЛЕМА, 2006. — Вып. 1 : сб. науч.-метод. ст. — С. 7—12 (0,3 п. л.).

5. Бабушкина Н. В. Разрешение омонимии *ing*-форм при машинном переводе специального текста: к вопросу о соотношении система МП — редактор // Известия Российского государственного педагогического университета им. А. И. Герцена : научный журнал. Аспирантские тетради. — СПб., 2006. — № 1 (18) [октябрь]. — С. 13—17 (0,35 п. л.).