

Профессиональный ПЕРЕВОД

Инструменты и технологии

Обзор рынка средств и технологий
памяти переводов

Системы управления переводом:
новые горизонты

Автоматизация редактирования машинного
перевода: технология регулярных выражений

Наносинтаксис: новое средство общения
человека и машины

Управление

Интернационализация и разработка —
две части одного целого

Комментарии специалистов

Обобщение опыта: уроки локализации

ПРОФЕССИОНАЛЬНЫЙ ПЕРЕВОД

№ 1 (13) февраль 2008

ПРОФЕССИОНАЛЬНЫЙ ПЕРЕВОД

Журнал для специалистов в области многоязыкового перевода

Журнал распространяется по подписке и в розничную торговлю не поступает.

Подписной индекс по каталогу Агентства «РОСПЕЧАТЬ» — 36111

Интернет-подписка — www.ITbook.ru

1

февраль 2008

Содержание

ИНСТРУМЕНТЫ И ТЕХНОЛОГИИ

- 3** Обзор рынка средств и технологий памяти переводов *Йост Цецше*
- 11** Системы управления переводом: новые горизонты
Бенджамин Б. Саргент
- 13** Краткий обзор средств и стандартов автоматизации перевода
Ив Савурель
- 16** Неточные соответствия на практике и в теории *Ричард Сайкс*
- 24** Автоматизация редактирования машинного перевода:
технология регулярных выражений *Рафаэль Гусман*
- 29** Наносинтаксис: новое средство общения человека и машины
Шрули Табер

УПРАВЛЕНИЕ

- 35** Интернационализация и разработка — две части одного целого
Алан Хорват

КОММЕНТАРИИ СПЕЦИАЛИСТОВ

- 41** Обобщение опыта: уроки локализации *Горан Нордлунд*
- 45** Web 2.0 — новое слово в локализации *Ультан О'Брайен*

ИЗ-ПОД ПЕРА МАСТЕРА

- 48** Ведьма (отрывок) *Грэгори Магвайр*
- 56** Тайный дневник Розовой Гвоздики (отрывок) *Лорен Виллиг*

Главный редактор	Михаил Жомин
Зам. гл. ред. по производству	Нана Чатынян
Научный редактор	Анна Филиппюк
Дизайн и верстка	Елена Козлова
Менеджеры по подписке	Мария Колчина, itjournal@rusedit.com, Ольга Астахова, itj@rusedit.com
Размещение рекламы	Екатерина Долгова, reclama@rusedit.com

Журнал «Профессиональный перевод» —
совместный проект издательства «Русская Редакция» и компании «Логрус».



ООО «Издательство
«Русская Редакция»

Адрес: 123290, Москва,
Шелепихинская наб., д. 32
Телефон: (495) 638-5-638
Факс: (495) 256-71-45
E-mail: info@rusedit.com
Веб-сайт: www.rusedit.com



Компания «Логрус»

Адрес: 115280, Москва,
ул. Мастеркова, д. 4, этаж 6
Телефон: (495) 510-6271
Факс: (495) 510-6270
E-mail: management@logrus.ru
Веб-сайт: www.logrus.ru

Журнал «Профессиональный перевод» содержит материалы из оригинального издания **MultiLingual Computing & Technology**, переведенные на русский язык и опубликованные с разрешения MultiLingual Computing, Inc.



Журнал зарегистрирован в Федеральной службе по надзору за соблюдением законодательства в сфере массовых коммуникаций и охране культурного наследия.

Свидетельство ПИ № ФС77-25499 от 24.08.06.

Тираж 2000 экз. Формат 60x90/16. Объем 4 усл.-печ. л.

Журнал выходит ежемесячно. Распространяется по подписке.

Подписной индекс Агентства «Роспечать» — 36111.

Интернет-подписка — www.ITbook.ru, телефон (495) 256-6691, e-mail: ITJ@mail.ru

© Издательство «Русская Редакция», 2008

© Компания «Логрус», 2008

Полное или частичное воспроизведение и распространение материалов, содержащихся в настоящем издании, допускается только с письменного разрешения издателя.

Обзор рынка средств и технологий памяти переводов

Йост Цецие (Jost Zetzsche)

Пора подвести итоги: каково положение на рынке средств памяти переводов сегодня, спустя два года после приобретения в 2005 году компанией SDL (разработчиком и владельцем средства SDLX, занимающего второе место на рынке) лидера отрасли — компании TRADOS?

Вопреки пессимистическим прогнозам, рынок систем памяти переводов еще никогда не был столь оживленным.

Ко времени написания данной статьи компания SDL выпустила восьмую версию Trados, STAR Transit — пакет обновления Service Pack 20 к текущей версии, Déjà Vu предлагает сборку 302 своей текущей версии. Это только несколько примеров, свидетельствующих об активном развитии существующих средств. Кроме того, за последние полтора года полностью подготовлены два новых приложения (MemoQ и Lingotek) и еще как минимум два должны появиться до конца 2007 года. В общей сложности основные функции средств среды для перевода (возможность работы с базами памяти переводов и двуязычными текстами, управление терминологией и

проектами, контроль качества) представлены сегодня на рынке более чем двадцатью вариантами коммерческих и бесплатных (с открытым исходным кодом) средств. Отдельные функции в области преобразования файловых форматов, управления базами данных и контроля качества реализованы в ряде менее масштабных приложений.

Как видим, масштабы рынка весьма и весьма впечатляющие.

И все же, по мнению Дональда А. Де Пальма (Donald A. DePalma, компания Common Sense Advisory), это не совсем так: «В 2006 году объем рынка средств автоматизации перевода и средств локализации — включая системы управления переводом и программы машинного перевода — составил приблизительно 100 миллионов долларов США. Если же проследить ситуацию на рынке разработки приложений в 90-х годах, то мы увидим большое количество средств с ежеквартальным доходом в 10, 25 и даже 100 миллионов долларов. Причем это показатели по отдельно взятым продуктам, а не по отрасли в целом».

Обратите внимание: эти данные охватывают множество приложений, таких

как средства локализации, программы машинного перевода и средства управления переводом, которые не входят в число названных выше двух десятков средств переводческой среды. Отсюда вполне закономерный вопрос: чем вызвано это стремление к разработке новых и совершенствованию существующих решений — чисто «академическим интересом» производителей или вероятным сценарием развития рынка, включающим новые экономические возможности?

Можно предположить, что истина находится где-то посередине.

Однако прежде чем углубиться в изучение современных тенденций и перспектив в этой области, стоит уточнить терминологию. Автор данной статьи придерживается мнения о необходимости замены терминов *средства автоматизированного перевода* (computer-assistant translation (CAT) tools) или *средства памяти переводов* (translation memory (TM) tools) на более емкий и потому более точный термин *средства среды для перевода* (translation environment tool). В статье «Средства перевода: круг замкнулся», опубликованной во втором выпуске этого журнала, подробно объясняются причины, по которым автор столь упорно отстаивает эту точку зрения. Поскольку в названной статье в общих чертах описывается также история развития технологии памяти переводов, здесь эти вопросы мы обсуждать не будем.

В той статье, написанной через шесть месяцев после приобретения компанией SDL средства TPADOS, приоритетными с точки зрения «непрерывного разви-

тия» отрасли были названы следующие направления:

- управление терминологией;
- управление контентом;
- модули потока операций;
- компоненты программ машинного перевода;
- формат обмена файлами для перевода XLIFF;
- ПО с открытым исходным кодом;
- технология обмена памятью переводов.

Рассмотрим теперь изменения, имевшие место в каждой из этих областей за прошедшие два года.

Управление терминологией

Управление терминологией в переводческой отрасли можно сравнить с пресловутой ложкой дегтя. По многочисленным отзывам специалистов, в целом работе с терминологией продолжает уделяться самое незначительное внимание (подробный обзор текущего положения в области терминологической деятельности опубликован в выпуске «Управление терминологией» за декабрь 2007 г.). И это происходит несмотря на то, что большая часть средств среды для перевода поддерживает функции управления терминологией, которые сегодня все более широко внедряются и в средства контроля качества.

Как ни странно, взгляды на терминологическую деятельность стали меняться не столько после применения традиционных терминологических модулей, сколько после обращения к принципам сегментации текста — разделения на менее крупные, чем предложение (при-

вычный уровень членения переводимого текста), единицы перевода. Производителям таких средств, как Lingotek, Similis и MultiTrans наконец-то удалось привлечь внимание представителей отрасли к важности небольших фрагментов текста (в среде Similis они называются *chunks*, букв. «куски») в составе более крупных структур. Этот принцип перенимают и другие производители, уже приступившие к разработке аналогичных решений.

В то же время в средствах с функцией сбора терминологии (компании SDL со средствами PhraseFinder и MultiTerm Extract, LogiTerm, MultiTrans и Similis), становящихся все более популярными некорпоративным пользователям, реализована возможность формирования в короткие сроки терминологических баз данных.

К сожалению, появление формата TBX (формат обмена терминологическими базами данных) не оказало серьезного влияния на отрасль. За исключением средства Heartsome, в котором этот стандарт внедрен наиболее полно, активно его поддерживают только несколько приложений.

Не так давно автору этой статьи довелось пообщаться с разработчиком автономного терминологического средства, который из-за сложности интеграции формата TBX с приложением решил реализовать функцию обмена терминологическими базами данных на основе гораздо более простого формата TMX (формат обмена памятью переводов). Возможно, решить эту проблему можно будет с помощью разрабатываемого в настоящий момент формата TBX-Lite.

Управление контентом

Как и полтора года назад, управление контентом остается многообещающим направлением разработок. Расширилась практика партнерских отношений между разработчиками средств перевода и поставщиками систем управления контентом.

Особенно активными в этом направлении были компании SDL, Idiom и Across, которые либо объединили усилия с производителями таких средств, как Documentum and Interwoven, либо добились высокой степени совместимости программных интерфейсов своих приложений с системами управления контентом.

Вместе с тем языковой отрасли еще только предстоит освоить возможности, которые открывает перед ней интеграция переводческих приложений со средствами разработки контента.

Особенно заметны эти преимущества в таких средствах, как AuthorAssistant компании SDL, Authoring Coach компании Sajan и crossAuthor компании Across. Эти средства позволяют создателям технических текстов работать с базами данных памяти переводов и составлять тексты так, чтобы при переводе максимально использовать имеющиеся в них совпадения с уже накопленными материалами. Все это сулит поставщикам реальные перспективы расширения спектра своих услуг, к которым могут быть добавлены услуги разработки документации, основанные на использовании предварительно созданной с помощью заказчика памяти переводов.

Модули потока операций

Включение в средства переводческой среды модулей потока операций — довольно свежая идея, вызвавшая немалый интерес после того, как данный автор опубликовал о ней статью — стало сегодня стандартным обязательным требованием к приложению для перевода. Если говорить о поставщиках языковых услуг, то они должны сегодня интегрировать свои требования к памяти переводов и работе с терминологией в более широкую структуру управления рабочими процессами. Самым ярким примером средств, объединивших в себе все три традиционных уровня переводческого процесса — переводчиков, поставщиков и потребителей переводческих услуг, — служат такие средства, как Idiom WorldServer, Logoport (компания Lionbridge), Across и SDL Trados Synergy. С другой стороны, в таких традиционных средствах среды для перевода, как Déjà Vu, Similis, MemoQ, MultiTrans, SDLX и STAR Transit, в настоящий момент обеспечивается взаимодействие между версией для поставщиков языковых услуг и находящимся в свободном доступе изданием для переводчиков и редакторов. Кроме того, привычным стал упрощенный интерфейс между некоторыми средствами по управлению проектами и аналитическими модулями, представленный в основных средствах переводческой среды.

Новое перспективное явление в отрасли — сотрудничество, основанное на интеграции компонентов систем машинного перевода в средства переводческой среды.

Новым направлением в отрасли, в успешном развитии которого можно не сомневаться, является укрепление действующих сегодня партнерских отношений между разработчиками средств среды для перевода и поставщиками решений по управлению проектами. Например, недавно возобновилось сотрудничество между MultiCorpora и Plunet.

Компоненты средств машинного перевода

В том же ключе развивается еще одна форма сотрудничества, основанная на интеграции компонентов систем машинного перевода в средства переводческой среды — довольно новое и перспективное явление в отрасли. Компании Across и Idiom заявили о заключении партнерских соглашений с производителем программ машинного перевода Language Weaver. Есть все основания ожидать развития этой тенденции, которая приведет в итоге к более тесному сотрудничеству подобного рода и еще более полной интеграции принципов машинного перевода в средства переводческой среды. Основанные на взаимной выгоде, эти отношения будут развиваться с обеих сторон: поставщики систем машинного перевода заинтересованы во включении в свои приложения эффективных компонентов памяти переводов и управления терминологией, а производители средств переводческой среды — в обеспечении простоты и устойчивости во взаимодействии и интеграции своих продуктов с программами машинного перевода. Существует и недо-

статок такого «симбиоза». Дело в том, что большинство средств среды для перевода основаны на принципе «языкового агностицизма», то есть невозможности создания машиной полноценного текста на естественном языке. И этот принцип может быть заменен (по крайней мере, насколько это позволяют машинные способы перевода) требующей меньших усилий обработкой текстов на наиболее распространенных языках, для которых существуют программы машинного перевода.

Будут развиваться и другие, менее формализованные способы интеграции — например, простой доступ через Word к программам машинного перевода, реализованный в таких средствах, как Wordfast и MetaTaxis.

Форматы обмена файлами: XLIFF

Формат XLIFF, который задумывался как первое средство полной передачи переводимого контента между поддерживающими его приложениями, существенных успехов не добился. Несмотря на то что формат XLIFF в разной степени поддерживается целым рядом приложений (в том числе большинством, если не всеми, средствами локализации, а также такими средствами, как Lingotek, SDLX, TRADOS), вопреки ожиданиям разработчиков он не пользуется заслуженным вниманием.

Впрочем, можно отметить ряд многообещающих признаков. Следуя примеру компаний Heartsome и Sun, разработавших для своих средств Open Language Tools редактор перевода

XLIFF (XLIFF Translation Editor), компании Lingotek и MultiTrans (в редакторе XLIFF) установили этот формат в качестве формата по умолчанию для всех (Lingotek) или для HTML- и XML-файлов (MultiTrans).

В то же время фактический статус стандартного формата для обмена файлами получили два других формата, разработанные SDL Trados: двуязычный формат RTF-файлов и — в силу ограниченности последнего — формат TTX, основанный на XML. Сложно найти программные средства, которые бы не поддерживали оба или хотя бы один из этих форматов, или в которых эта функция не предполагалась бы в будущем. Хотя степень управляемости этих форматов по сравнению с XLIFF ниже, функции обмена файлами представлены в них намного полнее.

Следующий фактор, на основе которого изначально сравнивались средства среды для перевода, тоже постепенно теряет былую актуальность. Речь идет о совместимости с многочисленными форматами настольных редакторских систем и программ обработки текста. Поскольку практически все средства переводческой среды поддерживают формат XML, а исходные файлы в большинстве случаев бывают или могут быть представлены в той или иной разновидности XML, дорогостоящая разработка фильтров для новых форматов, имевшая место, как правило, только в крупных программных средствах, стремительно утрачивает актуальность. В качестве примеров можно привести последние версии InDesign, Microsoft Office и, конечно же, OpenOffice.org.

Средства с открытым исходным кодом

Возможности формата XML позволили программам с открытым кодом поддерживать форматы таких приложений, как Word, Excel и PowerPoint, без предварительного преобразования формата файлов с помощью OpenOffice. Производители OmegaT, наиболее активно разрабатываемого средства с открытым исходным кодом, объявили о том, что следующая версия будет снабжена фильтром для работы с Office 2007. Это обеспечит еще большую популярность программы среди внештатных переводчиков (в мае этого года переводческий портал ProZ.com разместил удивительное сообщение — среди участников портала OmegaT является четвертым по частоте использования средством).

Обмен памятью переводов

Технология обмена памятью переводов за последние два года добилась, пожалуй, наиболее внушительных результатов — как с точки зрения усовершенствования конкретных переводческих приложений, так и в плане общих усилий со стороны участников отрасли.

В состав многих средств включены компоненты для интерактивного обмена памятью переводов непосредственно в процессе работы над переводческим проектом (TRADOS, SDLX, Across, Fusion, MemoQ, Idiom Workbench, Logoport, MetaTaxis, MultiTrans и другие). В средстве Lingotek возможность совместного использования памяти переводов

(или «индексов», в терминах этой системы) стала одним из краеугольных камней архитектуры приложения.

Компания Wordfast начала работу над проектом VLTМ (Very Large Translation Memory, «Очень крупная база памяти переводов»), в рамках которого масштабные, для самых разных языковых пар, базы памяти переводов, не имеющие ограничений на использование, стали доступны для всех пользователей Wordfast с центрального сервера компании. Не так давно эта практика была расширена за счет выделения на сервере места для частных переводческих проектов.

Помимо широко используемого формата TMX и его менее распространенной разновидности SRX (стандарт обмена правилами сегментации, обеспечивающий применение единых правил сегментации во всех приложениях и настройках пользователя), компанией MemoQ была представлена новая интересная технология «сегментации, управляемой памятью переводов». Суть этой технологии в том, что сегментация всех возможных соответствий определяется правилами сегментации базы памяти переводов.

Методы обмена данными

Другие, не связанные рамками конкретных приложений, инициативы представителей отрасли также свидетельствуют о постоянно растущем понимании ими потребности в технологиях обмена данными. В марте 2007 года Обществом потребителей автоматизированного перевода (TAUS) была организована конференция с участием представителей 26

международных организаций. Ее цель — «изучение возможностей для повышения уровня автоматизации перевода на основе совместной платформы обмена языковыми данными, в частности путем более интенсивного использования и совершенствования систем машинного перевода» (из пресс-релиза TAUS, www.translationautomation.com/downloads/TAUSSummitNewsRelease.pdf). На конференции обсуждались существующие в этой сфере проблемы (вопросы юридического характера, необходимость классификации данных и организации инфраструктуры). Рассмотрению этих актуальных вопросов будет посвящена следующая встреча, запланированная на осень.

Еще одной такой инициативой является метод лицензирования, практикуемый компанией TM Marketplace уже в течение примерно двух лет. Вместо того чтобы просто предоставлять в распоряжение конкурентов свои информационные активы, эта концепция позволяет компании учесть ценность данных, которые содержатся в сформированной за несколько лет памяти переводов, и за определенную плату выдавать лицензии на ее использование.

Например, компания General Motors предлагает другим представителям автомобильной промышленности, поставщикам языковых услуг, работающим в этой сфере, и разработчикам систем машинного перевода приобрести через систему TM Marketplace лицензию на использование памяти переводов. В ней содержится более четырех миллионов сегментов перевода в шести языковых парах (с английского на немецкий, ис-

панский (Испания и Мексика), французский (Канада), итальянский и голландский языки).

И третий проект — «модель на заказ» — реализуется без какого-либо прямого участия первоначальных владельцев информации. Огромное количество двуязычных данных в виде PDF-файлов, веб-страниц и в других форматах доступно в сети Интернет. Хотя данные из подобных источников не предназначены для использования в качестве памяти переводов, появление мощных средств их упорядочивания и накопленный в отрасли опыт позволяют сформировать из этих документов двуязычные базы памяти переводов. Последние могут быть «сделаны на заказ» и адаптированы для нужд конкретной отрасли и конкретного продукта. Официальный документ о правовом регулировании этого вопроса размещен по адресу www.tmmarketplace.com/whitepapers/align.

Практические выводы

Подводя итоги, развитие рынка средств среды для перевода по рассмотренным категориям можно кратко охарактеризовать следующим образом.

- Постепенное изменение в области управления терминологией в связи с расширением использования при переводе более мелких, чем предложение, фрагментов текста.
- Системы управления потоком операций (либо в качестве собственной функции приложения, либо на основе внешнего сотрудничества) стали стандартным компонентом среды для перевода.

- Отмечается тенденция к интеграции технологий машинного перевода и памяти переводов, как через партнерские соглашения их производителей, так и посредством функциональной интеграции.
- Воплощена в жизнь идея обмена данными между программными средствами, особенно на уровне файлов для перевода и памяти переводов.
- Доступ к базам памяти переводов значительно упрощен, что делает возможным одновременный доступ к данным для всех участников рабочей группы переводчиков.
- Технология обмена базами памяти переводов находит поддержку в виде различных инициатив со стороны представителей отрасли, не связанных требованиями конкретных приложений.

До настоящего времени рынок развивается достаточно успешно. Благодаря активной позиции многих разработчиков их продукты быстро адаптируются к требованиям рынка или, как в случае с использованием более мелких сегментов переводимого текста, влияют на развитие самого рынка. Растущая конкуренция, несмотря на предсказанный в 2005 году спад, выравнивание позиций участников рынка за счет открытости форматов и возможности обмена данными — все это позволяет рынку проявлять чрезвычайную гибкость в условиях ускоряющихся темпов развития. Тенденции к сотрудничеству производителей и интеграции программных средств на самых разных уровнях еще больше расширяют горизонты языковых тех-

нологий, а внедрение промежуточного программного обеспечения, такого как приложение Clay Tablet (www.clay-tablet.com), способно послужить еще одним толчком к дальнейшему развитию этого процесса.

На конференции Translation Summit (www.translationsummit.org) была создана рабочая группа, в задачи которой входит содействие в сокращении разрыва между требованиями пользователей средств перевода на каждом уровне и возможностями, которые предоставляет им переводческий инструментарий. Поле деятельности этой рабочей группы не ограничено вопросами разработки средств переводческой среды, а включает в себя все автоматизированные процессы в рамках потока операций перевода. Поэтому результаты этой работы должны оказаться чрезвычайно полезными для разработчиков переводческих приложений. В начале этого исследования будет проведен обширный опрос представителей отрасли, ход и результаты которого обязательно будут освещаться на страницах специализированных журналов и в веб-форумах.

Перефразируя заключительную фразу из обращения Билла Клинтона к Конгрессу США «О положении в стране» за 2000 год, можно сказать: «Пока наши мечты перевешивают наши воспоминания, мы будем молоды. Такова наша судьба и таково наше предназначение».

Йост Цецше — один из основателей Международной группы писателей (International Writers' Group) и компании TM Marketplace, автор книги «The Translator's Tool Box: A Computer Primer for Translators».

Системы управления переводом: новые горизонты

Бенджамин Б. Саргент (*Benjamin B. Sargent*)

В феврале 2007 года компания Compton Sense Advisory опубликовала обзор существующих систем управления переводом (*Translation Management Systems Scorecards*). Со стороны потребителей — поставщиков языковых услуг и компаний из других отраслей — последовали многочисленные отзывы о том, каких возможностей недостает в этих программных средствах. Если не принимать во внимание априори неосуществимые предложения, можно составить обзорный перечень характеристик систем управления переводом, внедрение или усовершенствование которых в ближайшем будущем представляется наиболее перспективным.

1. *Подключение к рабочей среде систем управления контентом (CMS)*. Для разработчиков и специалистов по управлению контентом удобнее осуществлять и извлекать готовый перевод непосредственно в среде CMS-средства, поскольку при этом не надо устанавливать и осваивать отдельное приложение. Естественным развитием идеи «контент-интерфейса», позволяющего перемещать контент в переводческую среду и из нее, стала технология подключаемого интер-

фейса. Впервые этот подход был осуществлен в средствах компании Translations.com на основе графического пользовательского интерфейса (в виде одного или нескольких окон), совместимого с Interwoven Team Site, Documentum Web Publisher и другими приложениями по управлению контентом.

2. *Совместная работа со средствами разработки материалов*. Специалисты по управлению документацией хотели бы обеспечить своих разработчиков материалов возможностью поиска в тексте совпадающих по смыслу фрагментов, перевод которых уже имеется в базе данных памяти переводов. Решающим фактором для этой функции является простота ее использования, так как техническим писателям в этом случае придется изменить сложившуюся методику работы — сначала найти для подготовленного абзаца соответствия в базе данных памяти переводов, удалить оригинальный текст и только затем переходить к следующему абзацу. Удачному примеру компаний SDL и Sajan, в средствах которых (соответственно AuthorAssistant и Authoring Coach) эта функция реализована,

должны последовать и другие разработки.

3. *Возможности управления бизнесом.* При оформлении заказа на приобретение переводческих услуг учитывается объем текста, соответствия которому уже имеются в памяти переводов. Наибольшей популярностью пользуются приложения, позволяющие сформировать заказ на приобретение «нажатием одной кнопки». Однако мало кто из поставщиков предоставляет такую возможность. Одна из причин заключается в том, что для этого необходима система именованных ресурсов с перечислением расценок по каждому из них, а для этого, в свою очередь, требуется предварительное внедрение функций управления ресурсами. Переводческие компании практикуют формирование счетов-фактур на основе аналогичным образом генерируемых данных. Переводческие отделы компаний также выставляют свои внутренние счета, согласно которым в рамках организации этим отделам компенсируются расходы на услуги внешних поставщиков. Разработчики в ответ на все эти доводы приводят пример многочисленных расходов того же порядка: на верстку публикуемых материалов, разработку локализации, контроль качества, управление проектами и прочее. Потребители, впрочем, проявляют большое упорство в своих требованиях, подстегивая тем самым конкуренцию среди производителей.
4. *Интеграция с другими корпоративными приложениями.* Эта функция, хотя ее реализация и рассчитана на долго-

срочную перспективу, со временем станет обязательным требованием к системам управления переводом. Возможность совместной работы с приложениями по управлению бизнес-процессами на уровне предприятия, ресурсами и финансами (бухгалтерской отчетностью и закупками) должна учитываться при разработке любой полномасштабной системы управления переводом. Средство ITC Organiser уже предоставляет такую возможность в отношении ряда корпоративных приложений: SAP, Quick Books и Crystal Reports. В ближайшем будущем следует ожидать появления на рынке новых компонентов-интерфейсов для бухгалтерских приложений, а также для средств автоматизации продаж, таких как salesforce.com и SugarCRM.

На сегодня основные характеристики систем управления переводом четко определены. Это функции потока операций и централизованного управления памятью, а также возможности веб-средств перевода и редактирования. Следующим шагом разработчиков, которые обеспечивают в своих приложениях все эти характеристики, должно стать устранение разрыва в функциях управления бюджетом и ресурсами, существующего между средствами управления переводом и другими корпоративными системами регистрации. Впрочем, этот вопрос — не более чем «болезнь роста» отрасли.

Бенджамин Б. Саржент — старший аналитик компании *Common Sense Advisory*, занимающейся исследовательской и консалтинговой деятельностью.

Краткий обзор средств и стандартов автоматизации перевода

Ив Савурель (Yves Savourel)

Чтобы не запутаться в разнообразных решениях, потребители отрасли языковых услуг настояли на разработке различных стандартов, которые позволили бы им работать со средствами автоматизированного перевода и создавать материалы более долговечные, чем средства, с помощью которых они созданы. Ниже перечислены основные стандарты, используемые в переводческой области.

Стандарт XLIFF. Формат файла XML взаимодействия локализациями (XML Localization Interchange File Format) позволяет хранить извлеченный текст и метаданные в общем формате, который можно переносить с одного этапа процесса локализации на другой. В дальнейшем эти данные можно вернуть в исходный формат. Стандарт XLIFF поддерживается техническим комитетом OASIS XLIFF (www.oasis-open.org/committees/xliff).

Многие средства автоматизированного перевода работают с документами в формате XLIFF. Некоторые из них, например Heartsome XLIFF Editor, разработаны специально для этого формата. Базовые настройки такого компонента TRADOS, как TagEditor, также обладают достаточной гибкостью для работы с ним.

Стандарт SRX. Формат SRX (Segmentation Rules eXchange) базируется на кодировке Юникод TR29 и на правилах на основе регулярных выражений, которые описывают способы разбивки предложений. Недостатком версии 1.0 было отсутствие детального объяснения того, как понимать эти правила. Это привело к появлению несовместимых вариантов в среде средств автоматизированного перевода. В версии 2.0 данная проблема решается посредством алгоритма толкования правил. Этот стандарт поддерживается группой OSCAR Special Interest Group при Ассоциации по стандартам в области локализаций (LISA, www.lisa.org/standarda/tmx).

Стандарт SRX уже внедрен в такие приложения автоматизированного перевода, как комплект Heartsome, средства XML-INTL и MemoQ. Некоторые из разработчиков приложений в настоящее время работают над его внедрением (например, OmegaT).

Стандарт TMX. Формат TMX (Translation Memory eXchange) делает возможным обмен памятью переводов между разными программными средствами. Этот стандарт также поддерживается группой OSCAR Special Interest Group при Ассоциации по стандартам в об-

ласти локализаций (LISA, www.lisa.org/standards/tmx). Некоторое время назад была выпущена версия 1.4b, а сегодня большинство средств перевода поддерживают функцию импорта и экспорта TMX-документов.

Использование стандарта TMX не всегда обеспечивает единообразие обмена данными, так как внедрения этого формата могут различаться. Кроме того, часто возникают проблемы, не связанные со стандартом TMX. Они превращают передачу памяти переводов в настоящую пытку. К этому могут привести, например, различия в сегментации предложений и обработке внутритекстового кодирования разных приложений.

Стандарт TBX. Формат TBX (Term-Base eXchange) — это стандартный формат обмена терминологическими базами данных. Этот стандарт также поддерживается группой OSCAR Special Interest Group при Ассоциации по стандартам в области локализаций (LISA, www.lisa.org/standards/tmx). Стандарт TBX является примером формата обмена машиночитаемой терминологией (MARTIF, ISO-12200) и базируется на метамодели системы разметки терминологических данных (TMF, ISO-16642) и на категориях данных, обозначенных в ISO-12620.

Средства автоматизированного перевода всегда использовали простые терминологические модели, таким образом, внедрение более сложных систем, таких как TBX, отходило на второй план. Однако некоторые приложения (комплект Heartsome suite и средства XML-INTL) все же поддерживают этот стандарт.

Стандарт GMX-V. Стандарт GMX-V (Global Information Management Metrics

eXchange — Volume standard) представляет собой спецификацию для расчета и хранения различных показателей, связанных с размером документа, например для подсчета слов. Этот стандарт также поддерживается группой OSCAR Special Interest Group при Ассоциации по стандартам в области локализаций (LISA, www.lisa.org/standards/tmx).

В отрасли, где оплата работы по переводу почти всегда производится по количеству слов, огромное значение приобретает определение единого способа их подсчета. Этот вопрос вызывает оживленные споры, потому что затрагивает интересы всех участников процесса. В настоящее время стандарт GMX-V внедрен в некоторые приложения, например в средства XML-INTL и в готовящуюся к выпуску версию 7 комплекта Heartsome suite.

Стандарт ITS. Этот набор тегов интернационализации был недавно рекомендован консорциумом W3C (www.w3c.org/international/its). Стандарт имеет встроенную функцию поддержки интернационализации для любого XML-документа. Некоторые аспекты стандарта ITS предназначены для средств автоматизированного перевода. Например, в нем предлагается стандартный способ отбора материала, подлежащего переводу в XML-документе. В отличие от большинства современных средств, фильтры которых основаны на ограниченных парадигмах поиска атрибутов просмотра, стандарт ITS базируется на языке XPath — стандартном механизме доступа к элементам XML-документа, который располагает более широким набором функций.

Особенность стандарта ITS заключается в том, что он по большей части ис-

пользуется разработчиками и создателями XML-контента на самых ранних этапах локализации. Таким образом, потребители отрасли языковых услуг могут потребовать внедрения стандарта ITS в средства автоматизированного перевода. Такие поставщики памяти переводов, как XML-INTL, уже начали внедрять этот стандарт. В средствах Heartsome поддержка этого формата планируется только в следующих версиях.

Стандарт OLIF. Обычно формат OLIF используется для хранения информации о лексических единицах и терминологии. Этот стандарт поддерживается консорциумом OLIF (www.olif.net), занимающимся производством различных средств для работы с данными в OLIF-формате, включая средства для программирования и утилиты для преобразования и сравнений.

Формат OLIF в основном используется системами машинного перевода, а потому его поддерживают лишь немногие средства автоматизированного перевода: несколько приложений для работы с терминологией и SDL Trados MultiTerm, в которых также возможен ввод файлов в OLIF-формате.

Стандарт Trans-WS. Разработкой спецификации Trans-WS (Translation Web Services specification) занимается Технический комитет веб-услуг по письменному переводу при комитете OASIS (OASIS Translation Web Services Technical Committee, www.oasis-open.org/committees/trans-ws). Trans-WS – это веб-услуги, направленные на автоматизацию организационных переводческих задач. Другими словами, он используется для определения условий платежа, оценки перевода, а в случае необходимости — для возврата переведенных данных.

Несмотря на то что стандарт Trans-WS не имеет прямого отношения к средствам автоматизированного перевода, его использование свидетельствует о возрастающей популярности интернет-технологий в решении задач локализации. В дальнейшем это может повлиять на структуру программных средств. Этап разработки стандарта Trans-WS приближается к завершению, и многие поставщики услуг по локализации (Lionbridge, Thebigword) и по управлению глобализацией (Idiom) проявляют к нему живой интерес.

Дополнительные аспекты средств автоматизированного перевода тоже можно систематизировать.

Внутритекстовое кодирование. При сравнении стандартов XLIFF и TMX становится заметно, что в них используются похожие конструкции внутритекстового кодирования (изолирование кода, выделенного жирным шрифтом, ссылок на изображения и т. п.). Однако между ними существуют и различия. В долгосрочной перспективе для представления извлеченного фрагмента текста лучше выбрать один способ. Также остается нерешенным вопрос, каким образом обрабатывать и представлять последовательности переключения кодов, используемые во многих форматах. На сегодняшний день не существует определенных условных обозначений для их кодирования в извлеченных фрагментах текста.

API-интерфейс средств доступа к переводческим ресурсам. По мере увеличения спроса на веб-службы и интерактивные приложения возникает потребность в новом способе передачи связанных с переводом данных по Интернету. Просто-го импорта и экспорта файлов недоста-

точно, необходима возможность делать прямой запрос на уровне фрагментов текста или более мелких единиц. Другими словами, не хватает стандартизированных способов обращения к памяти переводов, системам машинного перевода и терминологическим базам данных. В этом случае приложения обеих сторон (запроса и предоставления данных) смогли бы работать без проблем, даже находясь в разных средах. В некоторых приложениях уже сейчас можно обратиться с интерактивными запросами к

системам машинного перевода, памяти переводов и терминологическим базам данных, но в них используются разные методы доступа. Единый интерфейс значительно упростит разработку внедрений, использующих такие системы, и откроет перед всеми участниками обмена данными новые возможности.

Ив Савурель — разработчик локализационных решений в ENLASO Corporation в г. Боулдер, Колорадо. Он также возглавляет рабочую группу по набору тегов интернационализации консорциума W3C.

Неточные соответствия на практике и в теории

Ричард Сайкс (Richard Sikes)

Перед тем как приступить к ознакомлению с возможностями автоматизированной обработки неточных соответствий (fuzzy match), необходимо уточнить понятия сегмента и единицы перевода (translation unit, TU), дав им определение в контексте баз памяти переводов (translation memory, TM). После этого мы рассмотрим процессы измерения и использования неточных соответствий и определим, к каким этапам этих процессов применима компьютерная обработка и какие здесь могут возникнуть сложности. В заключение мы обозначим сферы автоматизированного перевода, где наиболее велика вероятность технологического прорыва.

Память переводов, сегментация, единицы перевода и неточные соответствия

Память переводов — это база данных, в которой собраны соответствующие фрагменты двух наборов документов, один из которых является переводом другого. Один из этих наборов документов представлен на языке оригинала, а другой — на языке перевода. Рассмотрим в качестве примера набор документов, состоящий из одного исходного документа и одного конечного документа. Набор также может состоять из нескольких исходных документов и нескольких соответствующих конечных документов. Так бывает, например, когда несколько документов

Документ	Исходный	Конечный
Электронная версия руководства пользователя	Английская версия (оригинал)	Французская версия (перевод)

посвящены одной теме, и важно обеспечить единообразие перевода.

Фрагменты каждого набора документов, занесенного в память переводов, называются *сегментами*; они могут быть разного размера (по характеру разбиения). Размер определяется правилами сегментации, которые программа использует при создании базы памяти переводов. В большинстве случаев наиболее удобной является сегментация на уровне предложения, особенно если речь идет о переводе с английского языка. Это связано с тем, что в английском предложении смысловая единица передается подлежащим, сказуемым, прямым дополнением и несколькими определителями. Предложение грамматически независимо от окружающего контекста. Тем не менее в некоторых случаях, которые мы рассмотрим далее, может потребоваться более крупное членение. Наиболее простые базы памяти переводов включают пары сопоставленных сегментов на исходном и конечном языках. Эти пары называются единицами перевода.

Сама по себе база памяти переводов особой пользы не приносит, но она важна при переводе новых документов. Если содержимое нового документа сходно с материалами, хранящимися в базе памяти переводов, соответствующие сегменты можно будет в той или иной степени использовать при переводе нового документа. Если фрагменты нового исходного документа в значительной степени совпадают с сегментами базы памяти переводов,

степень повторного использования будет высокой. Так бывает, когда переводятся материалы с большим количеством повторов — например, очередное переиздание пособия для ВУЗов или инструкции к медицинскому оборудованию.

Возможность повторного использования не ограничивается сегментами нового исходного документа, для которых имеются точные совпадения в базе памяти переводов. Если сегмент нового исходного документа лишь частично совпадает с сегментом памяти переводов, мы получаем так называемое «неточное соответствие» (*fuzzy match*).

Например, в старом исходном документе было предложение «The grass is green today». В базе содержится его перевод («Трава сегодня зеленая»). Если в новом исходном документе есть предложение «The tree is green today» («Дерево сегодня зеленое»), с ним частично совпадет уже существующая единица перевода. Степень совпадения будет зависеть от нескольких факторов, включая грамматический род слов *tree* и *grass*, морфологию прочих слов единицы перевода и другие языковые параметры.

Насколько точны неточные соответствия?

Встает вопрос: как же измерить степень соответствия? Существует несколько способов, и каждый имеет свои плюсы и минусы.

Давайте обратимся к нашему примеру:

The grass is green today.

The tree is green today.

Каждое из предложений содержит пять слов. Одно слово изменяется. Поэтому в простейшем случае можно считать, что соответствие составляет 80% (по крайней мере для исходных сегментов). Однако не факт, что и перевод при этом будет совпадать на 80%, — по причине уже упомянутых грамматических особенностей.

Рассмотрим перевод на немецкий язык:

Das Grass ist heute grün.

Der Baum ist heute grün.

В немецком варианте, к примеру, это приводит к изменению двух слов. Как в этом случае определять степень соответствия: на основе оригинала (80%) или на основе перевода (60%)?

Давайте теперь рассмотрим французский перевод:

L'herbe est verte aujourd'hui.

L'arbre est vert aujourd'hui.

Начнем с того, что в предложении теперь четыре слова. Поменялось подлежащее, но из-за изменения рода поменялся и определитель *verte*.

Считать ли это 50-процентным соответствием? Такая оценка, учитывая простоту предложения и относительно небольшие контекстные изменения, кажется заниженной. Может ли простое исчезновение буквы *e* в окончании рассматриваться как целых 25% разницы в степени соответствия? С точки зрения человека — скорее всего нет.

Эти простые примеры показывают, насколько сложно бывает определить степень соответствия и, следовательно-

но, параметры поиска неточных соответствий. Возможно, другой подход к разбиению для измерения степени расхождения таких слов, как *verte* и *vert*, был бы более обоснованным. Алгоритмы, достаточно распространенные при проверке орфографии, выявлении плагиата и т. п., выполняют сопоставление последовательностей символов на основе принципа «расстояния Левенштейна», известного также как «расстояние редактирования».

Расстояние Левенштейна. Национальный институт стандартов и технологии (NIST) дает следующее определение расстояния Левенштейна: «наименьшее число вставок, удалений и замен, необходимых для превращения одной последовательности символов в другую (или одного дерева в другое)». В данной статье мы будем рассматривать только последовательности символов, которые NIST определяет как «список символов ..., в содержательном смысле представляющий собой слово, фразу, предложение и т. п.».

В случае замены *verte* на *vert* расстояние Левенштейна определить просто: необходимо одно удаление, следовательно, расстояние равно 1. Аналогичным образом при превращении слова «забор» в слово «засор» результат будет тоже равен 1, так как опять происходит замена лишь одной буквы («б» на «с»).

Трансформацию слов, имеющих более значительные отличия, — например, *multilingual* в *multilateral*, — можно наглядно представить в виде матрицы, которая предоставляет один или несколько путей с минимальными затратами при подсчете совокупной дистанции для каждой комбинации букв, одновременно

Табл. 1. Матрица, иллюстрирующая простой пример расстояния Левенштейна

		-1	0	1	2	3	4	5	6	7	8	9	10	11
				m u l t i l a t e r a l										
-1		0	1	2	3	4	5	6	7	8	9	10	11	12
0	m	1	0	1	2	3	4	5	6	7	8	9	10	11
1	u	2	1	0	1	2	3	4	5	6	7	8	9	10
2	l	3	2	1	0	1	2	3	4	5	6	7	8	9
3	t	4	3	2	1	0	1	2	3	4	5	6	7	8
4	i	5	4	3	2	1	0	1	2	3	4	5	6	7
5	l	6	5	4	3	2	1	0	1	2	3	4	5	6
6	i	7	6	5	4	3	2	1	0	1	2	3	4	5
7	n	8	7	6	5	4	3	2	2	2	3	4	5	6
8	g	9	8	7	6	5	4	3	3	3	3	4	5	6
9	u	10	9	8	7	6	5	4	4	4	4	4	5	6
10	a	11	10	9	8	7	6	5	4	5	5	5	5	5
11	l	12	11	10	9	8	7	6	5	5	6	6	5	4

продвигаясь от начала к концу каждого из двух слов. Здесь четко просматривается путь, в котором замены начинаются только после прохождения отрезка *multil*. Следующих трех букв (*ing*) в слове *multilateral* нет, поэтому число замен каждый раз увеличивается на единицу. Последние две буквы (*al*) совпадают, и в итоге мы получаем число 4 (см. табл. 1).

В данном случае использовался простой путь вычисления на основании различий в написании. Этот метод работает не всегда. В качестве следующего примера рассмотрим трансформацию слова *acrobat* в слово *memory*.

Первые три буквы различаются, зато и в том, и в другом слове в четвертой позиции стоит буква *o*, что приводит к разветвлению пути. Таким образом, итоговое значение 8 может быть получено разными способами (см. табл. 2). Наличие нескольких путей не связано напрямую с компьютерной обработкой.

В целом, с человеческой точки зрения, более «затратный» путь может оказаться более наглядным и интуитивно понятным. Впрочем, в данном случае нас интересует лишь получение максимально короткого пути, который каждый раз оказывается равен 8 шагам.

Пока все вроде бы понятно. Однако в некоторых случаях простой расчет расстояния редактирования не будет являться точным отражением того, как человеческий мозг воспринимает изменения в словах. Рассмотрим еще один пример: простое изменение порядка слов во фразе *multilingual computing*. Кроме всего остального этот случай интересен тем, что многие буквы повторяются. Это можно наглядно представить, расположив две последовательности символов друг под другом — таким образом, чтобы соответствующие буквы находились на одном уровне. Подсчитав общее количество вставок, замен и удалений, получаем результат 18 (см. табл. 3).

Повторение 13 из 22 элементов (считая пробел) приводит к значительному

Табл. 2. Матрица, иллюстрирующая второй пример расстояния Левенштейна

		-1	0	1	2	3	4	5
				m e m o r y				
-1		0	1	2	3	4	5	6
0	a	1	1	2	3	4	5	6
1	c	2	2	2	3	4	5	6
2	r	3	3	3	3	4	4	5
3	o	4	4	4	4	3	4	5
4	b	5	5	5	5	4	4	5
5	a	6	6	6	6	5	5	5
6	t	7	7	7	6	6	6	6
7	i	8	8	8	8	7	7	7
8	c	9	9	9	9	8	8	8

Табл. 3. Большая матрица, иллюстрирующая трансформацию фразы multilingual computing в computing multilingual

		m	u	l	t	i	l	i	n	g	u	a	l	c	o	m	p	u	t	i	n	g						
	c	o	m	p	u	-	t	-	i	-	-	-	-	-	-	m	-	u	-	t	i	l	i	n	g	u	a	l
		-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21				
				c	o	m	p	u	t	i	n	g		m	u	l	t	i	l	i	n	g	u	a	l			
-1		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22				
0	m	1	1	2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21				
1	u	2	2	3	3	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20					
2	l	3	3	3	3	4	4	4	5	6	7	8	9	10	10	11	12	13	14	15	16	17	18	19				
3	t	4	4	4	4	4	5	4	5	6	7	8	9	10	11	10	11	12	13	14	15	16	17	18				
4	i	5	5	5	5	5	5	5	4	5	6	7	8	9	10	11	10	11	12	13	14	15	16	17				
5	l	6	6	6	6	6	6	6	6	5	6	7	8	9	9	10	11	10	11	12	13	14	15	16				
6	i	7	7	7	7	7	7	7	7	6	6	7	8	9	10	10	10	11	10	11	12	13	14	15				
7	n	8	8	8	8	8	8	8	7	6	7	7	8	9	10	11	11	11	11	10	11	12	13	14				
8	g	9	9	9	9	9	9	8	7	6	7	8	9	10	11	12	12	12	11	10	11	12	13					
9	u	10	10	10	10	10	9	10	9	8	7	8	8	9	10	11	12	13	12	11	10	11	12					
10	a	11	11	11	11	11	10	10	10	9	8	8	8	9	9	10	11	12	13	13	12	11	10	11				
11	l	12	12	12	12	12	11	11	11	10	9	9	9	9	10	11	11	12	13	13	12	11	10					
12		13	13	13	13	13	12	12	12	11	10	9	10	10	10	10	11	12	12	13	14	13	12	11				
13	c	14	13	14	14	14	13	13	13	12	11	10	10	11	11	11	11	12	13	13	14	14	13	12				
14	o	15	14	13	14	15	14	14	14	13	12	11	11	11	12	12	12	12	13	14	14	15	14	13				
15	m	16	15	14	13	14	15	15	15	14	13	12	11	12	12	13	13	13	13	14	15	15	15	14				
16	p	17	16	15	14	13	14	15	16	15	14	13	12	12	13	13	14	14	14	14	15	16	16	15				
17	u	18	17	16	15	14	13	14	15	16	15	14	13	12	13	14	14	15	15	15	15	16	16	16				
18	t	19	18	17	16	15	14	13	14	15	16	15	14	13	13	14	15	16	16	16	16	16	16	17				
19	i	20	19	18	17	16	15	14	13	14	15	16	15	14	14	14	15	16	17	17	17	17	17	17				
20	n	21	20	19	18	17	16	15	14	13	14	15	16	15	15	15	14	14	15	16	17	18	18	18				
21	g	22	21	20	19	18	17	16	15	14	13	14	15	16	16	16	15	15	15	16	17	18	18	18				

разветвлению и большому разнообразию путей прохождения через матрицу.

С человеческой точки зрения может показаться странным, что при простой перестановке слов дистанция оказывается значительно больше, чем при трансформации слова *acrobat* в слово *memory*. В целом, несмотря на точность вычислений, само по себе расстояние Левенштейна не может считаться самым удобным практическим решением для выявления пригодных к исполь-

зованию соответствий (в человеческом понимании).

Расстояние Левенштейна может определяться как для символов, так и для слов. В связи с этим для восточных языков удобнее ориентироваться на символы, а для языков, в которых границы слова обозначаются пробелами — на слова.

Давайте теперь рассмотрим следующий аспект. В идеале расчет неточного соответствия с помощью многоуровневого сопоставления (с учетом сегментов, слов и

расстояния Левенштейна или возможных вариаций) должен давать показательный результат. Намного точнее результаты были бы, если бы приоритеты при выборе характера разбиения (например, по словам или по частям слов) устанавливались с учетом особенностей конкретного языка. Такие системы пока не разработаны.

На практике разработчики средств памяти переводов не раскрывают, как именно производится расчет соответствий. Так, несмотря на распространенность средства TRADOS, алгоритм, используемый в этом приложении, остается производственным секретом. Аналогичным образом обстоит дело и с другими системами. Тем не менее переводческое сообщество ориентируется на неточные соответствия при расчете общих количественных показателей, рассматривая их в качестве источника сравнительных данных, основы для составления ценовых предложений по проектам и т. д. В связи с этим отсутствие универсальных общественных стандартов выглядит неуместным и вредным для отрасли в целом.

Неточные соответствия на практике

Итак, мы рассмотрели некоторые теоретические аспекты использования неточных соответствий. Давайте теперь посмотрим, как неточные соответствия используются переводчиками и какие элементы этой работы, традиционно выполняемой человеком, можно было бы автоматизировать.

Если разделить процесс преобразования неточного соответствия в контексте

обычного перевода на отдельные этапы, можно сделать вывод, что при работе с неточным соответствием переводчик последовательно выполняет три операции: идентификацию, поиск и синтез. Давайте рассмотрим каждую из этих операций более подробно.

Идентификация. Человеческий мозг способен определять структурные различия, а также замечать пропущенные или измененные слова и фразы в сходных языковых сегментах. При работе с небольшими объемами данных мы легко справляемся с этой задачей, однако по мере увеличения этого объема эффективность падает. Здесь на помощь приходит компьютер, обеспечивающий недостижимо высокую скорость синтаксического анализа огромных объемов данных при подборе вероятных соответствий для сегментов. Возможность просматривать только результаты такого анализа существенно сокращает объем материалов, с которыми приходится работать человеку.

В исходном и конечном фрагментах единицы перевода каждого найденного варианта соответствия для переводчика выделяются отличающиеся от нужного фрагмента и совпадающие элементы. На данном этапе компьютер предлагает переводчику более крупные сегменты для преобразования в соответствующие лингвистические эквиваленты, отделяя их от фрагментированных пар сегментов, для которых наиболее эффективным будет перевод без использования неточных соответствий.

Поиск. Чтобы внести нужные изменения в предложенный вариант и привести его в соответствие с оригиналом,

переводчику нужно будет выполнить поиск в хранилищах. Эти хранилища могут содержать языковые знания, полученные из словарей, справочников и терминологических баз данных, в результате проведения исследований или накопленного опыта, а также содержащиеся в уже переведенных частях этого документа. Такие задачи довольно просты для человека, и в этом случае обработка больших массивов материалов вызывает затруднения. Компьютер выполнит эту работу значительно быстрее и точнее. Возможности компьютерных систем в области быстрого и эффективного поиска потенциальных соответствий по огромному количеству элементов значительно превышают человеческие.

И все же при выполнении подобных задач существует область, в которой человек имеет преимущество перед компьютером — это способность к логическим выводам. Переводчик может предположить: «В этом разделе речь идет о судостроении. Значит, скорее всего, необходимо просмотреть только возможные элементы из других областей базы памяти переводов, связанных с судостроением. Кажется, я уже встречал что-то подобное в этом документе или в материалах для справки». Таким образом, к поиску добавляется контекст. Это предоставляет возможности для разумного сужения области поиска, чего невозможно достичь простым структурным сопоставлением, и позволяет переводчику переходить к другим единицам перевода, находящимся в той же базе памяти переводов или даже вне ее.

Традиционные базы памяти переводов не поддерживают автоматического

заимствования элементов из других единиц перевода, даже если они находятся внутри одной базы данных. Иначе говоря, переводчик может осуществлять поиск других соответствий, но при этом не предоставляются полные сведения о контексте, и результаты не подставляются автоматически в перевод, даже если они существуют и их можно найти. Концепция, охватывающая эти возможности, условно называется «расширенным сопоставлением» (*advanced leveraging*) и активно обсуждается в кругах, связанных с вопросами автоматизацией перевода. Появляются новые системы, поддерживающие расширенное сопоставление. Эти системы имеют возможность поиска за пределами отдельной единицы перевода, что расширяет диапазон потенциальных элементов для сопоставления, а также позволяет использовать другие единицы перевода и даже источники, расположенные вне основной базы памяти переводов. Это важный этап в развитии технологий памяти переводов, хотя значительно больший потенциал развития связан с другой сферой — с полностью автоматизированным синтезом неточных соответствий.

Синтез. Способность человека к обобщению разрозненных данных значительно выше, чем у компьютера. Он способен выбрать из ряда источников на языке перевода наиболее подходящее соответствие для непереведенной части фрагмента и привести его в соответствие с контекстом и лингвистическими требованиями конкретного языка. И справляется с этой задачей легко и весьма эффективно. Некоторые современные программные решения повы-

шают эффективность автоматического поиска в Интернете и позволяют сузить область выбора потенциальных соответствий, однако окончательный выбор и адаптация все еще выполняются человеком. Эти решения не могут считаться оптимальными, так как их применение помогает только в достаточно простых случаях, а при усложнении задачи способствует увеличению объема работ.

Эмулировать с помощью компьютера процессы синтеза, выполняемые человеческим мозгом, довольно сложно. Эта задача относится к классу сложных комбинаторных задач, называемых «NP-полными». Говоря простым языком, NP-полная задача — это задача, решение которой крайне сложно рассчитать, но очень просто проверить. Классический пример — задача о коммивояжере, который должен объехать множество городов и снова вернуться в точку отправления, преодолев при этом минимально возможное расстояние. Проверить, действительно ли расстояние получилось минимальным, просто: достаточно сравнить его с другими расстояниями, но сам поиск оптимального маршрута достаточно сложен. Так же дело обстоит и с поиском подходящих соответствий. Проверить правильность выбора, применяя логические и языковые правила, достаточно просто, однако для компьютера ограничение области употребления и определение того, какое именно соответствие использовать, весьма сложная задача.

После того как соответствие выбрано, нужно применить грамматические правила и убедиться, что все языковые нормы соблюдены. Переводчик делает

это инстинктивно, исходя из приобретенного опыта. Компьютеры достаточно успешно выполняют эту задачу уже около 20 лет: это функция проверки грамматики в программных средствах обработки текстов на базе ПК. О мастерстве в нашем случае имеет смысл говорить в отношении предыдущего этапа. Оно проявляется в выборе и адаптации составных частей фрагментов.

Расширенное сопоставление — это перспективно

В основе описанной концепции лежит способность человека с легкостью идентифицировать, находить и синтезировать составные части единицы перевода (такие как отдельные слова и фразы) на своем родном языке. Человеческий мозг легко справляется с этой многократно повторяемой задачей. Вместе с тем логика и опыт показывают, что с решением повторяющихся задач компьютер обычно справляется быстрее человека, и это помогает высвободить ресурсы для других видов деятельности. Итак, возможно ли столь же успешное решение описанного комплекса задач при помощи компьютеров? Технология, дающая положительный ответ на этот вопрос, — это поистине Святой Грааль в области автоматизации перевода.

Многие компании работают над проблемой расширенного сопоставления, пытаются улучшить качество неточных соответствий, а иногда даже исключить их вообще. В ближайшем будущем следует ожидать интенсивной и массовой интеграции технологий памяти переводов с технологиями основанного на

правилах и статистического машинного перевода, других технологий обработки естественных языков, а также со все более мощными функциями поиска, обеспечивающими расширенное сопоставление. Компания, которая сможет реализовать подобное сочетание технологий в удобной для использования

форме, надолго станет лидером в нашем двадцать первом веке.

Ричард Сайкс — внештатный инструктор и консультант в области управления локализацией. Работает в сфере локализации с 1989 года.

Автоматизация редактирования машинного перевода: технология регулярных выражений

Рафаэль Гусман (Rafael Guzmán)

Несколько лет назад Ив Шампольон (Yves Champollion) — переводчик, создатель программы Wordfast — писал: «Превратится ли переводчик в будущем... в редактора компьютерных переводов? К сожалению, да». Затем он поясняет: «Вполне возможно, что способ получить достойный машинный перевод появится раньше, чем мы думаем».

Несмотря на интенсивное развитие систем машинного перевода, наиболее серьезной проблемой для этой технологии по-прежнему остается уровень лингвистического качества, — с сожа-

лением констатируют Ариадна Фон-Литьос (Ariadna Font-Llitjós) и Джейми Карбонелл (Jaime G. Carbonell), сотрудники университета Карнеги-Меллон. В конце концов, результаты машинного перевода необходимо каждый раз редактировать вручную, исправляя грубые языковые ошибки. Как бы мы ни относились к этому факту, редактирование еще долго будет оставаться обязательным этапом работы над машинным переводом.

В этом смысле эффективность его была бы выше, если бы удалось максимально автоматизировать процесс редактирования. Это означает, что поль-

зователь должен иметь возможность не только вносить изменения в словарь и языковые правила, используемые программой, но и осуществлять в тексте машинного перевода автоматический поиск и замену определенных языковых моделей.

Чтобы как можно быстрее достичь этой цели, не столько разработчики, сколько переводчики и лингвисты должны открыть для себя и освоить потенциал технологии регулярных выражений. Использование этой технологии дает возможность выявлять и исправлять самые сложные языковые ошибки, часто встречающиеся в машинных переводах.

Стоит отметить, что данная функция поиска и замены фрагментов текста на основе регулярных выражений по крайней мере частично реализована в ряде средств по локализации программного обеспечения (например, CATALYST и PASSOLO) и редактированию памяти переводов (Olifant), а также во многих текстовых редакторах, таких как EditPad Pro, Notepad++ и UltraEdit. Между тем в широко используемых системах машинного перевода эта функция до сих пор отсутствует.

Регулярные выражения и языковые модели

Под *языковыми моделями* (англ. *linguistic pattern*, также можно встретить термины *шаблон*, *маска*, *образец*) в данной статье понимаются языковые ошибки, регулярно повторяющиеся в тексте машинного перевода. Например, при переводе на испанский язык названий

торговых марок, таких как Windows и Linux, программа каждый раз добавляет определенный артикль *el* (*el Windows*, *el Linux*). Это и будет языковой моделью, содержащей лишний артикль. В ходе редактирования эту ошибку необходимо будет исправить столько раз, сколько она встречается в каждом переведенном машиной тексте.

Что касается регулярных выражений (программисты часто сокращают этот термин до *regex/regexp*, от англ. *regular expression*), то их используют для поиска, извлечения или управления текстовой информацией на основе типичных моделей, в данном случае — языковых. В зависимости от модели синтаксическая структура регулярного выражения может быть как простой, так и достаточно сложной. Правилам построения регулярных выражений посвящено множество книг и электронных руководств, поэтому в данной статье этот вопрос не рассматривается. Вместо этого мы сосредоточим внимание на возможностях применения этой методики для автоматизации процесса редактирования машинного перевода.

Проиллюстрируем на приведенном выше примере принцип работы этой технологии. Исходя из описанной языковой модели, для поиска в результатах машинного перевода всех случаев подстановки артикля *el* перед словами *Windows* и *Linux* может быть создано следующее регулярное выражение:

```
Search=\bel (Windows|Linux)\b,
```

где `\b` обозначает границу между словами, а группа `(Windows | Linux)` — названия двух возможных торговых марок; группу можно расширить, отделяя каж-

дый элемент знаком |. Затем с помощью другого регулярного выражения эта модель заменяется на текст, не содержащий ошибок:

```
Replace=$1,
```

где \$1 обозначает все, что находится в скобках в первой группе (в данном случае Windows и Linux). Поскольку артикль *el* не включен в поле `Replace`, при каждом появлении *el Windows* или *el Linux* этот текст будет заменен, соответственно, на *Windows* или *Linux*.

Главная особенность этой методики в том, что регулярное выражение может быть построено для поиска и замены любой предварительно определенной языковой модели.

Практические рекомендации

Безусловно, каждую отдельно взятую языковую проблему решить невозможно. Вместо этого необходимо сосредоточить усилия на языковых моделях, но недостаток времени часто делает невыполнимой и эту задачу. Таким образом, решающим фактором становится частота той или иной модели, однако при этом следует учитывать также степень ее релевантности: на редактирование вручную хотя и редко встречающихся, но неясных по смыслу или синтаксически сложных конструкций требуется зачастую гораздо больше времени, чем на правку часто встречающихся моделей. Таким образом, подход к поиску, группировке и замене языковых моделей должен быть системным.

Хотя существует возможность поиска языковых моделей, которые можно будет использовать в будущем, одновременно

с редактированием текста вручную, целесообразнее выполнять эти операции по отдельности. В этом случае необходимо определенное время для тестирования программы, чтобы знать особенности ее работы при переводе различных видов контента. Разовый анализ — например, непосредственно перед запуском конкретного проекта — можно сравнить с авральным режимом работы пожарного. Для среднесрочной перспективы такой подход малопродуктивен, гораздо полезней для будущей работы будет анализ и обработка результатов по итогам каждого проекта.

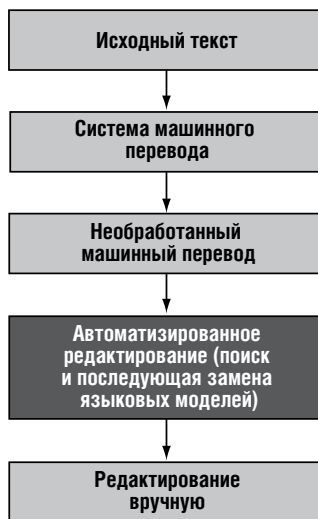
По мере выявления языковых моделей среди них выделяются различные категории, такие как «порядок слов», «грамматическое согласование», «орфография», «отсутствие слов в словаре программы» и так далее. На основе этой группировки может быть сформирована система показателей, которая позволит более точно определить основные проблемы. Для этого можно прибегнуть к представлению данных в табличном формате. Одно из преимуществ распределения языковых моделей по категориям заключается в том, что при недостатке времени их можно будет исправить на более позднем этапе работы.

Необходимо уточнить, что регулярные выражения не заменяют словари, включенные в системы машинного перевода. Напротив, они служат в качестве дополнения к последним. Другими словами, если конкретная языковая проблема может быть решена путем обычного обновления словаря приложения, использовать регулярные выражения нет необходимости.

При создании регулярных выражений следует учитывать, что они должны быть как можно более гибкими: одно и то же выражение должно быть применимо для максимального числа вариантов его употребления в тексте машинного перевода. Понятно, что чем точнее определена языковая модель, тем проще внедрение соответствующего регулярного выражения. Создание приемлемого регулярного выражения может потребовать от лингвиста творческого подхода. В этом отношении весьма полезным может оказаться выявление в тексте «узловых элементов» (в терминологии программистов — *anchor points*), таких как запятые, точки и обрамляющие теги. В любом случае внедрение регулярного выражения является завершающим этапом аналитического процесса.

Приступив к разработке регулярных выражений, необходимо внедрить процесс автоматического поиска и замены для каждой отдельно взятой языковой модели. Причем этапы этого процесса должны выполняться в строгой последовательности (см. **рис. 1**). Этот последовательный поиск можно сравнить с реализованной в приложении CATALYST функцией, когда с помощью технологии *ezParse* на основе созданного пользователем списка регулярных выражений в файле осуществляется поиск и извлечение соответствующих фрагментов текста. Однако, как сказано выше, такого рода функция в широко распространенных системах машинного перевода на сегодняшний день отсутствует. В этой ситуации компаниям необходим творческий подход при разработке собственных решений. В крайнем случае редакторы

Рис. 1. Автоматизированное редактирование машинного перевода существенно снижает объемы ручной правки текста и уровень связанных с ней затрат



всегда могут использовать функции поиска и замены текста, предусмотренные в текстовых редакторах, с возможностью использования регулярных выражений.

Наконец, рекомендуется всякий раз убеждаться в отсутствии ошибок как в построении регулярных выражений, так и в последовательности их использования. В противном случае возможны нежелательные последствия языкового характера. Например, на основе предположения о том, что в испанском языке признаком неопределенной формы глагола являются окончания *-ar*, *-er* и *-ir*, можно создать регулярное выражение, удаляющее артикли перед инфинитивами. В результате артикли будут ошибочно удалены и при некоторых существительных, например *el lugar*, *el primer* и *el router*.

Заключение

Редактирование машинного перевода можно в значительной степени автоматизировать за счет применения системы регулярных выражений для поиска и замены определенных языковых моделей в переведенном машиной тексте. Этот подход существенно снижает объемы ручной правки текста и уровень связанных с этим издержек. С другой стороны, чтобы переводчики и лингвисты освоили технику работы с регулярными выражениями и анализ результатов машинного перевода, требуется определенное время. Идеальным вариантом было бы включение соответствующего курса в учебные планы по подготовке переводчиков. Поскольку в системах машинного перевода не обеспечивается адекватной среды для редактирования, компании должны разрабатывать собственные решения, позволяющие осуществлять в тексте перевода последовательный поиск и замену языковых моделей на основе регулярных выражений.

В заключение следует отметить, что способ автоматизированного редактирования, описанный в данной статье, был протестирован с использованием системы машинного перевода на базе правил. По сравнению с системами, основанными на принципе статистических данных, в этом случае вероятность появления в тексте перевода регулярных языковых моделей выше. Поэтому можно предположить, что статистический принцип ра-

боты программ машинного перевода с этой точки зрения будет менее результативным. Однако этот вопрос нуждается в дальнейшем изучении.

Библиография

1. Butt, J. & Benjamin, C., 2004. *A New Reference Grammar of Modern Spanish*. Hodder Arnold, London.
2. Champollion, Y., 2001. Machine translation (MT), and the future of the translation industry, *Translation Journal*, Vol. 5, No. 1. Retrieved: May 10, 2007, from: <http://accurapid.com/journal/15mt.htm>.
3. Font-Llitjós, A. & Carbonell, J. G., 2006. Automating Post-Editing to Improve MT Systems. Paper presented at the AMTA Automated Post-Editing Techniques and Applications Workshop, Cambridge, 2006.
4. Фридл, Дж. (Friedl, J.). Регулярные выражения (*Mastering Regular Expressions*). — СПб.: «Питер», 2003. (Серия «Библиотека программиста»).
5. www.regular-expressions.info.
6. http://ru.wikipedia.org/wiki/Регулярные_выражения.

Рафаэль Гусман — сотрудник компании *Syntaxes Corporation* (Ирландия); работает в области систем машинного перевода для испанского языка, управления терминологией и вопросов лингвистического качества. Автор выражает благодарность доктору Джоанн Ротпурье (Johann Roturier), открывшей перед ним возможности использования регулярных выражений для целей, описанных в данной статье.

Наносинтаксис: новое средство общения человека и машины

Шрули Табер (*Sruly Taber*)

Можно ли избавить человека от необходимости изучать сложные языки программирования, заставив компьютер распознавать человеческую речь?

В 1968 году на экраны вышел фильм «2001 год: космическая одиссея», одним из основных персонажей которого был придуманный авторами робот HAL 9000. С тех пор воплощение в жизнь идеи о том, что компьютер может понимать человеческую речь, стало заветной целью ученых. Однако до настоящего момента ни одна попытка не увенчалась успехом — разве что на страницах научно-фантастических романов. Возможно, все дело в том, что попытки совместить языки человека и машины основывались на развитии компьютерных технологий, а широкие возможности, открытые когнитологией (наукой о мышлении) упускались из виду. Причиной для этого послужил недостаток знаний в области когнитивного раздела лингвистики.

Проблема

Почему люди не могут общаться с компьютерами на своем обычном языке? Как установили ученые, в человеческом мозге есть область, отвечающая за выполнение

связанных с речевой деятельностью задач, включая порождение и понимание речи. Способность понимать естественный язык буквально «встроена» в ДНК человека.

Из-за отсутствия этого компонента в компьютерной программе она и не способна распознавать естественный язык. Поэтому вопрос можно сформулировать по-другому: почему настолько сложно запрограммировать компьютер на понимание естественного языка? Дело в том, что с помощью этой «языковой способности» мозг кодирует все создаваемые им сообщения в цепочку языковых знаков. Второй человек, участвующий в общении, также обладает способностью к пониманию языка, с помощью которой он естественным образом расшифровывает полученное сообщение. А поскольку у компьютера такого «дешифратора» нет, его сложно «научить понимать» мысль, которую человек-оператор таким образом закодировал.

Каждое предложение обладает поверхностной и глубинной структурой. Глубинная структура одинакова для всех предложений с одинаковым значением, даже если они составлены на разных языках, в то время как поверхностная структура уникальна. Фразы «Джон про-

читал книгу» и «Книга была прочитана Джоном» имеют разную поверхностную, но общую глубинную структуру. Сам факт того, что человеческий мозг способен воспроизвести два одинаковых по смыслу, но разных по звучанию предложения подтверждает описанную идею кодирования: мозг извлекает хранящуюся в памяти информацию, а затем создает конкретные высказывания для ее передачи. Одна и та же мысль может быть сформулирована различными способами, что и препятствует пониманию человека компьютером. Если бы наша речь не была закодированной системой, существовал бы только один способ передачи той или иной информации, и в этом случае компьютер мог бы без труда понимать человека.

Решение: наносинтаксис

Синтаксис представляет собой раздел лингвистики о правилах и моделях объединения слов в словосочетания, фразы и предложения. Сам термин происходит от греческих слов *syn* (вместе) и *taxis* (последовательность). Для выявления смысловых зависимостей между словами в предложении в языкознании используется модель синтаксического дерева. В традиционной модели (также используется термин «дерево синтаксического разбора/анализа») каждому слову в составе предложения отводится своя узловая точка. В более подробной модели может быть представлен уровень элементов слова — морфем, таких как приставка, корень, суффикс. На морфологическом дереве каждый из этих элементов будет расположен в отдельной узловой точке.

Приставка *нано*, от лат. *nanus* (*карлик*), используется для обозначения количественных характеристик, намного меньших, чем общепринятые. В области синтаксиса это означает, что наносинтаксическое дерево включает в себя более мелкие элементы, чем синтаксическое. Слово, имеющее от двух до четырех уровней деления на синтаксическом дереве, в этом случае может быть разбито на более чем десять уровней.

В наносинтаксисе слово раскладывается на более мелкие, чем морфемы, элементы — образно говоря, «атомы». Кроме того, на наносинтаксическом дереве может быть отображен имплицитный уровень, то есть уровень подтекста, напрямую не выраженного в самом предложении. Например, в качестве такой узловой точки может быть представлено время (настоящее, прошедшее или будущее).

Каждое предложение обладает поверхностной и глубинной структурой. Глубинная структура одинакова для всех предложений с одинаковым, в то время как поверхностная структура уникальна.

Одно только слово *halachti* (произносится приблизительно [ха-лак-ти]), которое в переводе с иврита означает *я шел*, содержит в себе целый ряд утверждений: а) я — исполнитель этого действия; б) действие происходило в прошлом; в) действие уже закончено; г) действие представляло собой движение; д) подразумевается, что я шел куда-то; е) подразумевается, что я шел откуда-то. Все эти значения, заложенные в одном-единственном слове, невозможно отразить в традиционной модели струк-

турного анализа, вследствие чего объем передаваемой компьютеру информации в этом случае будет ограничен.

Таким образом, наносинтаксис дает лингвистам возможность рассматривать, анализировать и выявлять систему значений, скрытых в каждом конкретном предложении, с гораздо большей степенью детализации.

Наносинтаксис и обработка данных на естественном языке

В методах компьютерной обработки текстов на естественном языке (статистическая обработка, эвристический подход, метод сопоставления с образцом и т. д.) не задействуется все то множество значений, которым наделяется каждое слово в предложении. С помощью подобных методик невозможно полностью охватить систему значений того или иного высказывания, что ограничивает их практическую применимость в компьютерных программах. Отобразить всю систему значений, заложенных в предложении, можно только на основе модели полного синтаксического дерева. Возможность полного отображения смысла предложения дает наносинтаксический анализатор, в котором имитируется механизм понимания речи человеком. Наносинтаксис реализуется на уровне глубинной структуры, когда предложения с различными поверхностными, но одинаковыми глубинными структурами имеют одинаковое графическое представление. Это позволяет программе наносинтаксического анализа представлять фразы на естественном языке в структурированном виде. Ис-

пользуя эту информацию, компьютер может воспринимать структурированные данные и за счет этого выполнять устные команды, отвечать на вопросы и выполнять любые другие действия, указанные ему на языке пользователя, будь то английский, иврит или испанский.

В предложении из трех на первый взгляд простых слов «Я хочу читать» содержится огромный объем информации. Однако смысл этого предложения может быть понят программой и с помощью одного из названных выше методов вычислительной техники. Но что получится, если к этому предложению добавить еще одно слово: «Я хочу читать книгу»? Теперь оно состоит из четырех слов и конкретнее по своему значению. А предложение «Я хочу читать Джону», также состоящее из четырех слов, имеет другой смысл. Уже на этом этапе применение традиционных методов вычислительной техники становится затруднительным: для компьютера разница между книгой — «тем, что читают» — и Джоном — «тем, кому читают» — далеко не очевидна. Если же исходное предложение представить еще в двух вариантах: «Я хочу читать книгу Джону» и «Я хочу читать книгу о Джоне», мы получим два похожих по форме, но совершенно разных по смыслу предложения. В то же время для программы наносинтаксического анализа структурный разбор подобных предложений не представляет сложности.

С помощью методики наносинтаксиса компьютер представляет предложение не как набор слов, а как единое целое. Это позволяет программе правильно выполнять синтаксический анализ гораздо более сложных, чем приведенные выше,

предложений. В качестве примера рассмотрим следующее предложение.

«Джон посещал все бейсбольные матчи, а Джим — только часть из них». Что делал Джим? Джим посещал. Что посещал Джим? Бейсбольные матчи. В предложении эта информация прямо не выражена, она вытекает из предыдущего утверждения о том, что Джон посещал все бейсбольные матчи. Только такой целостный подход, «взгляд сверху», позволит компьютерной программе выявить смысл предложений подобного рода.

Попробуем объяснить разницу между традиционными методами обработки данных на естественном языке и наносинтаксической моделью. Представим, что американский турист, совершенно не говорящий по-немецки, потерялся в Берлине. Он спрашивает прохожего, как пройти к отелю и слышит в ответ: *Zwei links und drei rechts*. Согласитесь, ему это мало поможет... А если прохожий вместо этого даст туристу аэрофотоснимок Берлина, на котором точками обозначены места, где находятся сам турист и его отель? Это бы значительно облегчило бедняге жизнь. Так вот, наносинтаксис как раз и есть своеобразный аэрофотоснимок предложения.

Наносинтаксис в действии

Представим, что система «видео по запросу» — video on demand — выполняет команды на естественном языке, к примеру: «Я хочу посмотреть последний фильм с Бобом Рейнером». Или что персонажи новой игры для приставки Sony PlayStation 3 понимают человеческую речь: пользователю

достаточно произнести что-то вроде «Влезь на стену в глубине комнаты», и герой послушно выполнит эту команду. Или что график авиарейсов можно узнать, отправив обычное SMS-сообщение, например: «Во сколько прибывает рейс 323 в Лос-Анджелес?», и система отправит ответ прямо на мобильный телефон, с которого поступил запрос. Причем можно будет получить ответ на другом языке (в конце концов, на обоих языках это предложение имеет одинаковую глубинную структуру). Все эти и многие другие возможности открывает технология наносинтаксического анализа речи.

Одно из базовых правил наносинтаксиса заключается в отделении синтаксического анализа от логического и семантического этапов обработки данных на естественном языке. В ходе наносинтаксического анализа предложения «Как добраться до Нью-Йорка?» программе становится известно, что Нью-Йорк — это имя собственное, название конечного пункта маршрута. Однако это не означает, что ей известен город Нью-Йорк и как до него добраться. Эти сведения и предоставляет компьютеру функция составления логических выводов. На следующем этапе для выяснения маршрута программа может отправить запрос в соответствующую базу данных. Такая последовательность позволяет программе определять значение отдельных слов или терминов и контекст их употребления в соответствии с конкретной ситуацией: аббревиатура ПК может означать как «персональный компьютер», так и «Парижский клуб». При использовании других методов обработки данных на ес-

естественном языке значение слов должно быть предварительно заданным, что сужает сферу применения такого рода приложений.

Программа не в состоянии определить значение фразы в отрыве от контекста. Смысл предложения «Удалите то, что я хочу» невозможно понять без контекста даже человеку. Даже если к нему добавить пояснение: «Удалите то, что я хочу [, чтобы вы удалили]» или «Удалите то, что я хочу [сохранить]», без знания контекста, в котором употреблено словосочетание «я хочу», смысл предложения остается неясным. В то же время в решениях, основанных на технологии наносинтаксиса, функции синтаксического разбора и составления логических выводов полностью разделены. Смысл фразы может быть установлен на основе контекста — либо предварительно заданного, либо создаваемого в ходе диалога с текущим пользователем.

Возможности наносинтаксиса

Компьютер не может воспринимать человеческую речь, поскольку не способен расшифровывать закодированное в цепочку языковых знаков сообщение. Человек кодирует и расшифровывает эти сообщения естественным образом, благодаря способностям своего мозга. В отличие от синтаксиса, изучающего способы построения предложений, наносинтаксическая модель позволяет проводить анализ и раскрывать смысл предложения на значительно более глубоком уровне. Работа единственной в своем роде, недавно созданной программы наносинтаксического анализа основана на

принципе, аналогичном тому, которым пользуется человеческий мозг для понимания речи, благодаря чему в программе отображается вся система значений того или иного высказывания. Программа способна представлять неструктурированные высказывания на естественном языке в структурированном виде, поскольку действует на уровне глубинной структуры — общего смысла, которым обладают разные по форме предложения. Наносинтаксический анализатор рассматривает предложение в целом. В этой технологии этап синтаксического анализа отделяется от этапов составления логических выводов и семантической обработки языковых данных, что позволяет программе определять значение отдельных слов или терминов и контекст их употребления в соответствии с конкретной ситуацией.

Техническая реализация методики наносинтаксического анализа позволит совместить естественный (например, английский) и формальный (компьютерный) языки. Это открывает широкие возможности, которые могут быть реализованы в конкретных приложениях. Благодаря возможности задавать компьютеру вопросы о предмете поиска на естественном языке могут быть улучшены результаты работы поисковых систем, ускорена навигация веб-сайтов с большим объемом информации, усовершенствованы справочные службы, действующие посредством обмена SMS-сообщениями, значительно повышена результативность сетевой рекламы, а компьютерные игры могут стать гораздо более интерактивными. Наносинтаксическая модель может быть внедрена

в операционные системы, что позволит полностью изменить принцип взаимодействия человека и компьютера.

В области программ распознавания речи за счет понимания компьютером естественного языка может быть достигнуто более точное преобразование голосовых сигналов в текст: если программа «понимает» смысл сказанного, ей проще подобрать правильный вариант текста, согласующийся с входящим голосовым сигналом. В сочетании с технологиями распознавания речи функция «понимания» программой естественного языка позволит людям взаимодействовать с

бытовыми приборами в своем «умном доме», а также управлять настольным или карманным компьютером и мобильным телефоном посредством голосовых команд.

В конечном счете наносинтаксис является основой технологий будущего, которые полностью изменят способ взаимодействия человека и машины.

Шрули Табер — руководитель информационной службы компании Linguistic Agents; возглавляет проект Intelligent Action Engine.

Интернационализация и разработка — две части одного целого

Алан Хорват (Alan Horvath)

Выход на мировой рынок может стать серьезным испытанием для любой компании из-за связанных с ним временных и финансовых затрат, а потому исключительную важность приобретают его тщательное планирование и правильное проведение. Необходимо продумать процесс глобализации, а также следовать важному правилу — сначала закончить интернационализацию, и лишь затем приступить к локализации.

В данной статье рассматриваются проблемы, которые наиболее часто встречаются в процессе интернационализации. Здесь вы также найдете советы о том, как минимизировать временные затраты.

Значение слов

При описании процесса подготовки продуктов для выхода на зарубежные рынки используется множество специфических слов, и это сбивает с толку. Вот некоторые определения, которые проясняют картину.

Глобализация (g11n) включает в себя решение компании выйти на зарубежные рынки, а также интернационализировать и локализовать свои продукты.

Интернационализация (i18n) — это процесс разработки продукта, при котором заранее учитываются особенности языка, культурные традиции и инфраструктура других стран во избежание необходимости внесения изменений в продукт на этапе локализации.

Локализация (l10n) предполагает процесс адаптации продукта к условиям конкретного целевого рынка с лингвистической и культурной точек зрения.

Перевод — это преобразование слов из одного языка в другой. Опытный переводчик способен точно передать тончайшие оттенки, в то же время сохранив в переводе нюансы и стиль оригинала.

Интернационализация программного обеспечения

Предположим, вы создали программное средство, предназначенное для продажи в Соединенных Штатах. В вашей программе встречаются жестко закодированные термины, такие как *доллары* и *центы*, а также жестко закодированный символ \$. В качестве знака десятичной дроби вы использовали точку (.), а для разбивки многозначных чисел — запятую (,). Теперь компания планирует вы-

вести программу на зарубежные рынки, но использованные вами термины и символы не подходят для иностранных валют.

Существует два способа подготовки вашего программного средства для освоения новых рынков.

Первый заключается в том, что придется сделать копии программы и заменить термины и символы на обозначения, используемые в каждой отдельной стране.

А что делать, если будет обнаружена ошибка? Неужели придется вносить исправления в каждую версию программы? Если вы решите расширить функциональные возможности своего продукта, то внесение изменений в каждую из версий будет неизбежно.

Компаниям, планирующим выход на несколько рынков, этот способ однозначно не подходит. Изменение не подлежащих локализации файлов — это пустая трата времени.

И это еще не все. Поддерживает ли ваш программный код Юникод, благодаря которому можно работать с различными системами письма? Правильно ли программа выполняет сортировку и поиск символов во всех языках перевода? Прежде чем выходить на международный рынок, необходимо решить эти и другие задачи.

Второй, и гораздо более удобный способ, заключается в интернационализации программного обеспечения для выхода на мировой рынок с последующей локализацией.

Таким образом, необходимо убедиться в том, что продукт готов к интернационализации, и лишь затем начинать перевод его интерфейса, текста и файлов.

Десять советов по интернационализации программного обеспечения

Программное обеспечение — это компонент, которому следует уделить особое внимание. Ориентация на мировой рынок повлияет на каждое проектное решение как в программном коде, так и в пользовательском интерфейсе.

- 1) **Используйте функции и методы, поддерживающие Юникод.**
- 2) **Используемые средства сторонних разработчиков.** Необходимо с большим вниманием отнестись к выбору средств. Некоторые из них, например API-интерфейсы и встраиваемые дополнения, не поддерживают Юникод. Если вы не можете без них обойтись, постарайтесь выбрать правильные способы преобразования символов.
- 3) **Отделение пользовательского интерфейса.** Вынесите переводимые материалы. Жестко закодированные строки переводить довольно сложно: программный код постоянно меняется, и их нельзя будет переводить в процессе разработки. Используйте одну базовую основу кода для всех языков.
- 4) **Согласование слов.** Другой источник проблем при проведении локализации лежит в согласовании слов в строке. В разных языках порядок слов и формы множественного числа могут отличаться. Если прилагательные не будут согласовываться по роду, числу и падежу, это приведет к необратимым последствиям.
- 5) **Неоднозначность.** Составляя ресурсные файлы, не забывайте о перевод-

- чиках. В случаях, когда предложение неоднозначно, добавляйте комментарии. Это не только улучшит качество продукта, но и сократит время перевода.
- 6) **Увеличение размера.** Главное — помнить, что переведенные строки обычно имеют другую длину, чем строки исходного языка. Это особенно важно для диалоговых окон и меню, поэтому дополнительно предусмотрите по крайней мере 20–30% свободного пространства.
- 7) **Оформление.** Соблюдайте осторожность при создании значков и элементов графики. Не стоит вставлять в них текст, потому что перевод текста в картинке стоит довольно дорого и вдобавок может просто не поместиться. Старайтесь не использовать символов с культурным подтекстом: некоторые из них могут показаться жителям других стран непонятными или оскорбительными. На всякий случай следует проверить, будет ли ваш продукт работать на разных платформах, таких как PC, Mac, Linux и другие.
- 8) **Терминология.** Проверяйте единство терминологии. Если в вашей программе нет единства, то весь пакет будет еще более несогласованным. Средства управления терминологией помогут вам обеспечить ее единообразие.
- 9) **Тестирование региональных настроек.** Следует проверить работу вашей компьютерной программы в условиях целевого рынка, чтобы понять, решены ли все вопросы, связанные с региональной спецификой.
- 10) **Материалы для перевода.** Когда ваш продукт будет готов для перевода, необходимо подготовить соответствующие материалы. В этом наборе файлов должно находиться все необходимое для перевода и повторного компилирования языковых ресурсов, а также для тестирования приложения. Не стоит забывать о том, что при выходе вашей компании на рынки других стран могут возникнуть юридические сложности. Например, бывали случаи, когда компании-производители программного обеспечения встраивали в продукт свои стандартные условия пользовательских соглашений. А затем выяснялось, что использование некоторых из них неуместны в других странах. В большинстве случаев адвокатам компаний приходилось менять отдельные пункты соглашений. Следовательно, прежде чем начать процесс перевода, необходимо тщательно изучить законодательство стран целевого рынка.
- Обратите внимание, что существуют ограничения на экспорт программного обеспечения, содержащего технологию шифрования, а в некоторых странах на коммуникационное программное обеспечение распространяются особые правила.

Поддержка многобайтовых символов

Поставка продукции в азиатские страны многим компаниям кажется захватывающим и интересным опытом, но плата за ошибки может быть слишком высока.

В английском программном обеспечении для обозначения слов и чисел используется около 100 разных символов. Что касается азиатских языков, в них для отображения сообщений может использоваться более 10 000 символов.

Раньше для упрощения процесса перевода в программном обеспечении для хранения символов и текста использовались так называемые системы многобайтовых или двухбайтовых символов (DBCS). Сегодня их заменила кодировка Юникод, но в отдельных случаях может понадобиться поддержка многобайтовых символов.

Распространенной ошибкой при разработке программного обеспечения является использование сторонних продуктов, которые не предусматривают поддержку Юникода: библиотек ПО, динамически подключаемых библиотек (DLL), перемещаемых элементов управления (OCX) и т. п. Зачастую разработчики осознают это только после начала процесса перевода. В результате продукт приходится проектировать заново, чтобы все исправить, а это связано с дополнительными расходами.

Интернационализация документации

Планирование интернационализации документации занимает гораздо меньше времени. Вот несколько простых рекомендаций, которые помогут вам упростить локализацию документов.

Настольные редакционно-издательские системы. Выбрать одну настольную издательскую систему среди имеющегося разнообразия достаточно слож-

но. Ваш выбор зависит от того, будете ли вы использовать средства перевода, а также от типа документов и языка перевода. Необходимо решить, в каком виде будет представлен документ: печатном, электронном или в том и другом одновременно.

Если вы занимаетесь разработкой маркетинговых документов, наилучших результатов можно добиться при помощи таких систем, как FrameMaker, InDesign и QuarkXPress. Основным недостатком заключается в том, что процесс перевода таких документов сложнее. Однако ведущие средства памяти перевода превосходно справляются с файлами этого формата.

Кроме того, в азиатских странах довольно сложно найти опытного поставщика, использующего такие же настольно-издательские средства, так как цены на программное обеспечение в этих странах необычайно высоки, а программы трудны для освоения.

Если вы несмотря ни на что решили использовать эти приложения, следует помнить следующие правила.

- Старайтесь не создавать в документе несвязных текстовых окон. Обработка файлов, содержащих несвязные тексты при помощи средств перевода — довольно трудоемкий процесс. Лучше распределять текст по окнам. Тогда при переводе растянутый текст будет автоматически перемещаться из одного окна в другое, даже если оно находится на другой странице.
- Необходимо предусмотреть возможное увеличение размера текста, поэтому оставляйте на странице дополнительное место (20–30%). Об этом

следует помнить при интернационализации документов любого типа, но при работе с маркетинговыми материалами, где количество страниц обычно ограничено, это правило приобретает особую важность.

- Соблюдайте осторожность, выбирая снимки экранов для своего материала. Выбранный вами компонент вполне может оказаться последним в очереди на перевод, а потому локализованные экранные снимки будут еще недоступны во время подготовки к печати.
- Следует удостовериться, что вся графика, обрамленная текстом, связана с ним. Таким образом, если абзац перейдет на следующую страницу, изображения будут перемещаться вслед за ним. Если вы разрабатываете объемные технические документы, у вас будет большая свобода выбора. Хотя в данном случае предпочтительнее использовать такие средства, как FrameMaker и Microsoft Word для Windows, потому что в InDesign и QuarkXPress недостаточно функций для работы с документами большого размера.

Графика. Помните о том, что привлечение для перевода и создания графических изображений внешних компаний требует больших финансовых затрат, особенно для азиатских языков. Однако тщательное проектирование и качественная разработка графики сократит значительную часть расходов на ее локализацию. Использование графических изображений без текста снизит затраты более чем на 95%. Следует извлечь подлежащий локализации текст, внедренный в графический файл,

и поместить его в текст документа для повторного использования в памяти перевода. Замените извлеченный текст пронумерованными (а не обозначенными буквами) метками, расположив их по часовой стрелке. Затем сделайте перекрестные ссылки от чисел к тексту в главном документе. Это сделает возможным повторное использование графики во всех локализованных версиях.

Электронные или бумажные носители? Очевидным преимуществом документов в электронном виде является их стоимость. Создание, упаковка и доставка CD/DVD-версии обойдется гораздо дешевле, чем 15 толстых справочников. В случае необходимости книгу в электронном варианте нетрудно распечатать. Тем не менее отдельные части набора документов («Инструкция по применению» или «Руководство администратора») должны быть представлены в печатном виде.

Постраничное соответствие. Этой возможностью почти не пользуются, но она может быть полезна вашей организации. Ее главная задача — оптимизировать процесс осуществления технической поддержки. Допустим, у жителя другой страны возникли проблемы с вашим приложением. На месте проблеме решить не получается, поэтому требуется помощь более высокого уровня. Благодаря постраничному соответствию каждый пользователь будет направляться за решением к одной и той же странице документа, независимо от языка. Это также делает возможным создание стандартизированного упаковочного материала, так как все руководства будут одинакового размера. Выбрав этот путь,

будьте готовы к тщательному планированию процесса, а в крайнем случае и к тому, что придется пожертвовать эстетикой.

Альтернативный метод интернационализации документов. Рассмотрите системы управления информацией в качестве альтернативы традиционному подходу к созданию и обработке документов. Благодаря такой системе информацию можно вводить одновременно из любого количества мест, на любых языках. Существуют разумные способы хранения полученной информации, независимо от окончательного оформления. Информация может быть использована для создания публикаций практически в любом существующем формате: Help, HTML, DOC, XML, а также на разных языках. Суть этого подхода заключается в том, что создаваемая подобным образом информация уже интернационализована. В итоге значительно сокращаются общие затраты на перевод и локализа-

цию, а вместе с тем улучшаются качество и единообразие продукта.

Итоги

Интернационализация продукта играет первостепенную роль в процессе глобализации. При разработке продуктов необходимо учитывать особенности языка, культурных традиций и инфраструктуры других стран во избежание серьезных переделок продуктов. Стратегия проведения интернационализации, основанная на перечисленных выше правилах, поможет вашей компании сэкономить время и средства, а также достойно справиться с этой достаточно сложной задачей.

Алан Хорват — исполнительный директор компании STAR Group America, имеет степень бакалавра по деловому администрированию, специалист в области вычислительной техники и информатики.

Обобщение опыта: уроки локализации

Горан Нордлунд (Göran Nordlund)

Почему деятельность по локализации воспринимается компаниями как причина издержек, а не источник прибыли? Оглядываясь назад, на свой более чем 25-летний опыт разработки технической документации и ее локализации, я бы хотел не только обобщить свои наблюдения, но и попробовать ответить на вопрос, почему до сих пор многие компании видят в локализации «неизбежное зло», требующее больших затрат.

Как это обычно бывает, причину следует искать в отправной точке всего процесса, в данном случае — на этапе разработки технической документации. В 1980 году я начал карьеру технического писателя и заметил, что окружающим непонятны причины, побудившие меня предпочесть интересной работе инженера составление руководств для пользователей. В то время техническими писателями, как правило, становились те, кто не смог добиться успеха в основной специальности или по каким-то другим неопределенным обстоятельствам был вынужден сменить должность в организации.

Возможно, инженер я был не из лучших, однако истинной причиной был все-таки мой интерес к языкам и писа-

тельскому ремеслу. Да, конечно, многим это покажется странным.

По счастью, мой непосредственный руководитель в новой компании оказался дальновидным человеком. Все выпускаемые компанией материалы носили, по его мнению, маркетинговый характер, а потому качество содержания и графическое оформление считались столь же важными для технического руководства, что и для глянцевого проспекта. Как это часто бывает, таким «фантазерам» приходится тратить немало сил, чтобы убедить руководство предоставить ресурсы и оплатить затраты на соответствующий инструментарий и работу квалифицированных специалистов по составлению технической документации. Для этого понадобилось около пяти лет.

В основе такого непонимания лежит нехитрый принцип: «Написать руководство под силу кому угодно, а наши инженеры прекрасно владеют английским, так о чем же беспокоиться?»

Статус команды технических писателей упал до своего нормального — низшего в организации — уровня. Сегодня мало что изменилось. В компании, где я работал до выхода на пенсию, наш отдел состоял из высококвалифицированных сотрудников с опытом работы в медицинской сфере, языковыми и педа-

гогическими навыками, умением взаимодействовать с СМИ и доскональным знанием юридических вопросов. Тем не менее в наш адрес постоянно отпускали замечания: «Почему вы занимаетесь подобной работой при вашем уровне образования?», или ремарки по поводу недостаточно высокого уровня владения английским нашими техническими писателями (для которых английский язык был родным!); почему, мол, на печать 300-страничного руководства требуется несколько недель, когда это можно сделать за пару минут на обычном принтере; почему на перевод требуется так много времени и затрат, если даже чья-то там жена, сын или дочь могли бы перевести то же самое за несколько часов, к тому же практически бесплатно.

Пару лет назад мне встретилось явление одной шведской компании, занимавшейся продажами бытовой техники, о поиске «секретаря-помощника руководителя» в отдел маркетинга. Помимо обычной работы секретаря (заказа билетов, бронирования номеров в гостиницах и тому подобное) упоминались дополнительные обязанности этого сотрудника — написание руководств пользователя и перевод иноязычных документов! Просто «и жнец, и швец, и на дуде игрец». Надеюсь, зарплата на этой должности была соответствующей.

До тех пор, пока такой подход к подготовке технической документации сохраняется, у специалистов по локализации мало шансов призвать к серьезному отношению к своей работе или добиться признания ее важности.

Тем, кому приходилось слышать мои выступления на конференциях или чи-

тать мои статьи, известна моя главная заповедь: «Приведи в порядок язык оригинала, и ты решишь проблему перевода». Чтобы локализация была успешной, оригинальные документы должны быть написаны хорошим языком. Для этого нужны грамотные специалисты с языковыми и педагогическими навыками, для которых этот язык является родным, и удобные средства разработки материалов. Для нас это очевидно, но как убедить в этом руководство? Конечно, можно использовать такой аргумент, как повышение отдачи от инвестиций. Однако часто затраты окупаются не раньше чем через год, а то и позже, в условиях же современной экономики более важны ежеквартальные финансовые показатели. Кроме того, в этом вопросе есть ряд субъективных моментов. Что такое качество? Насколько важна удовлетворенность потребителей? Можно ли доказать, что они снова приобретут наше оборудование через несколько лет, потому что остались довольны прилагаемой документацией?

Взгляд на работу по переводу и локализации

Оставим вопрос разработки технической документации и обратимся к проблеме ее перевода и локализации. Предположим, язык оригинала «приведен в порядок». Уж теперь-то перевод пойдет «как по маслу» — или все-таки нет? Первый вопрос, который непременно задаст руководитель, выслушав ваше предложение: «Сколько это будет стоить?». За ним обязательно последует: «Сколько-сколько?! За что?». Создается впечатле-

ние, что с точки зрения руководства хорошо только то, что обходится дешево и делается быстро. По-моему, плохо переведенный текст, будь то руководство пользователя, красочный рекламный буклет, информация на передней панели бытового прибора или на экране компьютера, должен вызывать у потребителя подозрения: «Можно ли доверять продукции компании, которая допускает столь безграмотное использование языка?». Во многих странах плохой язык может быть расценен как оскорбление: если язык так плох, заслуживает ли внимания мысль, которую с его помощью хотят передать?

Как и в случае с разработкой технической документации, ее перевод считается простым делом, с которым может справиться любой. В конце концов, в современном мире люди свободно говорят на нескольких языках, а если нет, то у каждого найдется знакомый (жена или друг), который «в этом разбирается». Так зачем же тратить время и деньги на дорогостоящие услуги неторопливых переводчиков? Мне не раз приходилось слышать — мол, «брошу все и стану переводчиком: работа непыльная, а платят много». Такому человеку только и остается, что пожелать удачи.

Профессия переводчика требует серьезной подготовки, и часто эту работу недооценивают. Хороший переводчик должен свободно владеть по крайней мере двумя языками, иметь солидный багаж теоретических знаний и хорошо ориентироваться в своей тематике. Если вы специализируетесь на переводе инструкций для мобильных телефонов, вряд ли за один день вам под силу под-

готовить качественный перевод текста о фармацевтических препаратах или медицинском оборудовании. Связь между ценой и качеством вполне очевидна, и все же компании до сих пор ставят под сомнение обоснованность затрат на перевод. Но если продукт должен соответствовать самым высоким требованиям к качеству, почему то же самое не может относиться и к информации о нем, предназначенной для конечного пользователя?

Как донести до других важность своей работы?

Обобщая, можно сказать, что основная причина, по которой многие компании воспринимают деятельность по локализации как дополнительное бремя, заключается в непонимании ими характера этой работы. Что нужно сделать, чтобы это изменить? Конечно, не существует какого-то одного простого решения, способного изменить эту ситуацию, однако я бы предложил следующее.

- Убедите руководство повысить корпоративный статус деятельности по разработке и локализации технической документации.
- Требуйте выделения финансовых средств и времени на участие в конференциях по этой теме, в том числе международных. Не принимайте отговорки об отсутствии в бюджете компании достаточных средств.
- Проявляйте активность в сетевом общении специалистов (в том числе в различных сетевых конференциях), демонстрируйте результаты своей работы руководству.

- Проявляйте активность на конференциях, поборите свою стеснительность. Старайтесь каждый раз получить возможность выступить с презентацией. Консультативные советы всегда рады появлению интересных презентаций и новых участников.
- Делитесь опытом с коллегами, пишите статьи. Если вашу статью напечатают, разошлите этот номер журнала своим руководителям и коллегам.
- Измеряйте все, что поддается измерению. Даже если не поддается, все равно старайтесь вывести количественные показатели: в таком виде информацию проще донести до руководства.
- Пытайтесь выяснить, как окупаются издержки на локализацию на самых неожиданных участках работы, а не только в деятельности отдела локализации.

Следуя этим советам, легко доказать, что издержки на самом деле оборачиваются дополнительными преимуществами. Влияние таких факторов, как удовлетворенность потребителей, положительные отзывы покупателей о качестве сопутствующих материалов, меньшее количество звонков в службу поддержки и инцидентов, вызванных неверным толкованием инструкций по эксплуатации, возможно, трудно поддается измерению, однако это не означает, что его можно недооценивать.

Да и возможность предоставить потребителю все инструкции и справочные

материалы на его родном языке — это тоже эффективный маркетинговый инструмент.

Чем больше внимания я уделяю положительным моментам и тому, насколько очевидны преимущества по сравнению с необходимыми издержками, тем большее недоумение у меня вызывает тот факт, что после стольких лет существования отрасли нам все еще приходится бороться с этой проблемой. Уже давно пора признать: деятельность по локализации будет необходима всегда, и эта необходимость постоянно растет.

Сделайте трудности источником развития!

Подводя итоги своей многолетней работы в области локализации, мне бы хотелось поблагодарить всех тех замечательных специалистов, с которыми я познакомился на различных конференциях, а также читателей из самых разных стран за их положительные отзывы на мои статьи. На своем интереснейшем профессиональном пути я приобрел множество друзей, с которыми надеюсь встретиться в самом ближайшем будущем.

Горан Нордлунд — руководитель проекта по информированию и просвещению пользователей компании MAQUET Critical Care AB (Сольна, Швеция).

Web 2.0 — новое слово в локализации

Ультан О'Брайен (*Ultan O'Broin*)

Web 2.0 — это символ резких изменений, которым вскоре подвергнется сфера организованной локализации. Это то, чего мы так долго ждали. В 2005 году в своей революционной статье «Что такое Web 2.0?» Тим О'Рэйли рассказал о принципах работы этого средства, основанных на коллективном разуме, постоянных изменениях и массовом участии. Это поистине «веб как платформа», и не существует таких специализированных материалов, с которыми нельзя было бы здесь работать.

Web 2.0 пригоден не только для работы с английским языком. Это средство используется для перевода сайтов на основе технологии «социумных сетей» (вроде MySpace или китайских ресурсов MOP.com и WangYou.com). Ресурс YouTube планируется переводить на португальский, французский, итальянский, японский, голландский, польский и испанский языки. По результатам отчета компании Byte Level Research о глобализации Интернета за 2007 год, первое место среди глобализованных сайтов занимает Google, второе — Википедия. В Google переводы выполняются профессионалами лишь для нескольких основных языков, а остальную рабо-

ту выполняют добровольцы; переводы статей Википедии осуществляются исключительно добровольцами. Глобальный диалог во всемирной паутине и международное участие, равно как и связанные с ними ожидания, должным образом представлены такими инновационными проектами, как Meadan.com, сочетающими в себе машинный перевод и мгновенный обмен сообщениями.

Динамичность участия и разнообразие контента, характеризующие Web 2.0, ставят перед сферой «организованной» локализации, ее поставщиками и их клиентами вопрос об адаптации к реалиям меняющегося мира. Необходима соответствующая реакция и новые инновации со стороны отрасли, поскольку «социальный перевод» (то есть выполняемый бесплатно добровольцами) становится основным источником контента. Что же в действительности представляет собой социальный перевод?

1. Нулевое время на организационную подготовку: пользователи могут переводить англоязычный контент в реальном времени; пример — перевод добровольцами материалов для таких проектов, как OpenOffice.org и Википедия.
2. Конец симметричности контента веб-сайтов. Контент будет уникален для

каждого языка («истинная» локализация) и создавать его будут сами пользователи. Интернационализация и локализация проводятся только в отношении пользовательских интерфейсов и структур для перевода. Лидирующая сейчас модель, в которой английский язык играет роль базового, уходит в прошлое.

3. Постоянный пересмотр и обновление. Никаким системам управления глобализацией (GMS), средствам автоматизации и управления потоками операций не поспеть за скоростью возникновения, объемом и разнообразием пользовательского контента. Место GMS-систем займет человеческая изобретательность.
4. Исчезнет необходимость в проведении внутреннего контроля качества: такой контроль переходит в сферу деятельности заказчика. В соответствии с моделью «мудрости толпы» Джеймса Шуровьески, плохие переводы моментально заменяются более качественными — как это происходит в Википедии. Не нравится термин или его перевод? Возьми и замени его *сам* — *прямо сейчас*.
5. Быстро развивающиеся правила контента, в которых мало значения придается упорядоченной терминологии и стилю. Важнее становится донести смысл — даже если это предполагает отсутствие единообразия, использование сленга и разговорного языка. А как же иначе? По иронии судьбы сфера локализации, столь долго боровшаяся против «превращения в товар», теперь лишена такого понятия, как «определенный заказчик».

Существует только «текущий заказчик» — читатель. Хорошим подтверждением этого является всеобщее возмущение, вызванное попытками ввести правила и этические ограничения для блогов.

6. Проникновение на небольшие рынки осуществляется значительно быстрее благодаря нулевым затратам на локализацию и упразднению такого фактора, как время вывода продукта на рынок.
7. Последствия для заказчиков услуг локализации: сокращение времени для введения инноваций, создание меньшего объема контента, работа с большим разнообразием форматов, снижение затрат, а также концентрация на разработке и распространении средств для создания контента «глобальными» пользователями при международном участии.
8. Поставщики услуг локализации лишаются крупных надежных источников доходов («дойных коров»). Бизнес-модель здесь будет развиваться в направлении разработки и обеспечения потоков дохода, основанных на средствах и платформах. По сути в условиях перехода к открытым стандартам это может означать возникновение конкуренции между поставщиками и их заказчиками.
9. Появление новых форм конференций и других мероприятий в сфере локализации — бесплатных, в режиме онлайн, допускающих участие и совместную работу независимо от статуса. Гордиться гигантскими расходными бюджетами и счетами, выставленными заказчиком, стано-

вится так же нелепо, как огромными внедорожниками или большими объемами выброса углекислого газа в атмосферу.

Естественно, в определенных областях сохранится «традиционная» локализация. Заказчики из государственного сектора и межправительственных организаций с их малопонятными материалами и бюджетами (не говоря уже о деформирующем рынок большим объеме переводов для войны в Ираке) еще не скоро сойдут со сцены. Но рассматривать такие заказы как источник для исследований и карьерного роста будут лишь ленивые, с бюрократическим складом ума и отставшие от жизни — ведь это будет работа с далеким от идеала Web 1.0.

Итак, грядут захватывающие, но непростые времена. Вопросы, поднятые в данной статье, также обсуждаются в блоге www.multilingualblog.com. В предстоящих публикациях журнала Multilingual мы постараемся оценить

последствия «социального перевода» и глобального диалога — и, в соответствии с принципами Web 2.0, приглашаем вас принять активное участие в этом обсуждении.

Библиография

1. Anderson, Chris. *«The Long Tail: Why the Future of Business Is Selling Less of More»*, Hyperion, 2006.
2. O'Reilly, Tim. *«What Is Web 2.0?»*, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.
3. Джеймс Шуровьески (Surowiecki, James). *«Мудрость толпы: почему вместе мы умнее, чем поодиночке, и как коллективный разум влияет на бизнес, экономику, общество и государство»*, изд-во Вильямс, 2007.

Ультан О'Брайен — член редакционной коллегии журнала *MultiLingual* и сотрудник *Oracle Corporation*. Высказывания, содержащиеся в данной статье, не принадлежат компании *Oracle Corporation*.

Ведьма (отрывок)

Грегори Магвайр (Gregory Maguire)

ПРОЛОГ

Дорога из желтого кирпича

Высоко-высоко над страной Оз, будто подхваченная ветром пылинка, парила ведьма. Летние тучки и облака плыли мимо нее. Далеко внизу петляла Дорога из желтого кирпича, обезображенная зимними бурями и мотыгами вандалов, но неизменно ведущая в Изумрудный город. По дороге шли путники, избегая выбоин, обходя канавы и перепрыгивая через ямки. Они и не подозревали, что ждет впереди, а ведьма не собиралась их просвещать. Точно по перилам соскользнула она на метле с небес, стремительная, как ее верные слуги — летучие обезьяны, и примостилась на верхней ветке ивы, под которой путники устроили привал. Зажав метлу под мышкой, ведьма бесшумно полезла вниз, пока не различила фигуры. Тогда она замерла и стала прислушиваться.

Путников было четверо. Металлический блестящий человек с топором перебирал косматую гриву могучего зверя кошачьей породы — похоже, льва, — а тот ворчал и щерился от боли. Лежавшее рядом чучело развлекалось тем, что сдувало пух с одуванчика. Завеса листьев почти скрывала прислонившуюся к стволу девочку.

— Говорят, она гораздо хуже своей покойной сестры, — жаловался Лев. —

Всем ведьмам ведьма. Большая на голову. Психическая. Та еще картина.

— Ее кастрировали в детстве, — рассудительно сказал Железный Дровосек. — Она родилась гермафродитом или даже мальчиком.

— У тебя одно на уме, — проворчал Лев.

— Говорю, что слышал.

— Чего только не услышишь, — отмахнулся Лев. — Мне говорили, что ее бросила мать. Ее обижали в детстве. Она пристрастилась к лекарству от кожной болезни.

— Ей не повезло в любви, как и всем нам, — добавил Дровосек и печально приложил руку к груди.

— Она предпочитает женщин. — Страшила отбросил голый одуванчик и сел.

— Ее отверг женатый мужчина.

— Она и есть женатый мужчина.

От удивления ведьма чуть не свалилась с дерева. Она никогда не боялась сплетен, но что же о ней болтают по стране, если от первых встречных наслушаешься такого?!

— Она деспот. Страшный тиран, — убежденно продолжал Лев.

— Все-то тебе страшно, трусливое создание, — сказал Железный Дровосек и дернул Льва за гриву. — Я слышал, она отстаивает автономию для так называемых мигунов.

— Какой бы она ни была, наверняка она оплакивает погибшую сестру, — раздался голос девочки, такой глубокий и искренний, что у ведьмы мурашки побежали по коже.

— Ну-ну, не будем жалеть бедную ведьмочку, — сказал Железный Дровосек, притворно шмыгнув носом. — Без сердца у меня все равно не получится.

— Дороти права, — сказал Страшила. — Колдуньи тоже горюют.

Ведьму страшно разозлили непрощенные соблазна. Она поползла вокруг ствола, стараясь получше рассмотреть девочку. Ветер набирал силу. Страшила поежился и прижался ко Льву. Лев нежно обнял его в ответ.

— Будет гроза, — заметил Страшила. Где-то пророкотал гром.

— Ведьма летит, — сказал Железный Дровосек и ткнул пальцем в ребра Льва.

Перепуганный Лев с жалобным воем прыгнул на руки Страшиле и завалил его своим весом. Железный Дровосек рухнул сверху.

— Друзья, не укрыться ли нам от грозы? — спросила Дороти.

Окрепший ветер раздвинул завесу листвы, и ведьма увидела девочку. Она сидела, подобрав под себя ноги и обхватив колени. Это была не хрупкая куколка, а рослая деревенская девчушка в синем платье в белую клетку и переднике. На ногах у нее нервно поскуливал черный песик.

— Я понимаю, ты боишься грозы, — сказал Железный Дровосек. — После того, что ты перенесла, это естественно. Успокойся.

Ведьма вцепилась пальцами в кору дерева. Она все еще не могла разглядеть

девочкиного лица — только крепкие руки и темные волосы, заплетенные в косички. Кто она: опасный враг или безобидная одуванчиковая пушинка, случайно заброшенная сюда ветром? Казалось, достаточно посмотреть ей в лицо, и все станет ясно.

Но как ни тянулась ведьма, девочка постоянно отворачивалась, опасливо озираясь по сторонам.

— Гроза приближается, — обеспокоенно сказала она. — И быстро. — В ее голосе сквозили плаксивые нотки. — Я эти грозы знаю: как налетят — только держись.

— Здесь мы в безопасности, — напомнил ей Железный Дровосек.

— Ничего подобного! — возразила Дороти. — Дерево — самая высокая точка во всей округе, поэтому если ударит молния, то именно сюда. — Она прижала к себе собачку. — Помните, мы видели хижину? Поспешим туда. Пойдем, Страшила, побежали, не то ты первый сгоришь.

Девочка уже была на ногах и трусливо бежала по дороге. Спутники присоединились к ней во все нарастающей панике. Упали первые капли, и тут ведьма увидела — не лицо девочки, нет. Башмачки! Серебряные туфельки сестры переливались всеми цветами радуги даже в стущающейся грозовой темноте. Сверкали, как желтые бриллианты, как кровавый янтарь, как пламенные звезды.

Если бы она сразу заметила башмачки, то не слушала бы все эти бредни. Но девочка так неудачно сидела. Теперь же ведьма вспомнила про свою нужду. Башмачки! Разве мало она претерпела, разве не заслужила их? Она бы сию же минуту

упала камнем с неба и вырвала сокровище у нахалки, если бы только могла.

Если бы могла... Но гроза была для ведьмы страшнее, чем для принесенной ураганом девчонки и для соломенного чучела, которое могло вспыхнуть от молнии. Нельзя, нельзя было бросаться в открытый бой, когда вокруг так ужасно сыро. Ведьма забилась под могучие корни черной ивы, где ни одна капля ее не достанет, и приготовилась пережить грозу.

Ничего, она еще выберется. Жестокая страна Оз иссушила ее и отшвырнула прочь. Ведьму бросало то туда, то сюда, как семечко, слишком сухое, чтобы прижиться и прорасти. Но проклятье должно лежать на стране, а не на ней. Ведь, несмотря на искалеченную жизнь, разве не достались ей невиданные способности?

Не беда, что путники улизили. Она умеет ждать. Они еще встретятся.

ЧАСТЬ ПЕРВАЯ МАНЧУРИЯ

Колыбель зла

— Думаю, пришел мой срок, — сказала из постели женщина. — Смотри, как меня разнесло.

— Сегодня? — спросил муж. — Как раз в твоем духе: выбрать самое неподходящее время.

Он стоял у открытой двери и всматривался вдаль, поверх озера, полей и зеленых холмов, где едва виднелась деревня Закамышье и тянулся дым от утренних костров.

— Хуже день не придумаешь, — добавил он. — Естественно.

Жена зевнула.

— Можно подумать, я специально выбирала. С природой, милый, не поспоришь. Живот растет-растет, а когда расти больше некуда, приходится рожать. Тут уж ничего не поделаешь. Главное — не мешать.

Она приподнялась на локтях и выглянула из-за круглого живота.

— Такое чувство, будто я заложница в собственном теле. Пленница ребенка.

— Возьми себя в руки. — Муж подошел к постели, чтобы помочь ей подняться. — Считаю, что это испытание для твоего духа.

— Взять себя в руки? — повторила жена и рассмеялась. — Да разве ж это я? Так, одна оболочка, вместилище для растущего внутри паразита.

— Подумай обо мне, — попросил муж. Он и впрямь пытался уговорить ее сегодня не рожать.

— Фрекс, — решительно сказала она. — Когда извергается вулкан, ни один священник в мире не заставит его потухнуть.

— Что скажут мои собратья?

— О, наверняка что-нибудь вроде: «Брат Фрекспар, как ты позволил своей жене рожать первенца в столь ответственный день? До чего недалёковидно с твоей стороны, какая слабость. Нам придется лишиться тебя сана».

Она, конечно, дразнила мужа: снять с него сан было некому. Ближайший епископ находился слишком далеко, чтобы интересоваться здешними делами.

— Но нельзя же так не вовремя!

— Честное слово, Фрекс, ровно половина вины тут твоя. Ты все-таки тоже к этому причастен.

— Можно, конечно, рассуждать и так, хотя я сомневаюсь.

— Сомневаешься?!

Она запрокинула голову и расхохоталась. На шее напрягся мускул, спускавшийся к ямке над грудиной, как ручка серебряного половника. Даже после сна, с огромным животом, женщина была великолепно прекрасна. Распущенные волосы походили цветом на мокрую опавшую дубовую листву, играющую на солнце. Муж презирал ее за благородное происхождение, уважал за добровольный отказ от легкой жизни — и вдобавок к этому умудрялся еще и любить.

— Сомневаешься, что ты отец? — вдоволь насмеявшись, сказала она и уперлась рукой в подголовник. Фрекс взял жену за другую руку и помог сесть в постели. — Или что мужчины причастны к рождению детей?

Она поднялась — огромная, словно ходячий остров, и все еще посмеиваясь, вышла из дома. Одеваясь, муж еще долго слышал ее смех из уборной.

Фрекс расчесал бороду, намазал волосы маслом, зачесал их назад и закрепил заколкой из кости и сыромятной кожи. Его лицо сегодня должно быть открыто, а мимика выразительна и понятна на расстоянии. Он втер в брови немного сажи, провел румянами по щекам, подвел губы. Красивому священнику и доверия больше.

Во дворе, служившем им кухней, Мелена готовила завтрак. Она двигалась вперевалку, но не грузно, как все беременные, а словно гигантский воздушный шар, волочащий по земле веревки. В одной руке она несла сковороду, в другой — пару яиц и пучок лука. Тихо-тихо,

чтобы не услышал муж, она мурлыкала песню.

Фрекс тем временем облачился в рясу, застегнул белый воротник, обул сандалии и, достав из-под комода два письма от священника из деревни Суходрежье, тут же спрятал их за пояс, подалее от глаз жены. Если Мелена их увидит, еще чего доброго тоже захочет пойти в Закамьшье и поглазеть на предстоящее зрелище.

Пока Фрекс пробовал голос к выступлению, Мелена перемешивала яичницу деревянной ложкой. За озером звенели бубенчики коров, но она слышала не их, а что-то другое, внутри себя. Звук без мелодии, точно музыка из сна, которую не вспомнить, кроме как по оставленному приятному впечатлению. Может, это ребенок поет у нее в животе? Видно, будущий музыкант.

Из дома доносились слова Фрекса. Он импровизировал, разогревался перед боем, готовил доводы для спора, в очередной раз убеждал себя в своей правоте.

Как там пела когда-то ей няня?

Утром родился —

Горем умылся;

Днем рожден, значит

Горько поплачет;

Вечером роды —

Горе на годы;

Ночью рожденье —

Не исключенье.

Мелена вспоминала теперь песенку с улыбкой, как шутку. Горе — неизбежная составляющая жизни, а мы, тем не менее, продолжаем рожать детей.

«Дурочка! — в мыслях обругала ее няня. — Разве ты не знаешь, что мы ро-

жаем только в молодости, а когда нахлебався бед, наше благоразумное чрево в отращении засыхает?».

«А как же мужчины? — возразила Мелена. — Они-то не засыхают до самой смерти».

«Ох, эти мужчины, — вздохнула няня. — Уж на что мы, женщины, тугодумы, а эти вообще не способны учиться».

— Завтрак готов, — позвала мужа Мелена и выложила яичницу на деревянную тарелку. Ничего, ее сын не будет таким болваном, как остальные. Она научит его, как побороть горе.

— Настали трудные времена для нашего общества, — произнес Фрекс.

Для человека, осуждавшего мирские соблазны, он ел с большой изысканностью. Мелене нравилось смотреть, как ловко управляют с вилкой его пальцы. Она подозревала, что под маской праведного аскетизма скрывалась давняя мечта о роскоши.

— У нас что ни день, то трудные времена, — ответила ему жена, как возразил бы возможный противник. Дурачок, он не слышал сарказма в ее голосе.

— Мы стоим на распутье. Надвигается идолопоклонство. Наши вековые устои в опасности! Истина под сомнением! Добродетель забыта!

Он не столько разговаривал с ней, сколько готовил обличительную речь. Была в его характере какая-то отчаянность, которую он, в отличие от многих других мужчин, умел оборачивать на благо работе.

Опершись о стол, Мелена осторожно села. В голове у нее пел целый бессловесный хор. Неужели так всегда бывает перед родами? Можно, конечно,

спросить местных сплетниц-повитух, которые придут сегодня днем и начнут ворчать, глядя на ее раздутый живот, но Мелена боялась показаться смешной. Пусть ей не избавиться от городского произношения, которое соседки считали вычурным, так хоть скроет свое невежество в столь примитивных делах.

Фрекс заметил ее молчание.

— Ты не сердись, что я тебя оставляю?

— Сержусь?

Она удивленно подняла брови, будто даже мысль об этом не могла прийти ей в голову.

— Историю движут ничтожные жизни маленьких людей, а также могучие неведомые силы, — сказал Фрекс. — Нельзя одновременно следить и за тем, и за другим.

— Жизнь нашего сына не обязательно будет ничтожной!

— Не время спорить. Ты что же, хочешь отвлечь меня от священной работы? На Закамьшье надвигается страшная угроза. Если я закрою на нее глаза, то никогда себя не прощу.

Он говорил совершенно искренне. Когда-то Мелена влюбилась в Фрекса за его самоотверженную веру. Теперь за нее же она его ненавидела.

— Будут новые угрозы, — сказала она и в последний раз добавила: — Твой первенец родится только однажды, и, судя по всему, именно сегодня.

— Будут другие дети, — ответил муж ее же словами.

Мелена отвернулась, чтобы он не увидел ярость на ее лице. Может, и грешно злиться на мужа, но она старалась об этом не думать. Со священником в доме

ей и так хватало религии. Она угрюмо молчала. Фрекс ел.

— Это дьявол, — вздохнул он. — Сам дьявол приближается.

— Как ты можешь говорить такое перед рождением нашего сына? — возмутилась Мелена.

— Я про искушение в Закамышье. Неужели непонятно?

— Все равно, слова есть слова, и кое-кого поминать опасно. Я не прошу твоего безраздельного внимания, Фрекс, но чуть-чуть поддержки не повредит.

Мелена в сердцах бросила сковороду о скамейку перед домом.

— Взаимно. Думаешь, что мне сегодня предстоит? Как убедить прихожан отвернуться от манящих идолов? Наверняка я проиграю необычному зрелищу. Видишь, тебя сегодня ждет радость материнства, а меня — горечь поражения.

Даже эти слова Фрекс произносил с достоинством: пострадать за святое дело для него было верхом добродетели. Разве может с этим сравниться низкое, шумное и кровавое занятие — деторождение?

Он поднялся уходить. Свежий ветер с озера разгонял тянущийся к небу дым из печи. «Прямо водоворот», — думала Мелена, глядя, как кругами вьется дым.

— Счастливо оставаться, дорогая, — сказал Фрекс, уже принявший строгий наставнический облик.

— Счастливого пути, — вздохнула Мелена. Где-то глубоко внутри нее брыкнулся ребенок, и скрутило живот. — Я буду думать о тебе, моя надежда и опора. Смотри, чтобы тебя там не убили.

— На все воля божья, — ответил Фрекс.

— И моя тоже, — кощунственно бросила она.

— Свою волю направляй на то, что тебе подвластно, — назидательно изрек муж. Теперь он был священником, а она — грешницей. Не самая приятная роль.

— Прощай, — сказала Мелена и, вместо того чтобы провожать мужа, скрылась в вонючей уборной.

Часы Дракона времени

Фрекс хоть вида и не показывал, но за жену волновался. Он остановился у первой же рыбацкой хижины и переговорил с выглянувшим из двери мужчиной. Не может ли тот попросить кого-нибудь из женщин провести этот день и, если понадобится, ночь у Мелены? Фрекс был бы очень признателен. Рыбак согласился. Священник кивнул с вялой улыбкой на лице. Он знал, что его жену недолюбливают.

По пути вокруг озера в Закамышье Фрекс остановился у поваленного дерева и вытащил из-за пояса спрятанные письма. Их автор, дальний кузен Фрекса, не пожалел времени и ценных чернил, чтобы описать увиденного идола, Часы Дракона времени. Готовясь к праведной схватке, Фрекс перечитывал рассказ о боготворимых часах.

«Я пишу второпях, брат Фрекспар, пока свежи воспоминания.

Часы Дракона времени установлены на фургоне и в высоту достигают роста жирафа. По сути, это всего лишь передвижной балаган, с каждой стороны которого есть задернутая занавесом сцена. На плоской крыше фургона лежит механический дракон с зеленой кожей, серебряными когтями и глазами из драгоценных камней.

Кожа его покрыта чешуей из уложенных внахлест медных, бронзовых и железных кружков. Под кожей скрывается хитрый механизм, позволяющий дракону крутиться на пьедестале, махать крыльями, которые шумят, как кузнечные меха, и рыгать огненные шары вонючей серы.

На сценах под драконом, среди декораций, расставлены кукольные фигурки — карикатуры на крестьян и дворян, животных и героев сказок. Даже на святых, брат Фрекспар, на наших унионистских братьев. Какое кощунство! Фигурки движутся на шарнирах, выкатываются из дверей, кланяются, танцуют и всячески паясничают.

Кто придумал Дракона времени, этого лжепророка, посягнувшего на истинную веру? Управляют механизмом гном и горстка женоподобных юнцов. Кому от Дракона польза, кроме этих развратников?»

Во втором письме кузен предупредил, что Часы направляются в Закамьшье, и рассказывал об увиденном представлении.

«Представление началось струнной музыкой и шумом гремящих костей. Толпа со вздохом изумления подошла ближе. Раздвинулся занавес, и на сцене осветилось окно. В окне виднелась постель, где лежали две фигурки: муж и жена. Муж спал, жена вздыхала над ним. Жестом она показала, какой маленький у него инструмент. Толпа зашла от смеха. Покривлявшись, жена устроилась в кровати, а когда она захрапела, муж встал и украдкой выскользнул из дома.

Стоило ему выйти, Дракон пришел в движение. Он повернулся на своем постаменте и протянул к толпе лапу, указывая

когтем на землекопа Грайна, который всегда был верным, пусть и не очень внимательным мужем. Потом дракон поднялся на задние лапы и манящим движением пальцев выделил из толпы вдову Летту и ее незамужнюю дочь с большими, вытирающими зубами. Толпа отхлынула от этих троих, словно от прокаженных.

Дракон улегся и махнул крылом над другой сценой. Зажегся свет, и стало видно, как муж идет в ночи. Появилась кукла взломаченной женщины с пуницовыми, как бы от стыда, щеками, которая тащила за собой упирающуюся зубастую дочь. Они поцеловали мужа и стянули с него кожаные штаны. Под ними оказалось целых два достоинства: одно спереди, второе сзади. Вдова усадила дочь на коротенький отросток спереди, а сама предпочла более внушительный орган сзади. Куклы задержались, испуская блаженные стоны. Напрыгавшись, вдова и дочь слезли с развратника, снова поцеловали его и тут же одновременно двинули коленом, каждая со своей стороны. Муж завертелся, хватаясь за пострадавшие органы.

Толпа бесновалась. У настоящего Грайна выступили капли пота размером с виноградину. Летта пыталась смеяться со всеми, а ее дочери и след простыл. Тем же вечером распаленные соседи скрутили Грайна и стянули с него штаны, проверить, действительно ли у него то уродство, которое показал Дракон. Летту сторонились. Ее дочь так и не видели, и мы боимся, как бы не было беды.

Хорошо хоть Грайна не убили. Но кто знает, какой отпечаток наложило происшедшее на наши души. Душа — заложница в теле, и любая мерзость может ее искалечить. Ты согласен со мной?»

Иногда Фрексу казалось, что все маломальские волшебники в стране, каждая бродячая колдунья и каждый беззубый кудесник, словно по сговору обрушились на прежде заброшенную Вендову пустошь. Он знал своих прихожан: это были трудолюбивые, но простые и бедные люди. Чем дольше тянулась засуха, тем меньше оставалась в них веры в Единого Безымянного Бога. Часы Дракона времени покажутся им не столько чудом механики, сколько волшебством. Чтобы образумить народ, придется зывать к самым глубинам веры. А если они не устоят перед искусом и поддадутся на призывы к насилию — что тогда?

Нет, невозможно. Разве не он их священник? Разве не он лечил им зубы, хоронил детей и благословлял дома? Разве

не терпел унижений ради них, не бродил по деревням со спутанной бородой и кружкой для подавания, оставив несчастную Мелену одну? Скольким он для них жертвовал! Прихожане перед ним в неплатном долгу.

Расправив плечи и гордо подняв голову, священник зашагал дальше, не обращая внимания на кислый привкус страха во рту. Небо потемнело от песка и пыли, а свистящий по холмам ветер жалобно стонал, будто заранее оплакивал обреченное на провал начинание.

Maguire, Gregory, «Wicked». 1995.

Перевод Андрея Азова.

Источник: www.bakanov.org (отрывок публикуется с разрешения владельца ресурса www.bakanov.org).

Тайный дневник Розовой Гвоздики (отрывок)

Лорен Виллиг (*Lauren Willig*)

Пролог

Метро встало. Снова!

Поднявшись на цыпочки, я схватилась за поручень и тут же врезалась носом в плечо стоящего рядом мужчины. Француз, судя по чёрной водолазке и приторно-сладкому одеколону. Поспешно извиняясь, я шагнула в сторону, споткнулась о стильный зонт от Ральфа Лорена и упала на чьи-то джинсовые колени.

— Здрóрово! — подмигнул джинсовый, когда я попыталась подняться.

Как любят англичане это «здорово» и как ловко используют! В зависимости от ситуации оно может означать от «привет» и «спасибо» до «у тебя отличная задница». Я густо покраснела (румянец никогда мне не шел) и стала искать, куда бы спрятаться. Вагон битком набит усталыми, раздраженными лондонцами, спешащими домой. Между ними не протиснуться даже тощей змее, не говоря уже о здоровой американской девушке, которая в последние два месяца питалась исключительно жареной картошкой и рыбой.

Получается больше шестидесяти порций жареной картошки! Хотя, если живешь на съемной квартире, а кухня размером с горошину, тут не до кулинарных изысков.

Снова прислонившись к ухмыляющемуся французу, я уже в который

раз спросила себя, чего ради приехала в Лондон.

Я слишком долго просидела в гарвардской библиотеке Виденера, наблюдая за студентами, суетящимися у туннеля возле входа. Согнувшись вдвое под тяжестью своих ранцев, они больше всего напоминали муравьев. Пожалуй, не зря я получила стипендию на год исследовательской работы в Британской библиотеке! Больше никаких студенческих работ! Никаких слайдов и учебных фильмов! Никакого Гранта!

Грант...

Вспомнив о нем, я тут же внутренне содрогнулась. Грант! Еще одна причина, по которой я давлюсь в душном вагоне лондонского метро вместо того, чтобы смотреть микрофильмы в Виденере.

Я бросила Гранта. Ну, что-то вроде того... Естественно, обнаружив его в объятиях молодой аспирантки факультета истории, я не могла относиться к нему по-прежнему. Именно я стащила с пальца кольцо и швырнула в лицо изменнику в лучших традициях мыльных опер.

Вагон рывком сдвинулся с места, и пассажиры радостно загудели. Только бы снова не упасть на джинсовые колени! Одно падение еще может считаться неловкостью, а два подряд скорее всего примут за приглашение.

Все мужчины, которые меня интересуют в данный момент, давно мертвы: Очный Цвет, Пурпурная Горечавка, Розовая Гвоздика... Их имена возвращают в ту далекую эпоху, когда мужчины носили бриджи и сюртуки, обменивались колкостями поострее, чем концы их шпаг, и умели быть героями.

Очный Цвет спас от гильотины бесчисленное множество людей, Пурпурная Горечавка сводил с ума французское министерство полиции и сорвал по крайней мере два покушения на короля Георга III*, а Розовая Гвоздика... В период между 1803 и 1814 годами о нем писали все лондонские газеты.

Неудивительно! Он оказался самым неуловимым из всей тройцы.

Личности первых двоих, Очного Цвета и Пурпурной Горечавки, французы установили. Это оказались сэр Перси Блейкни и лорд Ричард Селвик. В результате оба ушли в отставку и дожили век в своих поместьях, потчюя многочисленную родню рассказами о подвигах во Франции. А вот Розовую Гвоздику так и не поймали.

Пока не поймали!

Именно это я и собиралась сделать — отыскать неуловимого шпиона среди архивов английских библиотек и наконец сорвать с него маску. Может, мне повезет больше, чем французским сыщикам?

Естественно, мой академический руководитель и не догадывался, чем вызван такой интерес. Обсуждая с ним

тему диссертации, я перефразировала ее по-иному. Пришлось прикрываться пробелами в историографии, значением шпионажа в формировании общественного мнения начала девятнадцатого века и другими заумными доводами. В результате работа была названа «Роль аристократического шпионажа в войнах с Францией 1789–1815 годов». Сухо и непонятно! «Почему мне нравятся мужчины в черных масках» звучит гораздо лучше, но ученый совет вряд ли одобрил бы такое название.

Для начала совершенно логичным казалось отправиться в Кембридж. Три аристократа, сводившие с ума французов, вне всякого сомнения, поддерживали переписку. В начале девятнадцатого века высший свет Англии был довольно тесным мирком, так что они наверняка делились знаниями и опытом. В библиотеке я наткнулась на обширную переписку между сэром Перси Блейкни и лордом Ричардом Селвиком. Уверена, в их письмах есть какой-то намек или даже неосторожная описка, которая приведет к Розовой Гвоздике.

Я не нашла ничего. Ничего, даже просмотрев финансовые отчеты поместья Блейкни за двадцать лет и счета Селвиков! Я не поленилась и посетила Государственный архив в Кью и, пройдя сквозь металлоискатель, наконец попала в отдел, где хранились данные министерства обороны девятнадцатого века.

Да, не зря разведку называют секретной службой! Ничего, опять ничего! В официальных отчетах не содержалось даже косвенного упоминания о «нашем розовом друге».

* Георг III — полное имя Георг Вильгельм Фридрих, король Великобритании и Ирландии, курфюрст Ганноверский, внук Георга II. Взошел на престол в 1760 году.

Я запаниковала: не хотелось писать о шпионаже, возводя его в аллгорию мужества и отваги. Значит, придется прибегнуть к последнему варианту. Итак, я сидела в читальном зале магазина «Уотерстоунз»^{**}, писала письма всем оставшимся в живых наследникам сэра Перси Блейкни и лорда Ричарда Селвика. Попасть в семейные архивы я и не мечтала, так велико было мое отчаяние!

Полузабытых историй о похождениях бравых прапрадедушек мне вполне достаточно. Расскажите мне хоть что-нибудь, и я пойму, где искать дальше!

Написав двадцать писем, я получила три ответа.

Владельцы поместья Блейкни ограничились расписанием работы музея сэра Перси, зато приложили приглашение на осенний фестиваль почитателей Очного Цвета. Да, очень здорово наблюдать, как туристы разгуливают в черных плащах, крутят моноклями и кричат: «Империя в опасности!».

Ответ хозяина Селвик-Холла былещё более обескураживающим. Письмо на украшенном фамильным гербом бланке дышало высокомерием и должно было поставить меня на место. Селвик-Холл — по-прежнему частная собственность и закрыт для всех посетителей без исключения. Предназначенные для просмотра документы представлены в Британской библиотеке. Хотя мистер Колин Селвик и не написал «Отвали!» открытым текстом, он явно имел в виду именно это.

Но ведь самое главное — результат, верно? А сегодня результатом было то,

^{**} «Дебретт» — ежегодный справочник английского дворянства, издаётся с 1802 года.

что миссис Арабелла Селвик-Олдерли ждет меня по адресу — поднимаясь по эскалатору станции в южном Кенсингтоне, я вытащила из кармана потрепанную ксерокопию письма — Онслоу-сквер, дом сорок три.

Шел дождь, как всегда, когда забудешь зонт.

Остановившись на пороге дома номер сорок три по Онслоу-сквер, я окинула себя критическим взглядом. Коричневые замшевые сапожки от Джимми Шу от дождя совершенно потеряли вид. Элегантная юбка «в елочку» съехала на бок, а на толстом бежевом свитере огромное коричневое пятно — результат неудачного столкновения с чьим-то кофе в Британской библиотеке.

Да, миссис Селвик-Олдерли будет очарована!

Торопливо поправив юбку, я нажала на кнопку звонка.

— Да? — протребезжал старушечий голос.

— Это Элоиза! — проговорила я. Ненавижу домофоны! Никогда не уверена, что говорю с тем, с кем нужно. — Элоиза Келли, по поводу Пурпурной Горечавки!

Дверь тут же открылась.

— Поднимайтесь! — велела хозяйка.

Запрокинув голову, я оглядела лестницу и, никого не увидев, нарисовала мысленный портрет миссис Селвик-Олдерли. Доброе морщинистое лицо, седые кудри, старомодный твидовый костюм, а в руках палка. Последовав указаниям свыше, я стала подниматься по лестнице и одновременно вспоминать небольшую речь, приготовленную накануне. Обязательно поблагодарю за любезное пригла-

шение, а потом улыбнусь и пообещаю сделать все, что в моих скромных силах, чтобы вырвать ее достойного предка из плена исторического забвения. Из уважения к старости нужно говорить громко, четко и не торопясь.

— Бедная девочка, вы похожи на выжатый лимон!

Элегантная женщина в темно-синем костюме из пушистого букле с оригинальным малиновым кашне смотрела на меня с явным сожалением. Белые, как снег, волосы, (ну, хоть в чем-то я не ошиблась) уложены в высокую сложную прическу, которая делала миссис Селвик-Олдерли похожей на королеву. Благодаря осанке она казалась очень высокой, даже выше меня (сто семьдесят пять сантиметров роста плюс пятисантиметровые каблук). Да, остеопорозом она явно не страдает!

Приготовленная речь вытекла из памяти, совсем как капли дождя с подола плаща.

— Э-э-э, здравствуйте, — запинаясь, проговорила я.

— Ужасная погода, правда? — миссис Селвик-Олдерли ввела меня в кремовое фойе и показала, куда повесить плащ. — Вы добирались от Британской библиотеки?

Вслед за хозяйкой я прошла в гостиную, раскисшие сапожки хлюпали по мягчайшему персидскому ковру. Легкий диван и два кресла, обитые вощёным ситцем, кто-то придвинул к мраморному камину, в котором уютно потрескивал огонь. Похоже, миссис Селвик-Олдерли любит читать: на журнальном столике много книг, но сегодня их сдвинули в сторону, освободив место для чаепития.

Взглянув на столик, хозяйка негромко вскрикнула.

— Печенье забыла! Сейчас вернусь, а вы устраивайтесь поудобнее!

Поудобнее... Легко сказать. Миссис Селвик-Олдерли, конечно, очаровательна, но в ее обществе я чувствую себя как школьница в кабинете директрисы.

Заложив руки за спину, я подошла к камину. Все как обычно: фотографии родственников, расставленные в хронологической последовательности. Справа покоричневевший портрет совсем молодой девушки с короткими кудрями в стиле конца тридцатых. Незнакомка задумчиво смотрела куда-то вдаль. Остальные снимки поновее и, в основном, групповые: по праздничному одетые люди, улыбающиеся в камеру фотографа. Судя по всему, семья большая и очень дружная.

Одна из фотографий показалась особенно интересной. Она стояла по середине полки, скрытая портретом двух маленьких девочек с незабудками в руках. В отличие от большинства снимков, на этом был только один мужчина верхом на лошади. Вьющиеся светло-русые волосы развевались на ветру. Изящно очерченные губы и высокие скулы напоминали миссис Селвик-Олдерли. Но если она по красоте и элегантности напоминала фарфоровую статуэтку, то мужчина так и излучал жизненную энергию. Даже с фотографии он улыбался так заразительно, что остаться равнодушной было просто невозможно.

Тут вернулась хозяйка с большим блюдом, наполненным шоколадным печеньем.

Я вздрогнула, будто меня поймали за чем-то постыдным.

Миссис Селвик-Олдерли поставила печенье на столик.

— Разглядываете снимки? Чужие фотографии притягивают, как магнит, вам так не кажется?

— Да, пожалуй.

Я робко присела на диван: жаль дорогой обивки, ведь юбка еще не высохла!

— Просматривать чужие фотографии — все равно, что читать книгу, — нерешительно проговорила я. — А тут у вас столько старых снимков... Поневоле начинаешь гадать, кем были эти мужчины и женщины, как жили...

— Именно потому люди увлекаются историей, — проговорила миссис Селвик-Олдерли, поднимая изящную фарфоровую чашечку.

Мы разговорились об английской литературе, и я почувствовала себя в своей тарелке.

Немного доброжелательности со стороны хозяйки, и я уже рассказывала о том, как увлеклась историей (следствие бесконтрольного «глотания» книг в нежном возрасте), о ситуации на историческом факультете в Гарварде (настолько сложной и запутанной, что и вспоминать не хочется) и причине, по которой я приехала в Англию. Рассказ стремительно приближался к тому, что было не так с Грантом (все!), когда я поспешно сменила тему, спросив миссис Селвик-Олдерли, не рассказывали ли ей в детстве о знаменитых шпионах девятнадцатого века.

— Ну конечно же, милая, конечно! — грустно улыбнулась миссис Селвик-Олдерли. — Мы только в шпионов и играли, по очереди примеряя роли Пурпурной Горечавки и Розовой Гвоздики. А вот моему кузену Чарльзу больше нравился

злой сыщик Деларош. Как здорово он имитировал французский акцент! Вылитый Морис Шевалье***! Сколько лет прошло, а мне до сих пор смешно. Чарли приклеивал усы, (мы были уверены, что все злодеи носят усы), сооружал из мамино платка накидку и носился по саду, осыпая проклятьями Пурпурную Горечавку.

— А вам кто нравился? — спросила я, вдохновленная ее рассказом.

— Естественно, Розовая Гвоздика!

Мы заговорщицки переглянулись.

— Вас ведь интересует именно Розовая Гвоздика? — многозначительно спросила миссис Селвик-Олдерли. — Вы, кажется, диссертацию пишете?

— Все правильно! — Я вкратце описала главы, которые собиралась посвятить похождениям Очного Цвета и искусству перевоплощения Пурпурной Горечавки. Об этих двоих информации было более чем достаточно... — А вот о Розовой Гвоздике ничего не нашла, — уныло закончила я. — В газетах только пишут, что ему поручались самые сложные задания.

— На что же вы надеетесь?

Я уставилась на дно чашки.

— Ну, это мечта любого историка! Найти потерянную рукопись, озаглавленную

*** Морис Шевалье — французский певец, киноактер. В 1912–18 гг. в театре оперетты, затем на эстраде (шансонье). Получил широкую известность в конце 20-х и в 30-х гг. С 1911 работал в кино, снимался в фильмах «Парад любви» (1931), «Веселая вдова» (1934), «Молчание — золото» (1947), «Век любви» (1956), «Канкан» (1962), «Фанни» (1963), «Дети капитана Гранта» (1965, роль Паганеля) и др.

ную «Как и почему я стал Розовой Гвоздикой». Мне бы хватило письма или сводки министерства обороны, проливающей свет на его личность. Да что угодно, что поможет понять, где искать дальше!

— Кажется, я смогу вам помочь.

Губы миссис Селвик-Олдерли дрогнули в чуть заметной улыбке.

— Правда? — оживилась я и выпрямилась так резко, что чуть не выронила чашку. — Вы что-то о нём слышали?

В светло-голубых глазах миссис Селвик-Олдерли загорелись огоньки.

— И не только слышала! — прошептала она, наклоняясь ко мне.

От возбуждения закружилась голова. Что же это? Старое письмо, скорее даже предсмертное послание, передававшееся из поколения в поколение? Сейчас его хранительницей является миссис Селвик-Олдерли... Но с какой радости она вдруг раскроет мне священную тайну семьи? Да, замечталась, пора вернуться с небес на землю.

Арабелла Селвик-Олдерли грациозно поднялась с диванчика.

— Пойдемте! — поманила она, оставив чашечку на журнальном столике.

Моя чашка громко звякнула о блюдце и чудом не разбилась, когда я последовала за хозяйкой к двум большим окнам, выходящим на площадь. Между ними висело несколько портретов-миниатюр, и на секунду я испугалась, что миссис Селвик-Олдерли просто хочет показать мне картины. Миниатюры неплохие, но причем тут я?

Справа от окон маленький восьмиугольный столик с розовой фарфоровой конфетницей, а слева книжные полки, тянущиеся через всю гостиную.

Однако миссис Селвик-Олдерли не интересовали ни столик, ни полки. К моему вящему удивлению она опустилась на колени перед массивным сундуком, стоящим прямо под миниатюрами. Декоративно-прикладным (или как его там называют?) искусством я особо не увлекалась, но уже достаточно походила по английским музеям, чтобы понять, что сундук относится к началу восемнадцатого века, если только это не очень удачная подделка. Крышку украшала мозаика из разных сортов дерева, изображающая цветы, птиц, а в самом центре большое райское дерево.

Из кармана миссис Селвик-Олдерли появился необычной формы ключ.

— В этом сундуке скрывается разгадка тайны Розовой Гвоздики, — объявила женщина, поднеся ключ к скважине.

Бронзовый замок щелкнул, и крышка тут же поднялась — несмотря на возраст, все механизмы работали отлично. Через секунду я уже стояла на коленях перед сундуком вместе с хозяйкой.

Первой реакцией было сильное разочарование: ни тебе старых газет, ни писем. Я с удивлением смотрела на выцветший веер, пожелтевший отрез украшенной вышивкой ткани и полуистлевшие останки букета, перевязанного превратившейся в лохмотья ленточкой. Там были еще какие-то древние безделушки, но я чувствовала себя обманутой и начала подниматься.

Никогда не стоит спешить с выводами! Тонкие с голубыми прожилками вен руки миссис Селвик-Олдерли потянули за бархатную обивку, аккуратно снимая первое отделение. Итак, в сундуке имелось двойное дно и второе отделение, а

в нем... Я снова встала на колени, побелевшие пальцы так и вцепились в украшенную мозаикой крышку.

— Это...это поразительно! — пролепетала я. — И все относится...

— Да, к началу девятнадцатого века, — договорила за меня миссис Селвик-Олдерли, любовно осматривая содержимое сундука. — Все данные разложены в хронологическом порядке, так что разберетесь без труда. — Женщина вытащила какой-то листочек и тут же отложила в сторону. — Вот это к делу совершенно не относится!

Она усмехнулась и на этот раз достала из сундука небольшую коробку из специального картона. Именно в таких в библиотеках хранят старые книги.

— Вам лучше начать с этого, — посоветовала миссис Селвик-Олдерли, — вместе с Амели.

— Амели? — переспросила я, осторожно расправляя шпагат, которым была перевязана коробка.

Миссис Селвик-Олдерли уже собралась отвечать, а потом вдруг осеклась и не без труда встала на ноги.

— Письма сами все объяснят, — заявила она. — Если что, я в кабинете, это следующая дверь направо.

— Но кто он такой, этот Розовая Гвоздика? — взмолилась я, вслед за хозяйкой поворачиваясь к двери.

— Почитайте письма и сами все поймете, — мягко проговорила миссис Селвик-Олдерли и вышла.

М-м-м! От досады я прикусила нижнюю губу и стала смотреть на коробку с письмами. Гладкий серый картон так и скользил под пальцами. Да, университетскому книгохранилищу до такого

далеко! К этому архиву относятся с трепетом и любовью. Неужели в моих руках разгадка тайны Розовой Гвоздики?

При любых других обстоятельствах я бы тут же разорвала шпагат. Но в гостиной миссис Селвик-Олдерли так тихо и спокойно... В камине уютно потрескивает огонь, а незнакомые лица с фотографий и миниатюр будто наблюдают за мной. Да, резкие движения здесь исключены.

Не стоит ждать слишком многого, подумала я, аккуратно развязывая шпагат. Миссис Селвик-Олдерли наверняка заблуждается или вообще выжила из ума. Ну, на сумасшедшую она, конечно, не похожа, скорее, просто смогла себя убедить, что знает, где скрывается разгадка тайны Розовой Гвоздики. Наверное, в коробке тексты песен «Битлз» или детские стихи одного из Селвиков.

Наконец мне удалось снять весь шпагат, картонная крышка упала на ковер, обнажая пожелтевшие листы бумаги. На первом же из писем стояла нацарапанная неровным почерком дата: «4 марта 1803 года».

Ничего себе детские стихи!

Вне себя от возбуждения я пролистала толстую пачку. Некоторые из писем сохранились лучше, другие хуже: кое-где чернила выцвели, а на изгибах строчки затерлись. На отдельных листах виднелись следы сургучных печатей, на иных совсем не было углов — их уничтожило время и неосторожные руки читателей.

Часть писем написана грубоватым мужским почерком, часть — изящной вязью, остальные вообще каракулями. Зато все они датировались 1803 годом.

Из моря слов я выхватывала отдельные фразы: «...чистой воды провокация», «...брат не решился бы...».

Я заставила себя вернуться на первую страницу и, устроившись на ковре перед камином, поправила юбку, глотнула остывшего чая и начала читать. Первое письмо было на довольно плохом французском, так что я разобралась без труда.

4 марта 1803 года.

Дорогая сестра!

Теперь, когда война закончилась, ты наконец сможешь вернуться в Отель де Балькур...

Willig, Lauren, «*The Secret History of the Pink Carnation*». 2005.

Перевод Аллы Ахмеровой.

Источник: www.bakanov.org (отрывок публикуется с разрешения владельца ресурса www.bakanov.org).



Producers of . . .

www.multilingual.com



The magazine
for
multilingual
communicators.

www.localizationworld.com



The conference
for
the localization
industry.

www.multilingual.com/news



The newsletter
for
staying
informed.

Special Offer:

A FREE half-year *MultiLingual* digital subscription (4 issues)
for new subscribers!

To take advantage of this special offer, register
online at www.multilingual.com/DLOGP1M