



ПЕРЕВОДЧЕСКИЕ ТЕХНОЛОГИИ ДЛЯ ЕВРОПЫ





ПЕРЕВОДЧЕСКИЕ ТЕХНОЛОГИИ ДЛЯ ЕВРОПЫ

УДК 81'25(082)
ББК 81.07я43п27
П 27

Руководитель проекта: Джанни Лаццари
Перевод с английского: Азгальдов Э.Г.
Редакторы русского издания: Е.И. Кузьмин, Е.В. Плыс

Переводческие технологии для Европы. – М.: МЦБС, 2008. – 64 с.
Издание на русском языке подготовлено по инициативе Российского комитета Программы ЮНЕСКО «Информация для всех» и Межрегионального центра библиотечного сотрудничества (МЦБС) при финансовой поддержке Федерального агентства по культуре и кинематографии и Московского бюро ЮНЕСКО.

HUMAN LANGUAGE TECHNOLOGIES FOR EUROPE

Работу заказал



Тренто (Италия)

Работу выполнил

ACCIPIO
CONSULTING

Ахен (Германия)

Данное исследование финансировалось Европейским Союзом в рамках интегрированного проекта TC-STAR (Technology and Corpora for Speech-to-Speech Translation). Руководитель проекта: Джанни Лаццари – IST-2002-FP6-506738, веб-сайт: <http://www.tc-star.org>

Апрель 2006 г.

Официальное уведомление: содержание данного доклада отражает исключительно точку зрения автора. Европейская комиссия и Консорциум TC-STAR не несут ответственности за использование любой содержащейся в нем информации.

Публикуется с разрешения автора.

ISBN 978-5-91515-014-9

© МЦБС, перевод и издание на русском языке, 2008

Предисловие

Богатство наших языков – вот что, по мнению многих, есть характерная особенность и главное достижение эволюции человека. Язык дает нам возможность передавать свои мысли, чувства, знания и выражать свою культурную самобытность. Все достижения человечества – наука и технология, философия, искусство и культура – все это стало возможным благодаря языку.

В Европе наличие разных языков – это объективная реальность. Европейский Союз считает языковое разнообразие неотъемлемой частью культурного наследия; вот почему принцип равенства языков зафиксирован в учредительных договорах Европы. С самого начала в основе его политики лежал принцип сохранения языкового разнообразия. Однако языковое разнообразие требует постоянных и значительных инвестиций. Так, европейские институты расходуют значительную долю своих текущих бюджетов на услуги письменного и устного перевода.

Для промышленности и бизнеса в Европе языковое разнообразие – это и проблема, и ценность, как было недавно отмечено Комиссией в ее первом Сообщении о новой стратегической концепции многоязычия¹. Можно с уверенностью сказать, что своевременное инвестирование в технологии многоязычной коммуникации способно обеспечить быстрый доступ к новым и растущим рынкам в любой части мира – а это чрезвычайно важный фактор успешного развития Европы на перспективу.

Способность получать и использовать информацию на разных языках жизненно необходима для отдельных граждан, власти и коммерции, и переводческие технологии способны заметно облегчить общение между людьми, органами управления и деловыми кругами. Европейский Союз совместно с государствами-членами на протяжении последних 20 лет финансировал ряд научно-исследовательских инициатив, которые способствовали накоплению знаний и ресурсов и формированию общеевропейской языковой инфраструктуры.

¹ http://europa.eu.int/comm/education/policies/lang/key/legislation_en.html

Сегодня Европа – один из самых передовых рынков для переводческих технологий и машинного перевода. Европейский Союз стремится к тому, чтобы для всех языков ЕС, а также основных коммерческих языков мира имелись все необходимые инструменты и ресурсы, с тем чтобы проложить путь к созданию в Европе всеобъемлющего многоязычного информационного общества. Через распространение многоязычных продуктов и услуг, например межъязыковых систем информационного поиска и машинного перевода, Европейская комиссия стремится достичь амбициозной цели, сформулированной как основное задание в инициативе на 2010 год, а именно: обеспечить равный доступ к информации для всех граждан Европы.

Многоязычие стало линией поведения на европейском уровне, и оно будет поощряться посредством различных мероприятий в рамках образовательных, учебных и научных программ, например по обучению языкам, исследованию языкового разнообразия, переводческих технологий и цифрового контента.

Настоящий документ – ценный отчет о современном состоянии, а также о проблемах и возможностях, открывающихся перед Европой в этой важной научной области. В не меньшей степени он служит и источником вдохновения для ученых, промышленников и политиков, и несомненно поможет сделать Европу еще более многоязыкой.



*Вивьян Рединг,
член Европейской комиссии по
вопросам информационного
общества и средств массовой
информации*



*Иен Фигель,
член Европейской комиссии
по вопросам образования,
профессиональной подготовки,
культуры и многоязычия*

Предисловие к русскому изданию

Языки представляют огромную ценность для всего человечества, так как включают в себе опыт и знания предыдущих поколений, способствуют самовыражению и общению, приобретению знаний, развитию наук. В то же время наличие большого числа языков может осложнять взаимодействие отдельных людей, научных учреждений, коммерческих структур, государств. Сразу после появления первых ЭВМ родилась идея использовать их для решения языковых проблем. Однако компьютерная технология была еще слишком слабо развита для того, чтобы создать пригодные для использования средства автоматического перевода.

Обучение компьютеров человеческой речи оказалось нелегкой задачей. Несмотря на прогресс в области компьютерной техники, машинный перевод до сих пор сильно уступает переводу, сделанному человеком. Но работа продолжается, и особенно активно она ведется в Европе, так как государства, входящие в состав Европейского Союза, бережно относятся к своему культурному и языковому наследию и не жалеют сил и средств на его сохранение и развитие.

Предполагается, что машинный перевод, хоть и не сможет сравниться по качеству, а тем более заменить перевод, сделанный вручную, тем не менее найдет довольно широкое применение – в особенности в тех сферах, где требуется быстрый и недорогой (или бесплатный) перевод текстов большого объема и при этом не предъявляются высокие требования к качеству. Кроме этого, компьютерные технологии могут сильно облегчить работу профессиональных переводчиков, обеспечивая соблюдение единообразия терминологии и повышая производительность труда.

Данная книга продолжает серию публикаций, подготовленных Российским комитетом Программы ЮНЕСКО «Информация для всех» и Межрегиональным центром библиотечного сотрудничества в рамках проекта «Развитие многоязычия в Интернете» и предназначена для широкого круга читателей.

Е.И. Кузьмин, Е.В. Плыс

Краткое резюме

Для нашей многоязыкой Европы межъязыковое общение и обмен информацией имеют принципиальное значение. Двадцать официальных европейских языков, т.е. 190 языковых пар, или 380 направлений перевода, означают, что любая межъязыковая деятельность, будь то в правительстве, в бизнесе или в обществе, требует затрат средств и труда. Если для некоторых видов взаимодействия и общения такие затраты сравнительно невелики, то для других они слишком значительны, что сводит взаимодействие на нет. В будущем ситуация резко изменится благодаря возможностям, предоставляемым системами автоматического перевода, которые не столь совершенны, как профессиональные переводчики, но зато более дешевы, работают быстрее, всегда доступны и вполне приемлемы для многих целей. Хотя следует ожидать, что эффективность традиционного перевода повысится, основное применение эти технологии найдут в сфере автоматизированного межъязыкового общения. Вначале перевод устной речи и машинный перевод будут занимать лишь узкие сегменты рынка, но затем масштабы их применения станут быстро расти, причем в основном независимо от уже существующей отрасли переводческих услуг. В качестве опорных технологий они будут стимулировать развитие торговли и экономики в Европе. Для Европы стратегически важно иметь такие переводческие технологии, которые в максимальной степени содействуют межъязыковому общению и информационному обмену.

В начале доклада показывается значение переводческих технологий, в особенности для Европы, и описывается современное положение дел. Европейская ситуация рассматривается в глобальном контексте на примере США, Индии и Восточной Азии. Анализируется современное состояние в науке и коммерции и намечаются рыночные перспективы. Беседы с политиками и специалистами из сферы науки и бизнеса расширяют представление о предмете и помогают лучше его понять.

Оглавление

1. Что такое переводческие технологии?	9
2. Значение переводческих технологий для Европы	10
2.1. Критический барьер на пути европейского внутреннего рынка	10
2.2. Международные языки	11
2.2.1. Главные международные языки	11
2.2.2. Языки, находящиеся под угрозой исчезновения	13
2.2.3. Размер имеет значение: основной, второй и третий языки и рыночные силы	14
2.3. Подробнее о Европейском Союзе	14
<i>Интервью с Карлом-Йоханом Лёнротом</i>	16
2.4. Общение	18
2.5. Следующий шаг в индустриализации: машины, обрабатывающие письменное или устное слово	19
2.6. Рынок	20
2.6.1. Перевод как фактор стоимости	20
2.6.2. Сложившиеся рынки локализации и перевода	20
<i>Интервью с Ренато Бенинатто</i>	23
2.6.3. От ручного перевода к машинному: резкое снижение затрат и улучшение доступности	25
<i>Интервью с Майклом Анобиле</i>	26
2.6.4. Поддерживающие и разрушительные технологические инновации	28
3. Азия, Европа и США: черты сходства и различия	30
3.1. Европейский Союз	30
3.2. Соединенные Штаты Америки	31
3.2.1. Стратегическая роль ПТ	31
3.2.2. Исследовательские программы	31
<i>Интервью с Джозефом Оливом</i>	33
3.3. Восточная Азия	35
3.3.1. Английский как общепринятый язык в Восточной Азии	35
3.3.2. Азиатские языковые пары и растущее значение китайского языка	35
3.3.3. Исследовательские программы	36
<i>Интервью с Юн-ити Цудзии</i>	37

3.4. Индия	39
3.5. Экономические граничные условия	39
3.6. Место действия для Европы	41
<i>Интервью с Жозефом Марьяни</i>	42
4. Современное положение дел	44
4.1. Переводческая деятельность сегодня	44
4.2. Технологии, применяемые в профессиональном переводе	44
<i>Интервью с Кевином Боленом</i>	46
4.3. Исследования в области перевода «речь в речь» и необходимых технологий	49
4.4. Проект TC-STAR	51
5. Сила опорной технологии	53
5.1. Нескончаемые потребности человека	53
<i>Интервью с Димитрисом Сабатакакисом</i>	56
6. Заключение	58
7. Используемые материалы и дополнительная информация	59
7.1. Библиография	59
7.2. Для дальнейшего ознакомления	60
7.3. Таблицы	60
7.4. Рисунки	61
7.5. Список сокращений	62

1. Что такое переводческие технологии?

Мы уже давно поняли, что технология способна коренным образом изменять окружающий мир и весь наш образ жизни. Оглядываясь в прошлое, мы видим, что мир кардинально менялся с появлением компьютеров, технологии оцифровки и компьютерных сетей. Впрочем, каждая из этих революций происходила постепенно, на протяжении ряда лет, и потому все эти необычайные изменения не имели шокового характера.

Сегодня нам хорошо известно об еще одной компьютерной технологии с колоссальным потенциалом, и можно утверждать, что это еще одна революция. Эта технология развивается довольно медленно, и многие научные темы разрабатываются уже столько лет, что некоторые потеряли всякую надежду. Необычайно трудно научить компьютер обрабатывать человеческую речь и язык (и в устной, и в письменной форме) так, как это умеет делать человек, а именно: говорить естественно, понимать сказанное (и подразумеваемое), обобщать содержание документа или разговора, находить звукозапись по содержанию, переводить с одного языка на другой. Мы бы хотели иметь возможность общаться с машинами с помощью голоса и языка, потому что это наши средства общения и мы желаем, чтобы компьютеры обрабатывали информацию именно в этой форме и так, как удобно нам.

Совокупность технологий, позволяющих осуществить все это, называ-

ется *переводческими технологиями* (ПТ). Самые заметные из них – это автоматическое распознавание речи, машинный перевод и преобразование текста в речь, но есть еще целый ряд других. Как и в случае предшествующих достижений в области вычислительной техники, построения сетей и оцифровывания, ПТ способны коренным образом изменить наше представление об информации и наши методы работы с ней, потому что мы сможем извлекать и обрабатывать закодированную в языке информацию совершенно по-новому. В центре внимания данного доклада лишь один аспект ПТ, а именно: способность этих технологий преодолевать языковой барьер, обеспечивать общение между людьми или обрабатывать неструктурированную информацию в виде текста на естественном языке. На основе базовых технологий *машинного и устного перевода*¹ формируется важный комплекс приложений, в который входят обобщение межъязыкового информационного поиска и усвоение данных. Еще один важный комплекс – межъязыковое общение, т.е. машинный перевод письменных текстов или перевод устной речи, иначе – *перевод из речи в речь* (устный перевод с устным или письменным выводом). Это настолько ново, что даже еще не придуман подходящий термин. Может быть, подходящим термином будет *технология межъязыковой информации и коммуникации*?

¹ Подробнее об этих технологиях см. в главе 4.3.

2. Значение переводческих технологий для Европы

2.1. Критический барьер на пути европейского внутреннего рынка

В Договоре о создании Европейского сообщества закреплены четыре фундаментальные свободы²: *свобода движения товаров, людей, услуг и капитала*. Внутренний рынок – главное достижение европейской интеграции – был реализован в конце 1992 года в виде зоны без внутренних границ для товаров и услуг. Торговля внутри Европейского Союза не облагается налогом, и национальные системы налогообложения обязаны уважать эти четыре фундаментальные свободы. Для обеспечения функционирования внутреннего рынка многое сделано в плане согласования нормативных актов и требований национального законодательства. Свободное движение товаров на едином рынке гарантируется *принципом взаимного признания*, который делает ненужным трудоемкий процесс полного согласования национальных законодательств государств-членов: товары, законно производимые в одной стране-члене Европейского Союза, не могут запрещаться к продаже на территории любой другой страны-члена, даже если они произведены согласно иным техническим условиям или стандартам качества, чем те, которым подчиняется собственная продукция этой

страны³. Тот же принцип распространяется на услуги.

Языковой барьер – последнее торговое препятствие на пути информационных услуг в Европе.

С введением общих налогов и сертификатов соответствия в принципе исчезли два существенных торговых барьера. С логистической точки зрения, распределение товаров все еще остается проблемой для вещественных изделий. Что касается информации, то с появлением Интернета произошел «квантовый скачок»: распространение информации настолько ускорилося и подешевело, что так и хочется сказать, что теперь ее можно получать немедленно и бесплатно. Если эти три барьера в основном преодолены, то остается еще четвертый барьер на пути торговли между странами, а именно: языковой барьер⁴. Межъязыковая и межкультурная коммуникация стали жизненно важными для торговли, особенно сегодня, с развитием глобализации экономики через Интернет. Затраты труда на представление товара на языке конечного потребителя колеблются в широких пределах и зависят от типа товара. Как правило, объем этой работы бывает велик в случае информационных услуг, что выражается в высоких издержках производства. Если представить это в виде одной таблицы,

² Статья 14 Договора.

³ Единственное допустимое исключение – доминирующий общий интерес, например, охрана здоровья, потребителя или окружающей среды – применяется в строго ограниченных условиях.

⁴ Это упрощение не учитывает, что помимо проблемы языка есть и другие, например, культурные различия, которые необходимо учитывать при выпуске товара или услуги на местный рынок. Впрочем, в большинстве случаев это мелочь по сравнению с затратами труда на перевод.

картина будет иметь примерно следующий вид:

	Товары	Информация
Налоги	☺	☺
Соответствие	☺	☺
Распределение	☺	☺
Язык	☺	☹

Таблица 1. Торговая конъюнктура для международной торговли внутри ЕС в отношении реальных товаров и информации (информационных услуг). Для информационных услуг языковой барьер является принципиальным барьером, который необходимо преодолеть.

Из четырех указанных барьеров самым существенным препятствием для предоставления информационного обслуживания между разными странами остается язык. Применение переводческих технологий, которые являются темой настоящего доклада, позволит устранить это серьезное препятствие – последнее значительное различие между нашим общим европейским рынком и крупным внутренним рынком, как, например, в США, что принесет существенную экономическую выгоду.

2.2. Международные языки

Сейчас в мире известно 6912 живых языков⁵. Многие из них находятся в Азиатско-Тихоокеанском регионе и в Африке (см. рис. 2).

2.2.1. Главные международные языки

Английский язык, по-видимому, является самым важным языком, но с точки зрения количества говорящих

Ранг	Носители языка (млн чел.)		
	Родной язык	Источ. А	Источ. В
1	Китайский	1,113	1,123
2	Английский	372	322
3	Хинди/Урду	316	236
4	Испанский	304	266
5	Арабский	201	202
6	Португальский	165	170
7	Русский	155	288
8	Бенгальский	125	189
9	Японский	123	125
10	Немецкий	102	98
11	Французский	70	72
12	Итальянский	57	63
13	Малайский	47	47

Таблица 2. Количество говорящих на основных международных языках как на родном (в миллионах человек) согласно двум источникам: (А) модель «engco» (English Company) [Gra] и (В) сравнительные цифры из «Этнолога» ([Gr1]; см. [Gra]).

на языке как на родном он уступает китайскому. Однако, если суммировать количество людей, для которых английский является родным, и для которых он является вторым языком (по 375 миллионов человек), и прибавить к ним людей, которые говорят на английском языке как на иностранном (750 миллионов), получаем внушительную цифру в 1,5 миллиарда человек⁶, говорящих по-английски.

В *Таблице 2* приведен полный список самых значительных языков с точки зрения количества говорящих на них как на родном. Эти данные взяты из литературы. Любопытно, что вместо

⁵ Источник: [Gor]. – Следует по крайней мере отметить, что существует такая проблема, как идентификация языка. Живой язык – это язык, являющийся родным хотя бы для одного человека.

⁶ [Cru] и другие источники. – Нужно заметить, однако, что цифры, приводимые в литературе, противоречивы и в значительной мере зависят от предполагаемого уровня владения иностранным языком.

	Язык	Уровень влияния
1	Английский	100
2	Немецкий	42
3	Французский	33
4	Японский	32
5	Испанский	31
6	Китайский	22
7	Арабский	8
8	Португальский	5
9	Малайский	4
10	Русский	3
11	Хинди/Урду	0,4
12	Бенгальский	0,09

Таблица 3. Уровень «мирового влияния» 12 главнейших языков согласно модели «engco» (см. таблицу 2). 100-балльный показатель соответствует позиции английского языка на 1995 год [Gra].

ожидаемого рангового списка из 10 или 20 языков в таблице указаны первые 13 языков⁷, по-видимому, для того, чтобы в списке оказались французский

и итальянский языки. Даже в сухой статистике, если речь идет о языке, непременно всплывают национальные чувства, культурные разногласия и эгоцентрический (или евроцентрический) взгляд. Язык тесно связан с культурой и нашим чувством принадлежности к какому-либо обществу.

Совершенно ясно, что количество носителей языка не соответствует нашему интуитивному ощущению его значимости. Например, французский считается значительным языком, и в европейских школах его преподавание стоит на втором месте после английского языка. Восприятие языка как значительного зависит и от других факторов, например от его экономической и политической важности (помимо фактора культурного наследия). Таблица 3 характеризует мировое значение некоторых языков по сравнению с английским. Интересно отметить, что значение

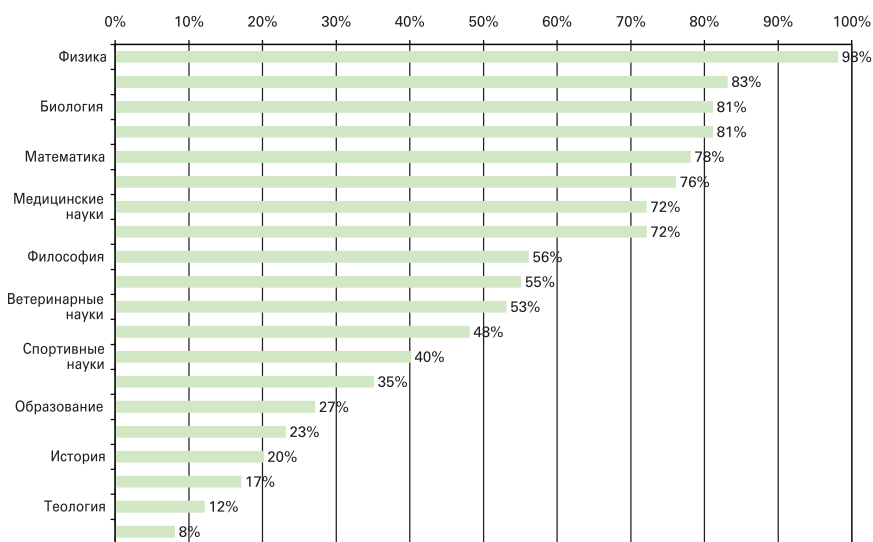


Таблица 4. Отрасли знания, в которых немецкие ученые считают английский своим рабочим языком [Gra].

⁷ Таблица взята из [Gra], где показаны первые 13 языков.

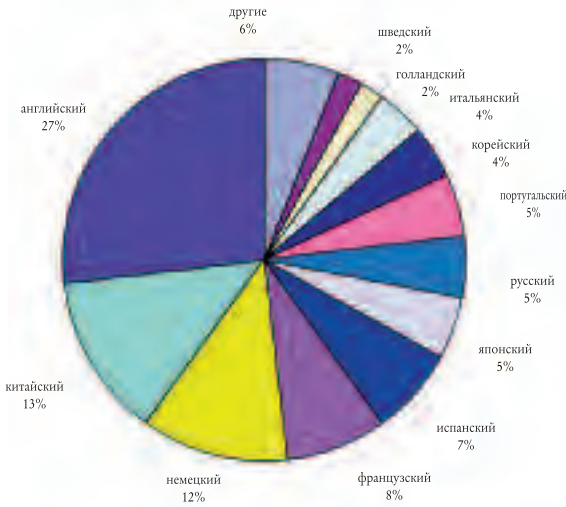


Рис. 1. Доля мировой книжной продукции на каждом языке. Английский – самый распространенный иностранный язык в книгоиздании; книги на английском языке выходят более чем в 60 странах мира [Gra].

того или другого языка меняется в зависимости от отрасли знания или сферы применения. В *таблице 4* дается конкретный пример.

2.2.2. Языки, находящиеся под угрозой исчезновения

Угроза языку проявляется в двух аспектах: количество говорящих на этом языке и количество функций, для которых он употребляется. Люди двуязычные, как правило, начинают общаться со своими детьми лишь на втором языке или все реже используют свой первый, родной язык. В «Этнологе»⁸ около 500 языков числятся почти исчезнувшими⁹. Угроза языкам вызывает беспокойство потому, что язык тесно связан с культурой, причем эта связь такова, что утрата языка почти всегда сопровождается социальной и культурной дестабилизацией.

Угроза языкам вызывает беспокойство потому, что язык тесно связан с культурой, причем эта связь такова, что утрата языка почти всегда сопровождается социальной и культурной дестабилизацией.



Рис. 2. Языки мира. Каждая точка соответствует первичному местонахождению живого языка, учтенного в «Этнологе».

⁸ «Этнолог» (GOR) – каталог, насчитывающий более 6700 языков, на которых говорят в 228 странах.

⁹ Почти исчезнувший язык: относится к языкам, на которых говорит меньше 50 человек или очень небольшая часть этнической группы.

(К удивлению автора данного доклада, еще одной проблемой является урон, который несет научное сообщество, изучающее такие языки!)

2.2.3. Размер имеет значение: основной, второй и третий языки и рыночные силы

Язык может находиться в невыгодном положении, даже без угрозы исчезновения. Возьмем какую-либо компанию, которая собирается распространить свой бизнес с местного рынка на международный. Поскольку стоимость локализации или адаптации к новому языку не коррелирует с численностью говорящих на нем, а является более или менее постоянной, можно говорить об основных языках (т.е. таких, на которых товары/услуги обязательно должны быть представлены) и вторых, а то и третьих языках, которые имеют второстепенное коммерческое значение. Хотя что считать вторым или третьим языком, зависит от обстоятельств, в целом можно сказать, что рыночные силы постоянно «карают» определенные языки, – как правило, те, на которых говорит незначительное число людей, или те, которые используются в экономически слабых странах.

Сходная мотивировка действует даже в случае, когда человек берется за изучение иностранного языка. Хочет ли немец скорее выучить голландский язык, который близок к немецкому и потому требует сравнительно

мало усилий, чем испанский, учитывая, что испаноязычный мир чрезвычайно велик по сравнению с миром, говорящим на голландском языке? (См. типичные цифры в *таблице 2.*) И зачем вообще учить голландский, если голландцы в большинстве своем все равно говорят по-английски?¹⁰ Выбор зависит от многих факторов. Говорят ли на этом языке в соседней или в очень далекой стране? Есть ли в этой стране второй язык, которым можно было бы пользоваться вместо основного?

Многие языки испытывают давление со стороны рыночных сил.

Язык тем привлекательнее, чем больше вероятность того, что придется говорить на нем. Понятно, что языки, на которых говорит мало людей, оказываются опять же в невыгодном положении.

2.3. Подробнее о Европейском Союзе

Язык тесно связан с культурой. Во многих случаях он является жизненно важной составляющей национальной самобытности. Неудивительно, что мы, европейцы, создавая Европейский Союз, приняли сознательное решение не вводить основной язык, а сохранить различные языки и предоставить им равные права. Хотя не следует недооценивать большое значение английского языка как *общепринятого языка* в сегодняшнем мире, все же мы

¹⁰ 91% голландского населения способен поддерживать разговор минимум еще на одном языке [ЕВ5]. Уместно заметить, однако, что недавно в Нидерландах был принят закон, по которому обязательным условием получения гражданства в стране является знание голландского языка. Это говорит о важности языка для культуры и самосознания.

Чешский	CS	eština
Датский	DA	Dansk
Голландский	NL	Nederlands
Английский	EN	English
Эстонский	ET	Eesti
Финский	FI	Suomi
Французский	FR	Français
Немецкий	DE	Deutsch
Греческий	EL	Elinika
Венгерский	HU	Magyar
Итальянский	IT	Italiano
Латвийский	LV	Latviesu valoda
Литовский	LT	Lietuviu kalba
Мальтийский	MT	Malti
Польский	PL	Polski
Португальский	PT	Português
Словацкий	SK	Slovenčina
Словенский	SL	Slovenščina
Испанский	ES	Español
Шведский	SV	Svenska

Таблица 5. Двадцать официальных языков Европейского Союза и их сокращенные обозначения [ELP]. Эрский (ирландский) язык станет 21-м официальным языком ЕС с 1 января 2007 г.

живем в многоязычном мире и язык имеет значение. Европейский Союз был задуман многоязычным, и в нем даже имеется должность Комиссара по вопросам образования, профессиональной подготовки, культуры и многоязычия.

Считается, что разные языки имеют равное значение, и определенные документы доступны (т.е. переведены) на всех языках, в частности законы и парламентские дебаты. Однако в целях экономии и ускорения работы установлены три рабочих языка, которые используются во внутренней деятельности Европейского Союза, а именно: английский, французский и немецкий.

Обобщая сказанное, у нас есть веские причины защищать свое культурное наследие, но сохранение многочисленных языков обходится дорого¹¹ как в экономическом смысле (стоимость перевода для того или иного товара), так и в смысле других затрат (затраты труда на изучение дополнительного языка). Любая технология, помогающая снизить эти издержки, способствует сохранению нашего культурного наследия.

¹¹ Само собой разумеется, что вариант отказа от перевода даже не рассматривается: гражданам необходимо понимать законы, а товары приходится адаптировать к разным языкам, чтобы они продавались. Наш довод заключается в том, что любые затраты, пусть и небольшие, это препятствие, и любое существенное сокращение затрат и улучшение доступности будет иметь положительный эффект.

Интервью с Карлом-Йоханом Лёнротом, генеральным директором Главного управления по вопросам перевода (ГУП)



Карл-Йохан Лёнрот, генеральный директор Главного управления по вопросам перевода Европейского Союза, Люксембург и Брюссель (Бельгия)

Карл-Йохан Лёнрот с января 2004 г. занимает пост генерального директора Главного управления по вопросам перевода. Предыдущие должности:

2000–2003 гг. – заместитель генерального директора Главного управления по вопросам занятости и социальным вопросам Европейской комиссии.

1996–2000 гг. – директор по стратегии занятости и Европейскому социальному фонду Главного управления по вопросам занятости и социальным вопросам Европейской комиссии.

1991–1996 гг. – директор Отдела занятости Международной организации труда, Женева.

1971–1991 гг. – различные посты в Министерстве труда Финляндии: научный сотрудник, зав. отделом планирования, зам. директора по услугам рынка труда, специальный советник.

1973–1977 гг. – Секретариат Скандинавского Совета министров, куратор вопросов занятости и миграции и трехстороннего сотрудничества. Консультирование ОЭСР, Скандинавского Совета министров и Посольства Финляндии в Стокгольме.

Образование: магистр политических наук (Университет Хельсинки, 1970 г.), магистр гуманитарных наук (Университет Висконсина, США, 1972 г.), Национальная школа администрации (ENA, 1983 г.).

Знание языков: финский, шведский, английский, французский и немецкий, начальное знание русского и испанского.

Г-н Лёнрот имеет богатый опыт международного сотрудничества (страны Северной Европы, ОЭСР, ООН, МОТ) и двустороннего сотрудничества (например, в Восточной Европе). Является автором свыше 40 статей, публикаций и т.п. по вопросам труда, занятости, миграции, социальных, политических и общественных проблем.

ГУП – довольно большая организация, не так ли?

Я бы сказал, что самая большая в мире. Мы работаем с 21 языком. Общая цифра, характеризующая переводческое обслуживание Европейского Союза, – это 1,1 млрд евро в год, или 1% всего бюджета Европейского Союза. Сюда входят устный и письменный перевод, причем обслуживается не только Европейская комиссия, но все другие учреждения, например, Европейский парламент. Звучит довольно внушительно, но в целом на одного гражданина приходится 2,55 евро.

Не является ли все это множество языков тяжким бременем для общества и экономики Европы?

Вопрос заключает в себе оценочное суждение. Предполагается, что мы говорим о бремени и издержках, а не о преимуществе. На самом деле многоязычие – это составная часть европейской социальной модели; мы обладаем культурным

разнообразием. Следует считать это богатством и динамическим элементом, а не помехой, и я бы сказал, что наличие нескольких языков позволяет открывать новые рынки. Политика многоязычия благотворна для нашего динамичного общества. Мы в Европе уважаем основные свободы и культурное разнообразие, и через это разнообразие люди воспринимают новые идеи и новые стимулы для развития экономики. Это бремя для тех, кто желал бы иметь свободу передвижения на едином рынке труда, но владеет всего одним языком. В этом одна из причин, почему ЕС старается поощрять знание языков и обучение языкам.

Каким образом ГУП решает проблемы, связанные с расширением ЕС?

За последний год мы приняли на работу более 500 новых сотрудников. Мне кажется, мы воспринимаем новый вызов очень хорошо. Трудность состоит в том, что нам нужно создавать новые термины и повышать качество переводов, потому что в новых языках отсутствуют некоторые термины европейской политики. В этом и заключается основная проблема. Мы в большинстве случаев переводим правообразующие документы с нуля, впервые, и иногда они оказываются довольно сложными для перевода. Интеграция этих новых языков проходит достаточно успешно.

Как, по вашему мнению, будет развиваться машинный перевод?

Машинный перевод как инструмент перевода и переводческая память находится пока еще в процессе развития, но, на мой взгляд, у него блестящее будущее. Это, впрочем, не означает уменьшения роли человеческого фактора. Развитие идет скорее в направлении некой комбинации, которую я назвал бы интеллектуальным переводом: переводчик обеспечивает качество, связность и правильную терминологию, а машинный перевод помогает понять смысл и повысить производительность.

Какой вы видите профессию переводчика?

Перевод как таковой – самая быстро растущая отрасль в мире. Глобализация и существование множества языков увеличивают потребность в многоязычном обслуживании, а это значит также, что данная профессия требует большего признания, ибо она очень важна. В условиях всех этих серьезных вызовов она тоже изменится. Переводчик больше не является простым преобразователем текста, становясь скорее лингвистом-консультантом, редактором. Так что сфера деятельности этой профессии будет все больше расширяться.

Мы работаем совместно с университетами и их объединениями в направлении создания учебной программы и курса обучения. Мы пытаемся разработать специальность, которую мы называем европейским магистром перевода – что-то вроде стандартизированной квалификации, которая будет признаваться всеми; это к тому же облегчит подбор кадров. Вот что я имею в виду, говоря об улучшении профессионализации в данной области.

2.4. Общение

Многие европейцы говорят на двух или нескольких языках. Но при этом около половины граждан Европейского Союза не говорит ни на каком другом языке, кроме собственного.

Разве не замечательно было бы, если бы европейцы, говорящие на разных языках, могли свободно беседовать друг с другом? Любая форма общения, пусть и на сниженном уровне, была бы достижением по сравнению с современным положением. Перевод не обязательно был бы идеальным или грамматически правильным. Такое требование весьма отлично от тех требований, которые предъявляются сегодня к профессиональному переводу. Как мы увидим позже, это имеет определяющее влияние на рыночные силы, и особенно, – на перспективы новых игроков на рынке.

Можно ли представить сегодня, что будет значить для нас беспрепятственный и недорогой перевод с естественного языка? По опыту мы знаем, сколь много было сделано с помощью технологии для улучшения общения между людьми,

для преодоления пространства и времени. Одним из первых крупных шагов в развитии человеческой культуры было создание письменности, которая позволяла передавать знания будущим поколениям. Преодоление пространства началось очень давно. Примером может служить передача информации с помощью оптических сигналов вдоль границ Римской или Китайской империй. Другие гигантские шаги: учреждение почтовой службы, появление телефона, трансатлантической линии связи, сотового телефона и Интернета.

Теперь нам предстоит сделать еще один большой шаг, а именно: внедрить мгновенный и дешевый перевод, роль которого – не преодолевать время и пространство, а служить мостом между культурами. Будет ли этот шаг спустя два десятилетия считаться столь же значимым, что и прочие? Нам трудно представить себе будущее, когда станет возможным мгновенный перевод речи и письменного текста по умеренной стоимости, потому что сегодня такой способ коммуникации попросту отсутствует. Все же вполне вероятно, что повсеместное распро-

Язык	Количество людей, говорящих на нем как на родном языке	Количество людей, говорящих на нем, не как на родном языке	Общее количество говорящих на языке
Английский	13%	34%	47%
Немецкий	18%	12%	30%
Французский	12%	11%	23%
Итальянский	13%	2%	15%
Испанский	9%	5%	14%
Польский	9%	1%	10%
Голландский	5%	1%	6%
Русский	1%	5%	6%

Таблица 6. Владение иностранными языками в ЕС: доля граждан ЕС, говорящих на соответствующем языке в ЕС (как на родном / как на втором или иностранном / на обоих). Обследование было проведено в 25 странах-членах ЕС, а также в странах, готовящихся к вступлению в ЕС (Болгарии и Румынии), странах-кандидатах (Хорватии и Турции) и среди турок-киприотов. – Источник: [EB5]

странение такого инструмента окажет влияние, сходное с тем, какое оказали телефон или Интернет.

Раньше преодолевались пространство и время; следующий большой шаг – связывать языки и культуры.

2.5. Следующий шаг в индустриализации: машины, обрабатывающие письменное или устное слово

Как было указано, повсеместное распространение системы недорогого перевода приведет к коренным переменам. В результате революционного скачка процесс, ранее выполнявшийся исключительно человеком, вполне успешно сможет выполняться машиной.

В 1960-х годах информационные технологии, вначале применявшиеся в правительственной, научной и военной сферах, стали проникать в коммерческую, вначале в банковское и страховое дело, а позже – практически во все сферы жизни людей. Машины научились обрабатывать информацию. Значительная часть заданий, прежде требовавших квалифицированных работников, теперь выполнялась машинами, притом быстрее, дешевле и эффективнее. Но в то время были возможны (или по крайней мере достаточно эффективны) лишь определенные виды обработки информации. Обработка лингвистической информации, не считая таких структурированных видов данных, как имена и адреса, оставалась по существу недостижимой. Если производство письменной информации, например газет, книг и вообще печатных материалов, в значительной мере обеспечивалось информационными технологиями, то применение этих технологий

для обработки входящей информации было весьма ограниченным.

С появлением Интернета произошли большие перемены. Невероятно выросла потребность в обработке языковой информации в разных аспектах: поиск, обобщение, перевод и систематизация. Широкое распространение самообслуживания клиентов в Интернете привело к громадному росту контактов пользователей с компаниями и правительственными органами. Одновременно возросли и требования клиентов, большинство из которых теперь считает само собой разумеющимся, что ответ на их письмо, отправленное по электронной почте, будет получен в течение суток. Поиск документов по словам (главным образом через Google) сейчас практикуется чуть ли не в каждом доме. Однако выявились два существенных узких места: во-первых, поиск по словам – всего лишь суррогат поиска информации, и желательно было бы иметь *семантическую сеть*, дающую возможность вести поиск по содержанию, а не по словам. Во-вторых, Интернет входит в жизнь все более широких кругов населения, что увеличивает потребность в локализации, особенно в отношении языков. Это создаст колоссальный спрос на масштабный и (или) оперативный перевод, в частности для создания многоязычных веб-сайтов, которые стали важнейшим стимулятором развития переводческой отрасли во всем мире, а также для межязыкового поиска документов; последнее направление является предметом значительных исследований, проводимых в таких компаниях, как IBM и SAP, Google и Yahoo.

В более широком плане развитие переводческой отрасли стимулируется глобализацией и Интернетом.

2.6. Рынок

Что касается экономической стороны перевода, то анализ существующего рынка переводов оказывается недостаточным. Нужно учитывать и другие важные аспекты. В Европе стоимость перевода и влияние этого фактора на европейскую экономику имеет первостепенное значение в том смысле, что помимо прямых затрат на перевод само существование языковых барьеров может мешать использованию открывающихся возможностей.

2.6.1. Перевод как фактор стоимости

Хотя мы дорожим культурным богатством Европы, отраженным в ее многочисленных языках, языковые барьеры – это обуза для нашей экономики, потому что они добавляют затраты на перевод к любой деятельности, преодолевающей языковые границы. Во многих типичных случаях эти затраты составляют лишь небольшую долю издержек производства: что-то в пределах от 0,25% до максимум 2%. Некоторые из этих издержек являются прямыми (хотя определить их зачастую бывает непросто), другие же – косвенными в том смысле, что они препятствуют образованию стоимости.

Трудно даже дать надежную оценку прямых издержек. К примеру, в крупной корпорации может быть предусмотрен специальный бюджет на локализацию веб-сайта. При этом вся мелкая переводческая деятельность, ведущаяся во множестве разных отделов, в этой цифре обычно не учитывается, а ведь она может составлять даже более крупную сумму.

С другой стороны, если считать по зарплатам переводчиков, это тоже не решение, потому что многие из них работают по совместительству или как внештатники. Ситуация, когда организация платит локализирующей фирме, которая в свою очередь платит субподрядчику, а тот – внештатнику, может затруднить оценку денежных затрат. Нужно ли учитывать денежные средства на всех этапах или только на первом? Таким образом, имеющиеся цифры варьируются в зависимости от используемой методики¹². Но можно с уверенностью сказать, что мировой переводческий рынок оценивается в пределах от 8 до 30 миллиардов евро. Вот две цифры, полученные от ЕС, которые характеризуют правительственные расходы:

- Ежегодно Европейский парламент тратит 300 миллионов евро, или 30% своего бюджета, на перевод всех парламентских дебатов и документов ЕС на 20 официальных европейских языков.
- Европейский Союз тратит 1,1 миллиарда евро в год, или 1% своего бюджета, на все услуги письменного и устного перевода.

ЕС тратит 1,1 млрд евро в год на услуги письменного и устного перевода.

2.6.2. Сложившиеся рынки локализации и перевода

Хотя люди имеют примерное представление о работе переводчиков, переводе книг или документов, в целом они не так много знают об отрасли, осуществляющей локализацию программных продуктов и ин-

¹² Подробнее об этом читайте в интервью с Ренато Бенинатто.

тернет-страниц, которые составляют значительную часть переводческой индустрии. Важно осознать, что их клиенты – компании и учреждения – на самом деле нуждаются не просто в переводе, а в «локализации».

Помимо переводческого аспекта, локализация программных продуктов также обеспечивает правильную работу программ (причем это относится и к онлайн-помощи и к документа-

ции), а также обеспечивает правильное форматирование письменных документов. Приходится учитывать различия в объеме переведенных текстов и по количеству слов, и по количеству знаков, а также разные направления письма. Кроме письменности слева направо и справа налево, существуют языки с двумя направлениями: если в тексте на арабском языке встречаются слова на латинице, приходится поддер-

Ранг	Компания	Местонахождение штаб-квартиры	Доход, млн долл.	Штат	Офисы	Статус
1	Lionbridge Technologies	США	377,1	4,000	50	Госуд.
2	Titan Corp.	США	285,4	n/a	n/a	Госуд.
3	SDL International	Великобритания	146,0	1,400	36	Госуд.
4	STAR AG	Швейцария	96,0	750	33	Частн.
5	RWS Group	Великобритания	63,4	350	7	Госуд.
6	SDI Media Group	США	60,3	200	20	Частн.
7	Xerox Global Services	Великобритания	60,0	200	4	Госуд.
8	Euroscript S.à.r.l.	Люксембург	54,5	600	9	Частн.
9	Transperfect/Translations	США	50,2	325	29	Частн.
10	CLS Communication	Швейцария	36,0	260	11	Частн.
11	Logos Group	Италия	36,0	150	17	Частн.
12	LCJ EEIG	Германия/Италия/ Бельгия/Испания	21,6	140	9	Частн.
13	Thebigword	Великобритания	20,0	122	7	Частн.
14	Hewlett-Packard ACG	Франция	20,0	65	6	Госуд.
15	Moravia	Чехия	19,0	350	11	Частн.
16	TOIN	Япония	19,0	105	5	Частн.
17	Merrill Brink International	США	18,5	120	4	Частн.
18	VistaTEC	Ирландия	18,2	123	3	Частн.
19	Transware	Ирландия	18,0	160	8	Частн.
20	McNeil Multilingual	США	17,2	105	9	Частн.

Таблица 7. Классификация 20 ведущих поставщиков переводческих услуг по прибыли за 2004 г. (В цифрах учтены два важных слияния 2005 г.: поглощение компании Bowne Global Solutions компанией Lionbridge и поглощение компании TRADOS компанией SDL.)

живать оба направления! Существует множество наборов знаков, для которых требуется поддержка, а для языков с наборами двухбайтовых символов (японского, китайского и корейского) нужна специальная программная поддержка. (Система Unicode позволяет решать эти проблемы, но может оказаться непригодной в тех случаях, когда речь идет о слишком большом объеме унаследованных материалов.) Вообще говоря, программные продукты, предназначенные для локализации, должны с самого начала разрабатываться соответствующим образом, с тем чтобы позже избежать лишних затрат. И последнее (но не по значению) – локализация затрагивает и различия между культурами. Нужно обладать творческим началом и умением поставить себя на место другого, чтобы передать понятие таким образом, чтобы оно прижилось в другом языке или в другой культуре.

Обычный способ локализации продукта – разделение его на текстовые сегменты, а также на элементы пользовательского интерфейса. Затем подлежащий переводу текст переводится, а пользовательский интерфейс и документация соответствующим образом модифицируются, чтобы нормально функционировать в выходном языке.

Основные сегменты рынка локализации – это локализация программного обеспечения и веб-сайтов.

По данным одного источника¹³, переводческая индустрия в 2005 году создала доход в размере 8,8 миллиардов долларов США. Сюда входит перевод с естественных языков, а также перевод с использованием различных технических средств и т. п. Рост этого рынка определяют сегодня два значительных сегмента: работа с многоязычными веб-сайтами и локализация программного обеспечения. Считая только компании, имеющие не меньше пяти сотрудников, на этом мировом рынке действуют 5 тысяч компаний.

Важнейшие технологии, используемые сегодня при переводе, это «переводческая память» (ПП), терминологические базы данных, а также программное обеспечение и инструменты для работы с многоязычными веб-сайтами и локализации программного обеспечения. Переводческая память – это широко известная (хотя и не очень широко используемая) вспомогательная технология, которая одновременно помогает снизить затраты и повысить качество, особенно в плане единообразия переводов, что очень важно, например, в технической и юридической сферах. Точно так же системы машинного перевода или перевода устной речи, используемые переводчиками как вспомогательные средства, по видимому, будут иметь благоприятное влияние и на стоимость, и на качество ручного перевода. Как указывается ниже в главе 2.6.4, где также приводятся терминологические дефиниции, в этом сценарии машинный перевод выступает как поддерживающая инновация.

¹³ Common Sense Advisory [CSA]. В других источниках приводятся гораздо более высокие глобальные оценки – порядка 30 млрд евро. Для данного доклада важны не точные цифры, а сам факт, что переводческая индустрия – широко известная и растущая отрасль, намного меньшая по размеру, чем другие отрасли, но дающая возможность осуществлять коммерческую деятельность гораздо более значительного масштаба.

Интервью с Ренато Бенинатто, исполнительным директором компании Common Sense Advisory

У Ренато за плечами 20-летний опыт работы на руководящих постах в индустрии локализации. Он работал в высшем исполнительном руководстве ряда крупнейших компаний отрасли, в последние годы – вице-президентом компании Alpnet Inc. и директором компании Berlitz Global-NET. Он делает акцент на стратегиях, которые стимулируют рост в мировом масштабе. Его специальность – обеспечивать успешную работу компаний на мировых рынках и основывать компании, выходящие за государственные границы.

В настоящее время он является партнером и ведущим аналитиком компании Common Sense Advisory, Inc., компании по маркетинговому исследованию рынка и консультированию, которая специализируется на индустрии переводов и локализации и имеет клиентов на всех континентах. Ренато курирует вопросы, связанные с предложением услуг и системой показателей компании.



Ренато С. Бенинатто,
исполнительный директор
и вице-президент по
консультативной практике
компании Common Sense
Advisory, Inc., Бостон, США

Я видел цифры, характеризующие объем мирового переводческого рынка, которые не создают единой картины. Почему так трудно получить эти цифры?

Оценка объема рынка – дело сложное, требующее в равной мере участия логики и скептицизма. Конечная цель – добиться достоверного приближения к реальному рынку.

Фактором, побудившим компанию Common Sense Advisory измерить объем переводческой индустрии, был отчасти мой прошлый опыт продажи переводческих услуг в двух переводческих компаниях открытого типа, а также личных контактов с крупнейшими покупателями переводов на рынке. Когда говоришь о рынке переводов с посторонними людьми, они неизменно видят громадные возможности, но также и разрыв между тем, что публикуется на всех языках, и тем, что выходит на их языке. Но когда ты на улице и продаешь эти услуги, действительность оказывается гораздо более суровой. Перевод – не стратегическая задача и потому не закладывается в бюджет. Мы придумали образное выражение – «соотношение перевода и туалетной бумаги», – которое характеризует долю затрат на перевод в бюджетах компаний.

Основной недостаток методов оценки объема рынка, проводимых людьми со стороны, заключается в том, что они исходят из информации, предоставляемой участниками рынка. Одна из основных особенностей рынка переводов – выполнение работ по субподрядам до четырех уровней. Например, поставщик многоязычных продуктов передает восточноевропейские языки на внешний подряд какой-то компании в Венгрии, которая в свою очередь передает их другим компаниям в Чешской Республике, Польше и Болгарии, которые в свою очередь передают их внештатному переводчику. У нас, в компании Common

Sense Advisory, мы считаем только первый субподряд, который отражает действительную сумму денег, заплаченную за работу конечным пользователем.

Какой методикой вы пользуетесь?

В нашем докладе «Бедняки на пиру глобализации», опубликованном в ноябре 2002 года, мы приводим такие данные: компании тратят на перевод от 0,25% до 2% доходов, полученных за рубежом (в зависимости от отрасли и размера компании). Это намного меньше, чем те 3% от общей выручки, из которых обычно исходят организации.

Чтобы получить наши цифры, мы испытали другие методы оценки размера рынка, чтобы проверить, согласуются ли с ними наши данные. Мы проанализировали численность и доходы переводческих компаний, действующих на рынке, численность переводчиков во всем мире, а также документально подтвержденные государственные расходы. Сводная таблица этих данных вселяет в нас уверенность в правильности собственных цифр.

Каковы же основные результаты?

Мы постоянно изучаем переводческую отрасль в целом и подробно анализируем ее с разных сторон. На сегодня наше основное открытие состоит в том, что переводческие работы выполняются на среднем уровне управления и если и достигают руководства компании, то только в том случае, когда что-то не получается.

Мы также установили, что при объеме рынка 9,5 миллиардов долларов (2006 год) он примерно соответствует по размеру мировому рынку велосипедов, и что несмотря на все разговоры о переводческих технологиях совокупный доход всех игроков на этом поле составляет всего 100 миллионов долларов США.

Так вы хотите сказать, что перевод и локализация имеют лишь незначительное коммерческое значение?

Нет. Как раз наоборот. Хотя перевод очень дешев, он позволяет компаниям внедряться на новые рынки и умножать доходы. Сокращение стоимости перевода будет содействовать реальной глобализации рынков.

Не могли бы вы дать примерные цифры по крупным пользователям переводческих услуг?

Как вам известно, в 2004 году, когда в ЕС было принято 10 новых стран, ГУП потратил 1,1 миллиарда евро, но сейчас у них цифры ближе к 800 миллионам. Канадское бюро переводов имеет доход меньше 200 миллионов долларов США. Среди частных компаний много тратят на эти цели Microsoft и Oracle. Затраты Microsoft составляют примерно 300 миллионов в год, а Oracle – около 200 миллионов долларов США. Автомобильные компании тратят от 10 до 35 миллионов долларов в год. На рынке очень мало компаний, расходующих более 3 миллионов долларов в год, и всем хочется продать им свою продукцию.

2.6.3. От ручного перевода к машинному: резкое снижение затрат и улучшение доступности

Если на время абстрагироваться от проблемы качества, развитие машинного перевода определяют два рыночных фактора: стоимость и доступность. Существенное на порядок сокращение затрат будет, без сомнения, способствовать широкому использованию переводов, поскольку существующий скрытый спрос еще далеко не удовлетворен. Сильный рост спроса весьма вероятен и правдоподобен, причем он наблюдался и в других случаях, например в отрасли авиаперевозок с появлением недорогих авиакомпаний и в отрасли телекоммуникаций с появлением систем передачи голоса по протоколу VoIP. Следует подчеркнуть, что экономическому эффекту этой технологии способствуют не только переводы как таковые: значительную долю экономического эффекта составляют новые типы торговых операций, ставшие возможными благодаря технологическому прогрессу. Второй движущий фактор – доступность. Сегодня большинство переводов выполняется не как устный перевод в реальном времени, а как письменный перевод в удаленных офисах. Материал на перевод обычно куда-то отсылается, и готовый перевод предъявляется после многочасовой или многодневной работы. Устный перевод в реальном времени существует, но об-

ходится дорого. В сравнении со временем ожидания порядка нескольких часов немедленное наличие текстового перевода открывает целый ряд возможностей для новых сфер применения, которые были бы невозможны, если бы постоянно имела место задержка.

Стоимость и доступность – два важнейших рыночных фактора, стимулирующих развитие машинного периода.

С точки зрения качества, машинный перевод еще много лет будет уступать переводу, выполненному человеком. Вследствие этого на различных сегментах рынка будет господствовать одно из двух предложений: перевод, выполняемый человеком, или машинный перевод. Какое из двух предложений будет доминировать на том или ином сегменте рынка, будет зависеть от особенностей и потребностей этого сегмента. Ручной перевод будет преобладать во всех областях, где абсолютно необходимо высокое качество. Напротив, машинный перевод займет нижнюю часть¹⁴ рынка, а также захватит господствующие позиции на новых рынках или сегментах рынка, которые формируются благодаря наличию малозатратной технологии перевода. Со временем, с повышением качества и производительности машинного перевода, он переместится в более дорогую часть рынка.

¹⁴ Нижняя часть рынка относится к качеству перевода. С точки зрения времени цикла обработки и доступа, автоматический перевод явно находится в дорогой части рынка. Предполагается, что в будущем он найдет очень широкое применение.

Интервью с Майклом Анобиле, генеральным директором LISA

Международный бизнесмен с более чем 25-летним опытом работы в секторе ИТ, Майкл Анобиле получил степень бакалавра коммуникационных наук в Сиракузском университете и учился в магистратуре по политическим коммуникациям в Университете Мэриленд. После переезда в 1980 году с семьей в Швейцарию, где он занял должность Европейского менеджера по повышению квалификации в компании Exxon Office Systems, он занимал в европейских и швейцарских компаниях высшие административные посты по ИТ и переводческим технологиям, занимаясь в основном вопросами глобального развития бизнеса и маркетинга.

Являясь членом-учредителем LISA (Ассоциация стандартов индустрии локализации) и генеральным директором с момента ее основания, он отвечает за оперативное руководство Ассоциацией, в том числе за программы взаимодействия с другими организациями в сфере стандартизации (например, МСЭ (ISO), Unicode, OpenI18N, W3C, OASIS и т.д.) и правительственными организациями (например, с Министерством торговли США, Министерством обороны США, ФБР, Национальным виртуальным центром переводов, Национальной ассоциацией стандартов, различными азиатскими, канадскими и европейскими организациями, работающими в сфере переводческих технологий, стандартизации и торговли), а также за организуемые ею международные форумы, учебные программы и проекты по маркетингу и связям с общественностью.



Майкл Анобиле,
генеральный директор,
Ассоциация стандартов
индустрии локализации (LISA)
Ромэнмутье, Швейцария

Что такое LISA?

Деятельность нашей организации контролируется ее членами, и в центре ее внимания государственные органы и неправительственные организации. Помимо сектора малого и среднего бизнеса, мы работаем на институциональном уровне с такими крупными организациями, как Всемирный банк, Макдоналдс, Кока-Кола, IBM, Европейское Главное управление по вопросам перевода, министерство промышленности Канады и Канадское бюро переводов. В некоторых из этих организаций насчитывается свыше тысячи переводчиков. Мы помогаем им осознать мировые перспективы, открываемые локализацией с точки зрения проблем бизнеса, технологии и технологических процессов. Сюда входят интернационализация, перевод и способы создания продуктов и услуг для мирового рынка.

Что бы вы порекомендовали для правильного осуществления локализации?

Согласно передовой практике локализации изделие должно разрабатываться для международного применения. Соответственно, решающими вопросами являются затраты на интернационализацию, жизненный цикл или обслуживание изделия, рынки и продажу. Такой «целостный» подход предполагает перевод информации. Следовательно, любые возможности автоматизации процесса перевода имеют принципиальное значение с точки зрения качества, одобрения конечного пользователя и себестоимости.

Как бы вы оценили спрос на перевод, если бы он был бесплатным?

Понятия не имею. Однако практика показывает, что если вы спросите клиента, купит ли он перевод, сделанный машиной, скорее всего, ответ будет «нет». Но если предоставить ему перевод, фактически сделанный машиной, в качестве свободного выбора, тогда он, скорее всего, скачает этот документ. Налицо дилемма: имеется некоторое сопротивление тому, что вы ранее назвали «разрушающей технологией».

Что дала переводу технология?

Один из самых значительных вкладов в развитие перевода – система «переводческая память» (ПП). Кстати, этой технологии уже двадцать лет. Смотрите, сколько времени ей понадобилось, чтобы утвердиться. Я от души аплодирую успеху фирм TRADOS, SDL и другим разработчикам ПП, например Atril и Logos, потому что они здорово поработали, чтобы широко внедрить систему переводческой памяти во многих деловых секторах. ПП строится на базовой концепции обработки текстов: если вы однажды что-то напечатали, вам больше никогда не придется снова печатать это. Так что когда вы переводите нечто, что вам еще понадобится в будущем, вам не придется переводить это заново. Конечно, необходимо контролировать контекст отдельных участков памяти и обеспечивать их применимость на разных платформах и в разных системах. Именно в этом заключается вклад ТМХ© (стандарта LISA на переводческую память) в развитие отрасли.

Как вы относитесь к машинному переводу?

МП очень экономичен и эффективен, когда он применяется в строго определенных областях, например, в базах знаний, центрах обработки вызовов (call centers) и технической документации. Оптимальным методом в этой отрасли считается контроль за процессами авторской разработки и управления терминологией. Необходимо понимать, каким образом конечный пользователь будет взаимодействовать с системой, а также понимать требуемые уровни информации и качества. Тогда можно ускорить процесс автоматизированного перевода, формируя и ведя конкретные терминологические словари. Результатом будут большая точность и более высокая восприимчивость. При правильном выборе ожиданий и ресурсов МП работает прекрасно.

Одна из целей данного доклада – повысить информированность о технологиях автоматического перевода, а также о переводе как таковом. Что бы вы могли сказать по этому поводу?

Крайне важно, чтобы наши политические лидеры и крупнейшие бизнесмены понимали ту роль, которую играет язык. Язык – это посредник, а не помеха. Он улучшает взаимопонимание и сотрудничество между людьми и культурами. Язык обеспечивает более широкий доступ к важной общественно-политической информации. Я согласен с тем, что нужно вести информационно-разъяснительную работу. Ведь это – не просто перевод. Это – общение и доступ к важным данным, которые могут помочь людям стать более восприимчивыми в экономическом, политическом и социальном планах. Хорошим примером является языковая политика Главного управления по вопросам перевода Европейского Союза, направленная на поддержку развивающихся рынков в Восточной Европе и в новых странах-членах.

2.6.4. Поддерживающие и разрушительные технологические инновации

Разве не странно, что порой громадные, влиятельные и хорошо управляемые корпорации почему-то не хватаются за очередную инновацию? Так вот, загадка уже решена, и сегодня мы понимаем динамику инновации намного лучше, чем десятилетие назад. Инновации обладают определенными свойствами, которые определяют их успех на рынке и их шанс превратиться в абсолютно нового чемпиона рынка. Согласно этим свойствам инновации делят на *поддерживающие* и *разрушительные*. Ниже приводятся их дефиниции.¹⁵

Большинство технологических инноваций способствует улучшению эксплуатационных характеристик товара. Такие инновации называются поддерживающими. Независимо от того, являются ли они поэтапными или дискретными, общим для поддерживающих инноваций является то, что они улуч-

шают эксплуатационные характеристики укоренившихся на рынке товаров согласно критериям, традиционно ценным основным покупателями на важнейших рынках.

Случается, однако, что инновации приводят к ухудшению эксплуатационных характеристик товара, по крайней мере в краткосрочной перспективе. Эти *разрушительные инновации* обычно ухудшают эксплуатационные характеристики товара на основных рынках, но они обладают другими особенностями, ценными определенными покупателями. Обычно такие покупатели являются новичками на данном рынке.

Поддерживающие инновации, как правило, не изменяют рыночный пейзаж, т.е. лидер рынка разрабатывает их (или выкупает позже) и остается лидером рынка. Поскольку разрушительные инновации уступают существующей технологии или существующему товарному предложению, они не представляют ценности для сегодняшнего лидера рынка при существующей конъюнктуре. Под действием опреде-

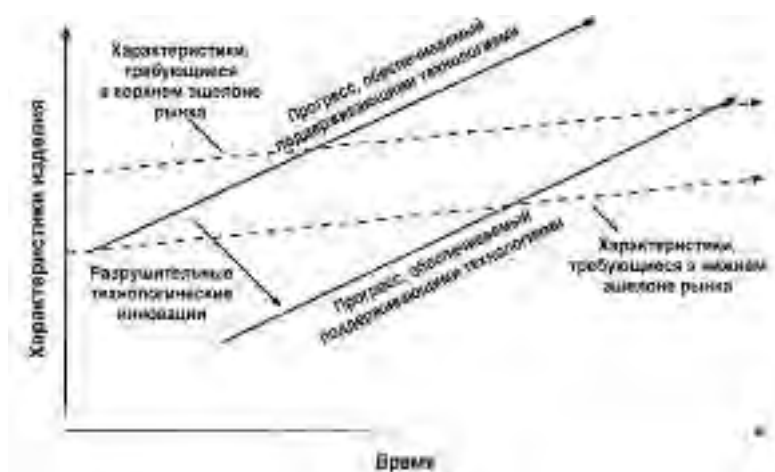


Рис. 3: Поддерживающие и разрушительные технологические изменения. – Источник: [CRa].

¹⁵ Два последующих параграфа близки к [CRa].

Привычная технология	Разрушительная технология
Мини-компьютеры	Персональные компьютеры
Ноутбуки	Портативные цифровые устройства
Галогенидосеребряная фотопленка	Цифровая фотография
Проводная телефония	Мобильная телефония
Microsoft OS и Microsoft Office	Linux и Open Office
Накопители на жестких дисках	Флэш-память

Таблица 8. Примеры разрушительных технологических инноваций вместе с соответствующей привычной технологией. – Источник: [Chr], [CRa].

ленных механизмов (которые и правдоподобны и количественно обоснованны) возникает ситуация, когда появляются новые игроки, которые расширяют эту рыночную нишу, не представляющую интереса для лидера рынка, пока она не разрастется до внушительных размеров. Любопытно, что рыночному лидеру не позволяет воспользоваться новой возможностью модель поведения, традиционно считающаяся хорошим стилем управления. Исторически сложилось так, что в случае разрушительной инновации лидер рынка оказывается не в состоянии сохранить свое положение и уступает место другому игроку.

Тогда перед сегодняшними рыночными игроками и политическими лидерами встает интересный вопрос: являются ли машинный перевод и перевод устной речи поддерживающими или разрушающими технологиями?

В случае систем перевода устной речи инновация является разрушительной: трудно понять, каким образом эта технология могла бы поддерживать переводчика-человека, и ее эффективность, хотя и улучшается, по-прежнему уступает устному переводу, выполненному человеком. Таким образом, ожидается, что эта технология будет внедряться на рынки совершенно иные, чем существующие сейчас, и будет постепенно расширять свою долю рынка. С другой стороны, эти рынки сбыта не

представляют интереса для переводчиков-людей (хороший пример – прямой перевод телевизионных передач).

Машинный перевод способен быть и поддерживающей и разрушительной инновацией, в зависимости от применения и от рынка, на который он выводится. К примеру, методы переводческой памяти сейчас применяются как подготовительное и вспомогательное средство для ручного перевода, причем они и облегчают перевод и повышают его качество. В этом контексте машинный перевод может рассматриваться как поддерживающая инновация: он улучшает то, что уже существует. Большинство же других применений машинного перевода будет резко разрушительным. Как и в случае перевода устной речи, услуги машинного перевода, например онлайнный перевод веб-сайтов, онлайнный перевод чатов и межъязыковое самообслуживание покупателей, уступают в качестве традиционному переводу. Однако ценовое преимущество обеспечит рост этих рынков, которые будут оставаться непривлекательными для любых традиционных служб перевода по причине низкой нормы прибыли.

Машинный перевод и перевод устной речи – в целом разрушительные инновации аналогично персональным компьютерам и цифровой фотографии: они начинают с одной ниши и позже набирают силу.

3. Азия, Европа и США: черты сходства и различия

В последние годы двумя важнейшими движущими силами в развитии переводческой технологии были события 11 сентября и расширение ЕС.

3.1. Европейский Союз

Языковая ситуация и потребность в переводческих технологиях в значительной мере уже описаны в главе 2.3.

Резюмируем основные положения:

- Европейское общество является многоязычным, и это не случайно;
- Все официальные языки Европы обладают равными правами;
- Прямой и обратный перевод с этих 20 официальных языков, образующих 190 языковых пар, требует значительной работы;
- Любая мера, облегчающая перевод, содействовала бы общению между гражданами Европы и расширяла бы торговлю внутри Европейского Союза;
- Любой компании, работающей на рынке ЕС, приходится адаптировать свои товары к региональным рынкам, учитывая местные языки.

Европейская комиссия заинтересована в исследованиях в области переводческих технологий и в течение ряда лет финансирует несколько научных проектов по машинному переводу.

Оперативно получить информацию о проектах можно на сайте IST (Information Society Technologies)¹⁶. В неполном списке числятся LC-STAR¹⁷, MATCHPAD¹⁸, METIS и METIS-II¹⁹, NESPOLE!²⁰, TC-STAR_p и TC-STAR²¹, TQPRO²² и TransType2²³. Европейская комиссия играла решающую роль в создании необходимых лингвистических ресурсов, финансируя множество проектов в этой области исследований.

Согласно 6-й Рамочной программе, Европейский Союз тратит 135 миллионов евро на многомодальные интерфейсы и переводческие технологии, т.е. примерно по 15 миллионов евро в год на переводческие технологии.

Европейское научное сообщество по своей природе хорошо подходит для решения многоаспектной задачи создания переводческих технологий. В Европе немало крупных государственных научных учреждений мирового уровня, что позволяет делить работу между ними и двигаться разными путями. Принципы многоязычия и культурного многообразия в целом понимаются и принимаются как в научной, так и в деловой сферах и при этом обеспечиваются политической инфраструктурой. Уже 10 лет действует инфраструктура языковых

¹⁶ IST project search: under <http://www.cordis.lu/ist/projects/projects.htm>

¹⁷ LC-STAR: <http://www.lc-star.com/>

¹⁸ MATCHPAD: <http://www.systransoft.com/R&D/Matchpad/index.html>

¹⁹ METIS-II: <http://www.ilsp.gr/metis/>

²⁰ NESPOLE! : <http://nespole.itc.it>

²¹ TC-STAR: <http://www.tc-star.org/>

²² via the IST page: under <http://www.cordis.lu/ist/projects/projects.htm>

²³ TransType2: <http://tt2.atosorigin.es/>

ресурсов – *Европейская ассоциация языковых ресурсов* (European Language Resources Association – ELRA).

3.2. Соединенные Штаты Америки

3.2.1. Стратегическая роль переводческих технологий (ПТ)

В Соединенных Штатах стратегическая роль ПТ существенно иная, чем в Европе. США имеют громадный внутренний рынок с фактически одним языком. И это справедливо, несмотря на то что в стране значительное латиноамериканское население, говорящее на испанском языке. Однако отсутствует сильная экономическая потребность в обслуживании этого экономически слабого слоя населения, и нет законодательно оформленного требования о поддержке этого языка. По существу, Соединенные Штаты представляют собой рынок с единственным языком.

Мировое господство английского языка как иностранного во многих отношениях выгодно для США, но таит в себе страшный недостаток в плане национальной безопасности. При таких обстоятельствах перевод, особенно с других языков на английский, играет решающую роль в сборе информации со всего мира и обо всем мире, будь то общая информация или сведения, относящиеся к разведывательным операциям.

Интерес США к ПТ в значительной мере определяется факторами национальной безопасности.

3.2.2. Исследовательские программы

После нападения на Всемирный торговый центр 11 сентября 2001 года сразу стало ясно, что необходимость в

переводчиках и людях, владеющих различными языками, должна резко возрасти. Это была одна из причин, почему Соединенные Штаты запустили грандиозную исследовательскую программу по сбору информации из источников на разных языках. Ее бюджет – более 50 миллионов долларов в год.

Эти средства, в несколько раз превышающие суммы, расходуемые Европейской комиссией, а также сама программа будут оказывать стимулирующее действие на научное сообщество. Если некоторые научно-технические достижения будут соответствовать конкретным запросам Европы, то с другими дело обстоит иначе: сбор сведений для военной разведки по целому ряду аспектов отличен от задачи перевода для многоязычного общества. Более того, фактически одноязычная структура американского внутреннего рынка также вызывает законные сомнения в том, что именно американский рынок переводческих технологий создаст то, что нужно Европе.

Не вдаваясь в подробности этих исследовательских программ, можно в целом констатировать, что США обычно финансируют небольшое число крупных проектов, в то время как ЕС финансирует крупномасштабные программы, состоящие из более мелких проектов. Сегодня флагманским проектом является GALE (Global Autonomous Language Exploitation), который решает задачу извлечения информации из разноязычных текстовых и звуковых документов в *неограниченной* области (т.е. речь идет и о переводе устной речи). Осуществляется и ряд других переводческих проектов, например, TransTac, STR-DUST и ACTD.

НЕСЕКРЕТНО

Ведомость обоснования строк бюджета по НИОКР, испытаниям и оценке							Дата Февраль 2005 г.	
Статьи ассигнований/бюджета Все оборонные программы, прикладные исследования				Номенклатура статей Информационные и коммуникационные технологии				
ЗАТРАТЫ (млн долл.)	2004 ф.г.	2005 ф.г.	2006 ф.г.	2007 ф.г.	2008 ф.г.	2009 ф.г.	2010 ф.г.	2011 ф.г.
Перевод с одного языка на другой	0,000	57,389	65,744	69,687	75,221	75,593	65,593	60,593

(U) Описание задачи:

(U) В проекте будет исследована и испытана мощная новая технология для обработки естественных языков, которая создаст принципиально новые возможности для удовлетворения самых различных нужд национальной безопасности. Благодаря этой технологии различные системы смогут: (а) автоматически использовать большие объемы устной речи и текстов на разных языках; (б) революционизировать взаимодействие человек – машина посредством английского и иностранных языков (как устных, так и письменных); (с) решать задачи расчета и принятия решений в напряженных, быстро меняющихся ситуациях; и (д) автономно подбирать, фильтровать, синтезировать и представлять релевантную информацию своевременно и в нужной форме. Данный элемент программы и проект сформированы в законопроекте по ассигнованиям Министерства обороны на 2005 финансовый год согласно намерению Конгресса. Финансирование на прошлый год выделено в бюджете PE 0602301E, Проект ST-29, и указано как докладная записка в каждой из приведенных ниже программ.

(U) Выполнение программ / Запланированные программы:

Представление ситуации и презентация	2004 ф.г.	2005 ф.г.	2006 ф.г.	2007 ф.г.
	(10,870)	(11,500)	(11,616)	(14,387)

(U) Имеются две программы, предусматривающие непосредственный устный перевод с одного языка на другой:

Программа CAST (Compact Aids for Speech Translation) обеспечивает тактическому бойцу возможность перевода лицом к лицу в реальном времени в боевой обстановке и в ходе гуманитарных операций на чужой территории. Программа решает проблему быстродействия и точности перевода в конкретных предметных областях. Первые опытные образцы программы CAST строились на простых словарях и словосочетаниях. Программа CAST позволяла прежде всего дать в распоряжение бойца систему быстрого одностороннего перевода (с английского на несколько иностранных языков). Главной из применяемых сегодня экспериментальных систем является Phraselator, разработанный Управлением перспективных исследовательских программ в области обороны. Система употреблялась в операции «Иракская свобода» и «Прочная свобода». Будущие версии сделают возможным более сложный, гибкий и гладкий перевод и перефразирование, они будут надежнее и более соответствовать нормальному диалогу между людьми.

Рис. 4. Бюджет Министерства обороны США на технологии перевода с языка на язык. Смета расходов Управления перспективных исследовательских программ в области обороны (DARPA) на 2005 финансовый год на исследования, разработки, испытания и оценку по всем оборонным программам. – Источник: DARPA.

Интервью с Джозефом Оливом, руководителем программы, DARPA

Д-р Джозеф Олив – руководитель программы Управления информационных технологий в DARPA. Сегодня в его портфеле одна крупная программа – GALE (Глобальная автономная эксплуатация языка).

Д-р Олив имеет более чем 30-летний опыт научно-исследовательской работы в центре Bell Laboratories. Он был лидером в синтезировании речи из текстовой информации и руководил командой мирового класса, занимавшейся системами компьютерного диалога и общением человека с компьютером. В роли Директора по изучению речи и Главного директора по технологиям в компании Lucent Speech Solutions он курировал окончательную доработку основных переводческих технологий Bell-Labs: «Распознавание речи», «Преобразование текста в речь» и «Опознавание говорящего».

Д-р Олив закончил Университет Чикаго с ученой степенью по физике (по специальности «Вычислительная атомная физика») и степенью магистра искусств по музыкальной композиции. По окончании университета он стал совмещать свой интерес к вычислениям и музыке, занявшись исследованиями в области акустики и обработки сигналов.

В 1974 г. д-р Олив получил от Национального фонда искусств грант на написание компьютерной оперы. В 1984 году он также получил от Bell Laboratories награду «Distinguished Member of Technical Staff».

Каковы научные направления и цели программы GALE?

Глобальная автономная эксплуатация языка (GALE) – это программа, которая будет выдавать нашим военным «отжатую», краткую и оперативную информацию. Поскольку источник может быть на иностранном языке, система GALE требует перевода и механизма дистилляции. Вводом может служить текст или устная речь, а выводом может быть полный перевод или «отжатый» ответ на запрос (не обязательно на естественном языке). Цель GALE – достичь такого уровня точности перевода и дистилляции, при котором система стала бы пригодной для военных целей. Программа GALE – продолжение двух программ DARPA по переводческим технологиям: EARS (программа транскрибирования) и TIDES (программа перевода, поиска, извлечения и синтеза информации). Предыдущие программы не были рассчитаны на построение сквозных систем перевода, но создали большой задел по переводческим технологиям.

Какова ваша позиция по вопросу о трансатлантическом научном сотрудничестве?

В Европе есть несколько групп, связанных с GALE. Они являются субподрядчиками при крупных командах, сформированных Главными исполнителями GALE, и сотрудничают внутри своих команд. В целом я считаю, что сотрудничество между американскими и европейскими/азиатскими учеными идет хорошо. Оно осуществляется на уровне правительств, корпораций и университетов.



Джозеф Олив,
руководитель программы,
Управление информационных
технологий (IPTO), DARPA,
Арлингтон, шт. Виргиния, США

Сколько государственных средств расходуется в США на переводческие технологии типа МП?

Общая программа DARPA по переводческим технологиям на 2006 год составляет примерно 50 миллионов долларов США. Я не думаю, что ее следует разбивать дальше, потому что переводческие технологии взаимосвязаны (или должны быть взаимосвязаны) и интегрированы, с тем чтобы добиться успеха в этой области. Я не располагаю информацией о расходах на ПТ других правительственных органов, но уверен, что DARPA не единственная организация, интересующаяся этой областью.

Насколько важны для США переводческие технологии, в особенности МП и родственные технологии?

Лично я считаю, что ПТ чрезвычайно важны для правительства США и военных. Нам необходимо общаться с союзниками и врагами и понимать их, а ведь многие из них не говорят по-английски. Также в связи с информационным взрывом нужно уметь находить важную информацию в «стоге сена».

Как организовано в США государственное финансирование исследований в области ПТ?

Мне не доводилось видеть в США никакого общего плана исследований по ПТ. Исследования поступают от различных правительственных органов и промышленности. Наблюдается широкое сотрудничество между государственными исследовательскими программами, корпоративными исследованиями и университетскими исследованиями, но оно носит скорее неформальный характер. Что касается DARPA, мы строим очень большие планы и ставим далеко идущие цели в отношении ПТ. И мы проводим периодические оценки для обеспечения прогресса. Мы работаем как путем сотрудничества, так и путем конкуренции. Наши команды нацелены на победу при оценках. Кроме того, мы увязали наши цели и оценки с полезностью технологии, а не просто с точностью измерения. ПТ имеют дело с языком, и в первую очередь их цель состоит в том, чтобы сохранить смысл в документе (устном или письменном).

Каков современный уровень развития переводческой технологии и с какими новыми научными проблемами придется столкнуться в этой области?

Технология МП в последние два-три года заметно усовершенствовалась. Этот прогресс в значительной мере обязан статистическим парадигмам в сочетании с методами оптимизации на основе BLEU. Впрочем, боюсь, что эта парадигма быстро исчерпывает свои возможности, если это уже не произошло. Я бы сказал, что машинный перевод находится сейчас примерно на полпути к тому, чем он должен быть. Хотелось бы увидеть появление новых идей для решения этой проблемы, пусть даже вначале результаты не будут конкурентоспособными. Необходимо, чтобы и МП, и технология автоматического распознавания речи взяли на вооружение комплексный подход, включив в себя другие методы ПТ, например информационный поиск, синтаксический анализ, извлечение и т.д. Хотя это тоже статистические методы, каждый из них работает по-своему, так что комбинированная система, использующая все эти технологии, могла бы улучшить результаты, что привело бы к решению задачи. Крайне важно еще теснее объединить автоматическое распознавание речи и машинный перевод, а не использовать первый лучший результат автоматического распознавания речи для запуска МП-компонента.

3.3. Восточная Азия

3.3.1. Английский как общепринятый язык в Восточной Азии

Учитывая разнообразие ее народов, культур и языков, Азия больше напоминает Европу, чем США. Но если Европа уже начала позиционировать себя как единое целое, то в Азии пока нет никакой ясной политики и нет ощущения себя как отдельной целостности. Страны общаются друг с другом в основном на английском языке, и широкое распространение в Азии английского как общепринятого языка говорит о том, что здесь существенно меньше спрос на прямой перевод языковых пар, например, тайский – японский. Преимущество заключается в том, что каждая страна в первую очередь должна гарантировать перевод на собственный язык с английского и с собственного языка на английский, когда на каждый язык приходится всего одна языковая пара, это упрощает ситуацию.

3.3.2. Азиатские языковые пары и растущее значение китайского языка

Однако использование в Азии английского как общепринятого языка общения имеет и обратную сторону. Возьмем в качестве примера задачу перевода между японским и китайским языками. Учитывая общие корни этих языков, слова в них имеют сходную этимологию. Было бы естественно воспользоваться этим обстоятельством, а также чертами культурного сходства. Применение английского

языка как посредника – не прямой способ, приводящий к ненужным осложнениям. Появилась новая тенденция: японские компании все чаще интересуются прямым переводом между японским и китайским языками. Это связано с бурным ростом китайской экономики и тем обстоятельством, что многие японские компании перебазируют производство в Китай. Интерес к переводу с китайского и на китайский растет и в мировом масштабе, и в Японии, и сегодня между Китаем и Японией более тесные отношения, чем в прошлом.

Английский является общепринятым языком в Азии. Растет значимость китайского языка.

Экономически сильная страна, язык которой имеет корни в древнекитайском, Корея поддерживает тесные связи и с Китаем и с Японией. Слова в японском и корейском языках в большинстве случаев имеют китайское происхождение, т.е. у них общая этимология, хотя в корейском языке не употребляются китайские иероглифы. Технические термины нередко вначале появляются в японском языке на основе китайских иероглифов, а затем воспринимаются в китайском языке. Это сходство в словарном запасе оказывается очень полезным при переводе; тем не менее эти три языка различаются произношением и совершенно различны в лингвистическом смысле. Учитывая эти общие культурные корни и тот факт, что торговля и контакты между этими тремя государствами сильнее,

чем с другими странами Азии, можно предположить, что на следующем этапе развития для трех языков – китайского, японского и корейского – будет непосредственно осуществляться прямой и обратный перевод, в то время как другие азиатские языки будут взаимодействовать главным образом через английский.

На следующем этапе для китайского, японского и корейского языков чаще будет осуществляться прямой и обратный перевод.

3.3.3. Исследовательские программы

В Японии пик исследований в области переводческих технологий пришелся на 80-е годы, когда на эти цели выделялись большие государственные средства, а компании проявляли большой интерес к проектам по машинному переводу, например к проекту EDR (*Electronic Dictionary Project*). Когда обнаружилось, что связанные с машинным переводом большие ожидания пока что не сбываются (вспомним европейский проект *EUROTRA*), интерес и государственное финансирование пошли на убыль. Сегодня имеются признаки возрождения интереса.

Учитывая господствующее положение английского языка, в исследованиях по переводу основной акцент делается на перевод с английского на родной язык и наоборот, что по существу приводит к тому, что научно-исследовательские программы осуществляются на национальном уровне. Межнациональные азиатские проекты было бы довольно трудно осуществить, поскольку в Азии для этого нет подходящей организации. Однако некоторые аспекты научной работы требуют международного сотрудничества, которое, как можно надеяться, будет осуществляться и поддерживаться азиатскими ассоциациями, например недавно основанной *Азиатской федерацией по обработке естественного языка* (*Asian Federation of Natural Language Processing*)²⁴.

Для стимулирования исследований в области перевода между языками Азии было бы весьма важно накопить языковые ресурсы для этих языковых пар. Учитывая требуемые затраты, а также необходимость совместного использования ресурсов, естественный путь к достижению этих целей – через международное сотрудничество. Следующим важным шагом на этом пути было бы создание специального органа оценки.

²⁴ <http://afnlp.org>

Интервью с Юн-ити Цудзии, директором Национального центра по разработке текстов в Манчестере и профессором в Манчестере и Токио



Юн-ити Цудзии, директор Национального центра по разработке текстов, Манчестер (Великобритания), профессор Школы информатики при Манчестерском университете, профессор факультета компьютерных наук Токийского университета

Профессор Юн-ити Цудзии в июле 2005 года был назначен директором Национального центра по разработке текстов (National Centre for Text Mining) и профессором по разработке текстов в Школе информатики при Манчестерском университете (Великобритания). Он также является профессором по обработке естественного языка на факультете компьютерных наук Токийского университета (Япония). В области обработки естественного языка он работает с 1976 года. Начав с машинного перевода, он расширил сферу своих исследований на грамматические формализмы для практических приложений обработки естественного языка, синтаксический анализ на базе HPSG-грамматики, извлечение информации и разумные ответы на вопросы. Его исследовательская группа по обработке естественного языка в Токийском университете недавно успешно применила глубокий синтаксический анализатор для получения семантических представлений всех рефератов в системе Medline (1,4 миллиарда слов).

Он считается признанным авторитетом в области разработки биотекстов, машинного перевода и обработки многоязычных текстов, и активно содействует паназиатскому сотрудничеству. Его приглашают давать консультации, читать лекции и делать основные доклады на многочисленных крупных конференциях как в области биоинформатики, так и в области обработки естественного языка. Он постоянный член ICCL (Международного комитета по вычислительной лингвистике) с 1992 года, был вице-президентом (2005 г.) и президентом (2006 г.) ACL (Ассоциации вычислительной лингвистики), президентом (2003–2005 гг.) IAMT (Международной ассоциации машинного перевода) и вице-президентом AFNLP (Азиатской федерации по обработке естественного языка).

В восьмидесятые годы машинный перевод был темой масштабных научно-исследовательских программ, но их результаты не оправдали ожиданий общественности. Как вы смотрите на эти работы сегодня?

Действительно, работы 80-х годов не оправдали ожиданий большого потенциального рынка. Как участник некоторых из этих проектов я был сказал, что мы испытали разочарование. Однако эти работы позволили создать прочную базу для исследований и разработок в области МП. Благодаря этим работам в Японии на рынке по-прежнему активно работают 6–7 поставщиков систем МП. Но важнее, как мне кажется, то, что наши тогдашние замыслы далеко опережали время. Иначе говоря, для достижения наших целей не было еще необходимых технологий. Полагаю, что теперь, когда мы располагаем технологиями, которых нам не хватало в то время, мы сумеем воскресить эту область. Например, одна и та же задача синтаксического анализа предложения с помощью сложных грамматических формализмов, на решение которой в то время требовались часы, сейчас занимает меньше секунды времени, когда мы выполняем ее с помощью нашей программы в Токийском университете.

Переводческие технологии, в том числе машинный перевод, должны представлять интерес для Азии. Так ли это воспринимается в разных странах?

К сожалению, нет. В отличие от Европы английский язык здесь давно считается единственным международным языком общения, и поэтому люди не думают, что имеет смысл работать с «местными языками» типа китайского, корейского, японского и т.д. Конечно, это неправильно, и люди начинают это понимать. Благодаря Интернету мы вдруг осознали, что существует громадный спрос на обработку «местных языков» и что наш местный язык на самом деле вовсе не местный.

В чем основные различия между азиатскими и европейскими языками? Насколько близки друг другу главные азиатские языки?

Все зависит от того, как мы определяем европейские и азиатские языки. Европейцы утверждают, что языки Европы разнообразны, и я соглашусь с этим. Но разнообразие азиатских языков колоссально и, как я думаю, намного превосходит в этом смысле европейские языки. Многие языки Индии, Среднего Востока, Малайского архипелага и др. совершенно не связаны с языками Дальнего Востока, китайским, корейским, японским. При этом китайский и японский/корейский языки, хотя и имеют общие словники благодаря долгой истории культурных обменов, принадлежат к совершенно разным языковым семьям.

Как бы вы охарактеризовали в целом ситуацию во всех странах Азии?

Я не подходящий человек для ответа на этот вопрос, но ситуация в азиатских странах опять же отличается разнообразием. Существует немало языков, для которых нет согласованных правил транскрипции, не говоря уже о стандартизации кодов символов. Впрочем, поскольку во многих странах, например Индии, Таиланде, Китае и т.д., наблюдается ускорение технологического прогресса, эти страны имеют более или менее сходный технологический уровень. Нам встречается немало интересных научных статей, опубликованных учеными из этих стран.

Наблюдается ли в Азии какая-либо международная деятельность в направлении координации научных исследований в этом секторе или даже сотрудничество в мировом масштабе?

Два года назад мы создали научную ассоциацию, Азиатскую федерацию по обработке естественного языка (AFNLP) с целью поощрять сотрудничество и координировать работы. Есть и другие инициативы. Но по сравнению с Европой мы еще очень далеки от идеала.

Какие следующие шаги нужно предпринять для содействия развитию ИТ и МП в Азии?

Необходимо более широкое участие государства. ЕС играет существенную роль в развитии НИР в области ИТ и МП. До конца 80-х годов Япония играла ведущую роль не только в технологии, но и в финансировании регионального сотрудничества в этой сфере. С тех пор ситуация резко изменилась. Мы видим несколько стран, которые могут внести не только интеллектуальный, но и финансовый вклад в развитие отрасли. Однако у нас нет никакого паназиатского правительственного органа, который координировал бы эту деятельность. Научное сообщество готово к сотрудничеству, но нам нужны, например, финансирующие органы, которые поддерживали бы все азиатские проекты.

3.4. Индия

Можете ли вы представить себе страну с таким же языковым разнообразием, как в объединенной Европе? Так вот, в Индии²⁵ с ее 35 штатами насчитывается 22 языка, которые считаются официальными, то есть утверждены конституцией. Мало того, каждый язык имеет в среднем по 20 диалектов. Кроме этого, все языки принадлежат к трем разным языковым семьям, и к тому же имеются разные письменности. Государственный язык – хинди, но в индийской конституции записано, что английский язык тоже может использоваться для официальных целей. В городах люди часто знают три языка: язык штата (например, хинди, бенгали и т.д.), хинди (официальный язык Индии) и английский. Примерно 30% населения говорит на хинди, и примерно 5% достаточно хорошо владеет английским языком.

Все официальные документы в столицах штатов должны быть на трех языках (на английском, хинди и языке штата). Самое распространенное направление перевода – с английского на хинди и с английского или хинди на соответствующие языки штатов. Но, как и в других странах, большинство переводов выполняется людьми, а их недостаточно, чтобы

удовлетворить спрос. Если говорить о технических средствах, то система «переводческая память» не имеет широкого применения, но перевод часто поддерживается электронными словарями и такими средствами, как морфологические анализаторы.

И общество, и правительство проявляют интерес к автоматическому переводу, и государство финансирует индийские исследования по машинному переводу. Международные компании типа IBM, Microsoft, Google и Yahoo инвестируют в МП, хотя он пока еще не работает с индийскими языками. Поскольку эти компании, как правило, используют статистический подход, существует множество параллельных фондов²⁶, но они не являются общедоступными. Начиная с 2006 года этот недостаток будет устранен, когда возникнет новый государственный орган – LDC-IL²⁷ (Консорциум лингвистических данных по индийским языкам – *Linguistic Data Consortium for Indian Languages*).

3.5. Экономические граничные условия

Экономические граничные условия применительно к переводческим технологиям различны в Европе и в США. На самом деле рыночные условия для систем распознавания

²⁵ Выражаю благодарность профессору д-ру Пушпаку Бхаттачария (Индийский технологический институт, Мумбай) за предоставленную информацию о ситуации в Индии.

²⁶ Еще один параллельный фонд – собрание текстов на двух версиях языка вместе с информацией о том, какие предложения связаны друг с другом. (Например, первые два предложения в языке А могут соответствовать первым трем предложениям в языке В и т.д.).

²⁷ <http://www.ciilcorpora.net/ldcil.htm>

речи и машинного перевода в США и ЕС прямо противоположны.

Вначале давайте рассмотрим автоматическое распознавание речи. В данный момент самая интересная с экономической точки зрения сфера использования автоматического распознавания речи и диалоговой технологии – самообслуживание клиентов, иначе говоря, автоматизация центров обработки вызовов, обслуживаемых людьми. Что касается этого рынка и этой сферы, Европа отличается от США в двух отношениях. Во-первых, если в США широко применяются так называемые интерактивные автоответчики (IVR)²⁸, то в Европе они не так распространены. Американцы уже довольно давно пользуются этими системами, нередко несколько утомительными, и в целом были вполне довольны переходом с DTMF-сервиса (двухтонального многочастотного набора номера) и жестких меню на речевые диалоги²⁹. В Европе, где в некоторых странах информационно-справочные службы бесплатны и предоставляют весьма высокий уровень обслуживания, введение автоматических систем часто рассматривается звонящими как шаг назад. К тому же США – это большой и достаточно однородный рынок с одним языком, гораздо больший, чем любой одноязычный ев-

ропейский рынок. В результате в США сильнее эффект масштаба, поскольку разработка устройства речевого диалога требует значительных усилий. С учетом этого неудивительно, что автоматическое распознавание речи больше прижилось в США, чем в ЕС.

Прямо противоположны экономические условия для машинного перевода: они трудны для США, но очень благоприятны для Европы в отношении и устной речи, и машинного перевода. Это связано с тем, что в Европе говорят на разных языках, в то время как любая система, используемая для громадного внутреннего американского рынка, охватывала бы лишь малую долю всей деятельности. Заниматься коммерцией в Европе значит быть многоязычным.

Благоприятные экономические условия для европейских компаний, предоставляющих услуги МП.

Также имеются мощные экономические силы, поддерживающие переводческую отрасль в Азии, но вследствие господствующего положения английского языка местный рынок, а также перспективы для переводческой индустрии меньше, чем в Европе. Учитывая текущую ситуацию и положение компаний, мы бы предположили, что при

²⁸ IVR означает «интерактивный речевой ответ». В классических системах IVR после нажатия звонящим кнопок проигрываются речевые сообщения.

²⁹ В часто цитируемом отчете «Nuance Communications - Market Research: Nuance Speech User Scorecard» (май 2000 г.) говорится, что у 87% респондентов высокая общая удовлетворенность распознаванием речи, намного выше, чем удовлетворенность DTMF-системами. Хотя по крайней мере этот результат обоснован и совпадает с собственным опытом автора, следует проявлять осторожность, ибо первоначальный отчет больше не присутствует ни на сайте компании Nuance, ни в той части Интернета, которая доступна для Google.

обычных обстоятельствах Европа будет играть здесь ведущую роль.

3.6. Место действия для Европы

Обобщая сказанное, стратегическая роль ПТ, а также ситуация в трех больших регионах выглядит следующим образом.

Европа: Языковая технология научно необходима с экономической, политической и культурной точек зрения. Преодоление языкового барьера будет способствовать общению и развитию экономики. Хотя ПТ уже являются предметом серьезных исследований в Европе, учитывая стратегическое значение этой технологии для Европы, она должна получить гораздо более высокий приоритет на повестке дня НИР.

США: Применение ПТ диктуется преимущественно военными сообра-

жениями и борьбой с терроризмом. На исследования и технологию ПТ выделяются сегодня очень значительные средства.

Азия: Главный приоритет – перевод с английского и на английский. Отсутствие общей политической идентификации и инфраструктуры мешает Восточной Азии взять на себя ведущую роль. Большой спрос на перевод наблюдается в Индии.

Что касается разработок и коммерческого использования, Европа ощущает в этом острую потребность и в то же время находится в выгодном положении. Учитывая общий характер данной технологии, ее промышленное внедрение возможно и в других регионах мира. С другой стороны, нельзя рассчитывать на то, что какие-то внешние поставщики удовлетворят наши потребности.

Интервью с Жозефом Марьяни, директором Департамента Министерства научных исследований Франции

Научная деятельность Жозефа Марьяни связана с переводческими технологиями, многомодальным общением человек – машина, распознаванием речи, языковыми ресурсами и вопросами оценки.

Он был президентом Европейской ассоциации языковых ресурсов (ELRA), президентом Европейской (ныне Международной) ассоциации речевой коммуникации (ISCA), членом правления Европейской сети по языку и речи (ELSNET), и координатором франкоязычной сети FRANCIL.

Д-р Марьяни был директором LIMSI и руководителем его Отдела коммуникации человек – машина (1989–2001 гг.), членом Научного совета Национального центра научных исследований (CNRS), председателем Консультативного совета по информационной науке и технологии CNRS и членом Комитета по оценке INRIA.

С 2001 года является директором Отдела информационно-коммуникационных технологий в Министерстве научных исследований Франции, где курирует исследовательские программы в области телекоммуникаций, программотехники, мультимедиа и нанотехнологий, в том числе конкретную программу по переводческим технологиям.



Жозеф Марьяни,
директор Департамента
информационно-
коммуникационных технологий
Министерства научных
исследований Франции
и старший научный сотрудник
LIMSI-CNRS, Париж (Франция)

По поводу Европы, США и Азии: в чем вы видите сходство между ними и в чем различие?

Соединенные Штаты с их преимущественно одноязычным внутренним рынком рассматривают многоязычие в первую очередь с точки зрения обороны и безопасности. Там все понимают по-английски, но с иностранными языками у них туго! В Азии, как и в Европе, множество языков, и там тоже ощущается коммерческая потребность в многоязычии, но это все же не общий рынок, как в ЕС. Для Европы, где в 25 странах – членах ЕС говорят более чем на 20 языках, многоязычие крайне важно. Помимо экономического аспекта, который делает многоязычие совершенно необходимым, существуют также политический, культурный и общественный аспекты. Это делает нашу ситуацию совершенно особенной.

Как вы относитесь к сотрудничеству с США или Азией?

Международное сотрудничество в совокупности со здоровой конкуренцией полезно для научно-технического прогресса, и я ценю и поддерживаю те инициативы, где наши страны объединяют усилия в науке и инфраструктуре для решения сложных проблем в области ПТ. С другой стороны, международное сотрудничество не может заменить программу удовлетворения наших собственных потребностей. Что касается многоязычия в Европе, то лидером должен быть ЕС – никто другой не может и не будет этого делать за нас.

Что нужно для того, чтобы укрепить позиции Европы?

Несмотря на довольно значительный объем работы, проделанной в прошлом, полагаю, что уровень финансирования научных исследований и технологий пока не соответствует масштабу вызова. Хотя Комиссар по вопросам образования, культуры и многоязычия Иен Фигель недавно подчеркнул значение многоязычия для Европы, переводческие технологии фигурируют в виде крошечной доли работ в области ИКТ в проекте 7-й Рамочной программы, потерявшись в конце технологического направления «Моделирование, визуализация, взаимодействие и смешанная реальность». А темы, отобранные для возможных крупных мероприятий в рамках Статьи 169, это «Научные исследования в Балтийском море», «Помощь престарелым» или «Метрология», каждая из которых весьма достойна, но, по моему мнению, имеет меньшее стратегическое значение для Европы, чем проблемы языка. Переводческие технологии – это ведь такая тема, которая лучше всего согласуется с принципом совместных действий в европейском масштабе, потому что идея охватить все технологии и все европейские языки слишком грандиозна, чтобы Европейская комиссия могла осуществить ее в одиночку. Ее вполне могли бы совместно разрабатывать страны – члены ЕС со своими языками и собственными программами. Если государства – члены должны в первую очередь учитывать то, что характерно для их языка (или языков), например языковые ресурсы (речевые и текстовые массивы, словари) и адаптированные к конкретному языку технологии, то Европейская комиссия могла бы заниматься в первую очередь аспектами, независимыми от конкретного языка, общей координацией работ, разработкой и оценкой базовых технологий и стандартами.

Европа все еще нуждается в инфраструктуре для оценивания переводческих технологий, сравнимой с Национальным институтом стандартов и технологий (NIST) в США, и финансирующие органы при отборе проектов должны учитывать результаты оценки опорных технологий, с тем чтобы не допускать финансирования разработки продуктов, явно требующих более совершенной технологии.

Это было бы дополнением к Европейской ассоциации языковых ресурсов (ELRA), только что отметившей свое десятилетие.

Да, действительно. Ситуация с языковыми ресурсами и оценкой переводческих технологий стала намного лучше благодаря таким постоянно действующим органам, как NIST и LDC (Консорциум по лингвистическим данным) в США или ELRA в Европе. Но проблемы, стоящие перед Европой, требуют большего. Я выступаю за создание Агентства переводческих технологий на уровне Европейской комиссии. Это должна быть постоянная структура, которая координировала бы работы Европейской комиссии и государств – членов ЕС по созданию необходимых переводческих технологий для многоязычной Европы и которая вывела бы Европу на передний край ПТ во всем мире. Тем самым она использовала бы языковые проблемы Европы в своих интересах.

4. Современное положение дел

4.1. Переводческая деятельность сегодня

Хотя без технологии никак не обойтись, предоставление услуг – это в конечном счете перевод, а также локализация.

Переводческая деятельность весьма разнообразна. Ее сфера – от перевода аттестата зрелости для абитуриента до работы с громадным веб-сайтом компании, входящей в список 500 крупнейших фирм (Fortune 500), и от языковых пар сходных языков (например, голландский и немецкий) до совершенно разных языков (например, итальянский и китайский), причем переводимые тексты могут представлять собой стандартный текст, художественную прозу или материал узко технического содержания.

Перевод осуществляется частично в тех учреждениях или компаниях, которые нуждаются в нем, а частично передается внешним переводческим службам. Двадцать крупнейших поставщиков переводческих услуг занимают всего 16,3% рынка³⁰, так что этот рынок действительно сильно фрагментирован. Если считать только компании со штатом минимум пять человек, оказывается, что в мире 5000 фирм, предлагающих переводческие услуги. Многие переводчики работают внештатно, причем даже крупные переводческие службы и крупные пользователи переводов привлекают к этой работе внештатников. Возьмем для примера

ГУП (Главное управление по переводам при Европейской комиссии): в нем доля работ, выполняемых внештатниками, повысилась с 11,8% в 1992 году до 23,0% в 2004 году.

4.2. Технологии, применяемые в профессиональном переводе

Важнейшие технологии, применяемые сегодня в переводческих службах, это переводческая память и терминологические базы данных.

Совершенно естественно, что техническая сложность труда переводчиков зависит в очень большой степени и от той компании, где он или она работает, и от характера переводимого материала. Если переводятся самые разные и недлинные тексты – это одна ситуация. Совсем иная ситуация, когда речь идет, например, о больших пакетах прикладных программ, которые необходимо локализовать в новой версии, причем уже существует локализация предыдущей. В этом случае имеет смысл учесть предыдущие версии. Вот почему в некоторых сферах перевода широко используется система «переводческая память» и существуют автоматические методы проверки, обеспечивающие унифицированность перевода слов. В качестве подспорья для единообразного и качественного использования терминологии применяются терминологические базы данных.

Машинный перевод еще не нашел широкого применения, но уже ис-

³⁰ Источник: Common Sense Advisory [CSA].

пользуется в некоторых системах. Машинный перевод подключается к работе, когда нужно очень оперативно получить хотя бы сырой перевод. Например, в Европейском Союзе сырые переводы используются для облегчения внутренних контактов. Когда же материал идет вовне, перевод выполнятся либо полностью от руки, либо результат машинного перевода выправляется переводчиком.

Такие технологии, как ПП и терминологические базы данных, не только сокращают расходы, но также повышают качество и единообразие перевода.

Необходимо понять, что «переводческая память» и терминологические базы данных не только повышают про-

изводительность труда переводчика, но также улучшают качество и согласованность его работы. Отмечалось, что машинный перевод используется в качестве сырого ввода для ручного перевода с целью повысить эффективность труда. Но главное преимущество машинного перевода заключается в быстром переводе текстов, которые необходимо понять для каких-то производственных целей, причем стоимость перевода, а еще чаще – срок выполнения перевода, исключает использование других методов. Поток оперативного онлайн-контента, требующего немедленного перевода, к тому же довольно небольшими кусками (например, ленты новостей), быстро растет.

Интервью с Кевином Боленом, главным специалистом по маркетингу компании Lionbridge

Кевин Болен руководит в компании Lionbridge направлением глобального маркетинга, связанным с определением перспектив развития рынка, управлением решениями, конкурентным позиционированием и продвижением марок. Г-н Болен перешел на работу в Lionbridge в 2005 году в результате поглощения компании Bowne Global Solutions, где он занимал должность вице-президента по маркетингу с момента прихода туда в начале 2002 года. Помимо руководства функцией глобального маркетинга, г-н Болен также занимался вопросами сбыта в Восточном регионе американского континента. До прихода в компанию Bowne он руководил функцией маркетинга в фирме LexiQuest, Inc., разрабатывающей программное обеспечение для предприятий и специализирующейся на переводческих технологиях. До прихода в LexiQuest г-н Болен шесть лет проработал в компании IBM Global Services, где занимал посты в разных сферах, от консультирования и экономического развития компании до маркетинга. Его последняя должность – старший управляющий маркетингом, курирующий разработку решений мирового масштаба и стратегическое планирование для отраслей розничной торговли, расфасованных товаров, транспортировки и производства.

Г-н Болен имеет степени бакалавра делового администрирования по специальности «Международный бизнес», которую он получил в Пейском университете, и магистра делового администрирования по специальности «Маркетинг и управление», полученную в Школе бизнеса Стерна при Нью-Йоркском университете.



Кевин Болен,
главный специалист по
маркетингу компании
Lionbridge, Уолтем,
Массачусетс (США)

Могли бы вы охарактеризовать в двух словах рынок локализации?

Рынок локализации сильно фрагментирован, и тысячи мелких компаний конкурируют друг с другом в региональном, вертикальном или функциональном планах. Одни из них предлагают полный комплект программ для локализации, другие же предлагают чисто переводческие услуги, нередко в качестве субподрядчиков для более крупных поставщиков услуг. Ряд крупных фирм также предоставляют дополнительные услуги, такие, как авторская разработка, художественный дизайн и устный перевод или вспомогательное программное обеспечение для управления процессом и активами локализации.

Каковы основные движущие силы рынка, в том числе определяющие его рост?

Компании, стремясь найти новых покупателей или сократить эксплуатационные издержки, продолжают расширяться в мировом масштабе, переводя производство в более дешевые регионы. Выход на эти новые рынки требует

локализации товара, маркетинга и другого оформления упаковки, подтверждающей или нормативной документации, а также материалов для обучения торгового персонала, специалистов заказчика и взаимодействия сотрудников.

Рынки Китая или Индии, которые формируются покупателями, ведут активную работу по локализации широкого круга товаров и услуг, причем по тем языкам и культурам, которые традиционно не обслуживались устаревшими предприятиями, что создает новые источники дохода для поставщиков услуг. С другой стороны, ускоряются жизненные циклы изделий, а это значит, что все чаще появляются все новые характеристики и новая информация. Это создает новые возможности для локализации, но при этом требуется, чтобы поставщики услуг пересмотрели свои производственные модели, ибо длительность цикла обработки и пересылки файлов становится критическим фактором в конкурентной гонке.

Каковы самые значимые технологии, используемые при переводе естественного языка?

С точки зрения производительности ведущей технологией по-прежнему является «переводческая память» (ПП). А традиционные модели лицензирования и настольные приложения выходят из употребления в эпоху интернет-ориентированных программных моделей, создаваемых по запросу. ПП на базе Интернета позволяет множеству переводчиков трудиться одновременно над одним проектом, эффективно используя работу друг друга, чтобы мгновенно улучшить коллективный результат.

Терминологические средства также являются ключевым обеспечивающим элементом, потому что они позволяют добиться большего единообразия по широкому диапазону материалов и производственных групп. Способность встроить это средство контроля языка в производственный процесс обеспечивает точность и одновременно ускоряет производство, что является важнейшим фактором в сегодняшнем «мире по запросу».

Каков ваш опыт машинного перевода?

Компания Lionbridge благодаря приобретению Bowne Global Solutions является сейчас владельцем «Барселоны» (Barcelona) – одного из самых уважаемых механизмов машинного перевода, основанного на системе правил. Хотя статистические системы и системы на базе примеров перспективны в исследованиях, их фонд наличных языков и областей все еще слишком ограничен, чтобы представлять практический коммерческий интерес на сегодняшнем рынке. Lionbridge уже успешно применял свой механизм «Барселона» на целом ряде клиентских проектов, что позволило сократить затраты времени и средств на переводческий компонент сводного проекта локализации. Необходимо тщательно определять сферу действия этих проектов, потому что система все же требует специального словаря и выработки пра-

вил, чтобы достигнуть уровня качества, при котором работу можно передавать на просмотр постредактору, в отличие от ручного перевода. Требуется достаточный объем слов, чтобы гарантировать прибыль на средства, затраченные на первоначальную настройку системы.

Качественная локализация и недорогой автоматический перевод ориентированы на разные рынки. Собираетесь ли вы также затронуть зону «разрушительных технологий»?

Lionbridge уже предлагает бесплатный веб-сайт для машинного перевода и дает возможность клиентам использовать МП для перевода текстов, за которые в противном случае они бы даже не взялись по причине дороговизны ручного перевода. В будущем, когда МП станет более конкурентоспособным, клиенты будут иметь возможность выбрать один из нескольких уровней решения; впрочем, качество – очень субъективное понятие. Мы убеждены, что клиенты стремятся иметь более комплексное решение, в котором использовалось бы все лучшее, что могут дать технологии «переводческой памяти», доступные через Интернет, машинный перевод и ручной перевод/постредактирование, для того чтобы обеспечить качество на уровне ручного перевода, но оперативнее и дешевле, чем сегодняшняя модель «переводческая память/ручной перевод». Это позволит выйти на новые рынки быстрее и при лучшей рентабельности инвестиций.

4.3. Исследования в области перевода «речь в речь» и необходимых технологий

В последние десятилетия тысячи ученых во всем мире занимаются изучением человеческого языка в его устной и письменной форме. Давайте же взглянем на различные технологии.

Автоматическое распознавание речи, или создание письменной последовательности слов, оказалось крайне сложной научной проблемой, на деле куда более сложной, чем даже игра в шахматы на уровне шахматиста. Научное сообщество последовательно изучало вначале более простые, а затем более сложные проблемы. Первоначально технология работала с небольшими словарями, системами, зависящими от говорящего в спокойных условиях и со вводом изолированных слов. Сейчас она уже развилась до уровня распознавания связной речи, независимой от говорящего и на основе очень больших словарей. Хотя население пока еще мало осведомлено об этой технологии, задача диктовки текста в профессиональной обстановке сегодня нередко поддерживается функциями распознавания речи; впрочем, все еще случаются ошибки распознавания. Теперь внимание научного мира направлено на следующую задачу, а именно, на произвольную речь говорящего. Произвольная в данном случае означает, что человек говорит без намерения быть распознанным автоматической системой. Типичный сценарий исследования – распознава-

ние разговора между людьми, ведущегося естественным, непринужденным языком. Сегодня системы распознавания речи нередко являются одним из компонентов многомодального перцептивного интерфейса³¹.

Преобразование текста в речь (TTS), иначе синтез речи, это порождение речи из письменного текста. Хотя системы TTS уже давно научились давать вполне внятный выход, последнее крупное достижение, «конкатенативные системы, использующие метод выбора единицы», всего несколько лет назад привело к созданию более естественно звучащих систем. Системы TTS играют важную роль в повышении мобильности взаимодействия человек – машина благодаря внедрению систем перевода устной речи. Если заранее записанная речь может применяться в основном к неизменяемому содержанию, связывая группы слов в предложения, то в случае переменного содержания это невозможно, и тогда требуется TTS. В центре внимания текущих работ по преобразованию текста в речь такие вопросы, как быстрый синтез различных голосов и исследование возникновения эмоций и контрастивной интонации.

Машинный перевод обеспечивает перевод письменного текста с входного языка на выходной. В последние десятилетия на решение этой крайне сложной проблемы было потрачено немало сил. Как и в случае распознавания речи, где примерно в 1990 году произошла смена методологических

³¹ Примеры – европейские проекты CHIL (Computers in the Human Interaction Loop, <http://chil.server.de/>) и AMI (Augmented Multi-party Interaction, <http://www.amiproject.org/>).

парадигм и методы на базе правил уступили место статистическому подходу, в исследованиях по машинному переводу тоже существуют два принципиально разных подхода. Более ранний подход почти полностью строится на методах лингвистики и детальном кодировании человеческих знаний, в то время как более поздний подход в очень большой степени управляется данными и использует варианты статистических методов, хорошо проявивших себя в распознавании речи. При любом из этих подходов новейшие системы обеспечивают переводы естественных текстов, которые все еще уступают по качеству ручному переводу и могут применяться лишь как сырые переводы, отражающие основное содержание текста вперемешку с ошибками. Для лингви-

стических методов характерно то, что они дают хорошие результаты на одних предложениях и совершенно не справляются с другими, в то время как при статистических методах ошибки распределяются более равномерно, т.е. все переводится более или менее прилично, но отличные результаты встречаются редко. Мы постарались дать правильное представление достигнутой на сегодня эффективности этих систем, но следует заметить, что в этой области наблюдается стремительный прогресс, а международная конкуренция и объективное эталонное тестирование также движутся вперед быстрыми темпами, особенно что касается систем, управляемых данными³².

Перевод устной речи (SLT) обеспечивает перевод устной речи на

Рис. 5: Функциональная схема перевода «речь в речь» (SST). Речь на входном языке преобразуется в текст в модуле автоматического распознавания речи (ASR). Текст вводится в компонент МП, который называется модулем перевода разговорного языка (SLT), что связано с тем, что в тексте содержатся ошибки распознавания и используются характеристики устного, а не письменного языка: грамматически неправильные предложения, заминки, неудачные начала. Его выход – текст на выходном языке – служит вводом в модуль «текст в речь» (TTS), порождающий речь на выходном языке. Для сохранения исходных характеристик голоса в модуль TTS вводится дополнительная информация



³² Сравните с [Neu].

входном языке в текст на выходном языке. Системы SLT, как и следует ожидать, состоят из двух основных частей: компонента распознавания речи и компонента машинного перевода, но это еще не все: в тексте встречаются ошибки распознавания, которые приходится исправлять. Другая трудность заключается в том, что в тех сценариях, по которым работают исследователи, анализируется такая речь, которая специально не предназначена для распознавания системой, т.е. речь, звучащая на телевидении, на совещаниях или в беседе двух человек.

Перевод «речь в речь» (SST). Объединяет все вышеуказанные методы. Здесь речь на входном языке переводится в речь на выходном языке, так что люди могут разговаривать друг с другом на разных языках, используя компьютер в качестве переводчика. Чтобы достичь этого, система SLT располагается каскадом с системой TTS, а также принимаются некоторые дополнительные меры. Во-первых, на выходном языке следует сохранить индивидуальные свойства говорящего, наряду с просодическими особенностями, фигурирующими во входном языке. Эти особенности, разумеется, не различимы ни в какой промежуточной письменной форме. Сейчас эта очень сложная проблема решается в рамках европейского проекта TC-

STAR, который посвящен решению задач перевода «речь в речь» применительно к дебатам в Европейском парламенте. Интересно отметить, что, хотя это и очень амбициозная задача, достигнуто уже достаточное высокое качество переводов, которое передает суть парламентских выступлений на незнакомом входном языке.

4.4. Проект TC-STAR

Перевод «речь в речь» – амбициозная научная задача. Чтобы упростить ее и сделать несколько более осуществимой, в рамках первых исследовательских проектов по этой теме в 1990-е годы (уже по свободной речи) работа велась в ограниченных областях, таких, как составление расписания встреч³³.

TC-STAR – первый совместный научно-исследовательский проект, решающий проблему перевода «речь в речь» в неограниченной области.

Нелегко описать, сколь велик этот шаг – переход от ограниченной области к неограниченной, но если вам когда-нибудь приходилось пользоваться карманным разговорником, чтобы заказать завтрак в португальской гостинице, а затем вы пытались послушать парламентские дебаты по какой-нибудь теме, тогда

³³ Большая работа в этом направлении проведена в рамках немецкого проекта Verbmobil (1993–2000 гг.) с совокупным бюджетом 53 миллиона марок. Международная инициатива TC-STAR, крупный мировой консорциум из 20 партнеров, разрабатывающий проблемы перевода «речь в речь», был организован в 1991 году и функционирует до сих пор.

вы получите некоторое представление. TC-STAR³⁴ (*Technology and Corpora for Speech to Speech Translation*) – первый совместный научно-исследовательский проект, в котором решается задача перевода «речь в речь» в неограниченной области. Кроме перевода китайских телепередач на английский язык, который был выбран с тем, чтобы иметь привязку к международным ориентирам, в рамках проекта TC-STAR ведется работа над речами в Европейском парламенте, произносимыми на английском и испанском языках носителями языка и теми, для кого эти языки не родные. Тематика речей – все, над чем работает Европейский парламент.

Эта деятельность сосредоточена на перспективных исследованиях по всем базовым технологиям перевода «речь в речь» (SST): распознаванию, переводу и синтезу речи. Цели проекта весьма амбициозны: осуществить прорыв в исследованиях по SST, с тем чтобы существенно сократить разрыв в качестве перевода, выполняемого человеком и машиной. Первая половина этого трехлетнего проекта уже дала очень не-

плохие результаты в решении данной сложной задачи, но еще многое предстоит сделать.

Помимо научной работы, по проекту TC-STAR создана инфраструктура, необходимая для ускорения прогресса в этой области. Собраны сведения, необходимые для управляемой данными методики, которая используется в проекте, и реализована инфраструктура оценки, построенная на сравнении конкурирующих образцов. Этот ориентированный на оценку подход гарантирует, что научные достижения признаются как таковые и что методы, созданные разными партнерами по проекту, могут сравниваться и соответствующим образом утверждаться. Наряду с комбинацией сотрудничества и конкуренции, этот подход призван максимально ускорить научный прогресс. В то же время консорциум оценивается во внешнем научном сообществе, и партнеры по проекту получили высшие баллы в ходе недавних оценок, проведенных на международных семинарах по переводу устной речи³⁵.

³⁴ TC-STAR (<http://www.tc-star.org>) – это интегрированный проект (IP), т.е. крупный проект внутри 6-й Рамочной программы.

³⁵ Международный семинар по переводу устной речи. IWSLT 2005, Питтсбург, США <http://www.is.cs.cmu.edu/iwslt2005/>. IWSLT 2004, Киото, Япония <http://www.slt.atr.jp/IWSLT2004/>

5. Сила опорной технологии

5.1. Нескончаемые потребности человека

Итак, ситуация на сегодня следующая. Машинный перевод только что начал применяться на практике, перевод «речь в речь» находится на переднем крае научных исследований, а множество родственных технологий находится в стадии разработки. С другой стороны, существует громадный скрытый спрос, который необходимо удовлетворить, и мы знаем, что когда-нибудь технология удовлетворит этот спрос. Можем ли мы представить себе будущее, пусть даже расплывчато?

Чтобы не увязнуть в деталях рыночных прогнозов, которые, возможно, пришлось бы пересматривать каждые два года, попробуем опереться на такие аргументы, которые указывают на значимые силы, определяющие спрос и предложение на будущем рынке. Существуют некоторые тенденции, в целом признаваемые как специалистами, так и неспециалистами, а именно: тенденции к мобильности, формированию сетей и распределенной работе компьютеров. Если мы поймем главные движущие силы технологического и коммерческого характера, цепочки приращения стоимости (точнее, сети приращения стоимости) и динамику рынка, мы можем

быть достаточно уверены, что наше представление о будущем не слишком неточно. И помните об одном трюке, который делает прогноз еще надежнее: лучший способ предсказать будущее это придумать его³⁶.

Если посмотреть на реальные истории успешного внедрения на рынок новых технологий, обнаруживаются два элемента, которые обычно совпадают: новая технология и старые потребности человека. Типичный пример – мобильная телефония, которая удовлетворяет нескончаемую потребность человека общаться с другими людьми и его желание оказывать влияние даже на расстоянии. В детстве у вас уже была связь с помощью коротких записочек на бумаге, которыми вы обменивались в школе, – почти в реальном времени, хоть и не равномерная; сегодня дети используют SMS-сообщения. А может быть, правильнее было бы сравнить сегодняшние SMS с телеграммами прошлого? Между ними есть сходство: в обоих случаях короткий текст передается по электронному каналу непосредственно адресату. Однако вследствие огромной разницы в цене и совершенно разных сценариев применения было бы неправильным считать SMS современной формой телеграммы. Попытка предсказать объем SMS-трафика на основе количества телеграмм, отправлявшихся двадцать лет назад,

³⁶ Фраза приписывается Алану Кею, пионеру персональных компьютеров и специалисту по пользовательским интерфейсам.

была бы совершенно несостоятельной. Точно так же было бы неверно и крайне неточно оценивать сегодняшнюю степень общественного пользования системой Google на основании данных о поиске информации на бумажных носителях или в базах данных двадцатилетней давности. Но даже в прошлом было ясно, что любая технология, удовлетворяющая первичные, по существу нескончаемые потребности человека, найдет признание: люди нуждаются в общении с другими людьми (телефон, мобильный телефон, электронная почта, SMS, чаты); люди желают осваивать пространство (автомобиль, самолет, также средства связи); люди жаждут информации и развлечений (Интернет, телевидение).

Нескончаемые потребности человека: общение с людьми, освоение пространства и времени, жажда информации и развлечений.

Существующий рынок перевода и локализации мало что говорит нам о том, что произойдет, если перевод будет доступен в реальном времени и по весьма умеренной цене. В очень многих ситуациях платить по одному евро за строчку переведенного текста – это уже не соответствует никакой добавленной стоимости! Если бы удалось сократить затраты на один или два порядка, это дало бы колоссальный рост использования переводов, пусть даже и не идеального качества. Человек ощущает сильную потребность общаться и получать информацию. Рынки постоянно меняются. Стоимость мигрирует из одних цепочек приращения стоимости в другие. Ста-

рые требования встречаются с новыми возможностями, и рождаются новые продукты и рынки.

Что бы произошло, если бы можно было получить перевод в реальном времени по разумной цене? Каков был бы эффект от подобного перевода текстов и речи, документов, веб-сайтов, видео, потокового контента? Что, если языкового барьера совсем бы не стало?

Мы не должны строить предположения на основании сегодняшнего положения дел в области перевода. Современный объем переводов – это лишь верхушка айсберга по сравнению со скрытым спросом на переводы. Разрешите обрисовать, чего бы хотелось лично мне.

Мне бы хотелось смотреть программы Аль-Джазира с субтитрами на своем языке, чтобы иметь представление о том, что думает арабский мир. Когда мы говорим по-английски с итальянскими и испанскими коллегами, нам иногда не помешало бы иметь какое-то подспорье, чтобы чуть расширить «пропускную способность» нашего канала связи. Мне нелегко читать французский документ, особенно если я не знаком с терминологией, так что хотя бы приблизительный перевод на английский или немецкий существенно ускорил бы процесс чтения. Во время отпуска в Португалии мне бы хотелось понимать меню и уметь обменяться парой дружеских фраз с местными жителями.

Представьте себе мир, в котором перевод можно получить недорого и немедленно.

Вы бы хотели просмотреть китайский веб-сайт? Приготовить угощение по чудесному рецепту, полученному во время отдыха в Испании, но который, увы, записан на испанском? Какие у вас запросы?

Нам, европейцам, будет все легче общаться друг с другом, и мы будем сближаться. Применение переводческих технологий принесет значительные экономические выгоды. Сегодня везде, где только возможно, автоматизируются контакты с конечными пользователями, в том числе составление многоязычных списков часто задаваемых вопросов; здесь и находится применение для ПТ. У компаний значительная часть нематериальных активов хранится в головах сотрудников и компьютерных файлах, созданных ими, причем на многих языках. Для любой международной компании жизненно необходимо по-

лучать информацию через языковые границы. Это будет интересный рынок, и такие компании, как IBM и SAP, Yahoo и Google, уже готовятся к нему. Перевод в реальном времени любого источника данных (лент новостей, динамического веб-контента, видео), межъязыковой поиск документов, более простой выход на рынки иностранных языков, особенно для специалистов, а также малых и средних предприятий, – возможности для Европы поистине безграничны.

Внедрению автоматического перевода будут способствовать три мощных рыночных фактора:

- дешевизна;
- работа в реальном времени;
- способность к автоматической обработке.

Любого из этих факторов достаточно для того, чтобы двигать бизнес.

Интервью с Димитрисом Сабатакакисом, генеральным директором компании SYSTRAN

SYSTRAN – лидер рынка по созданию программ и решений для настольных, корпоративных и интернет-систем, которые облегчают коммуникацию в 40 направлениях перевода (на 20 языковых парах) в 20 областях. Ведущие мировые корпорации, интернет-порталы и государственные учреждения отдают предпочтение программным продуктам компании SYSTRAN, которая имеет более чем тридцатилетний опыт НИР в этой области. Применение продуктов и решений SYSTRAN улучшает многоязычную коммуникацию и повышает производительность труда пользователей на рынках «бизнес – служащий» (B2E), «бизнес – бизнес» (B2B) и «бизнес – потребитель» (B2C), поскольку они предлагают решения на базе языков в реальном времени для целей информационного поиска, обработки контента, онлайн-поддержки клиентов, внутрифирменной связи и электронной коммерции.

Димитрис Сабатакакис родился в 1962 году в Афинах (Греция). Он закончил Страсбургский университет по специальности «Экономические науки» и вначале работал в финансовой сфере, а затем в промышленности. Поддержанный инвесторами, он взял под контроль и руководил оздоровлением компании Gachot, которая в 1995 году была продана группе компаний KEYSTONE/TYCO. С февраля 1997 года г-н Сабатакакис руководит компанией SYSTRAN.



Димитрис Сабатакакис,
генеральный директор
SYSTRAN, Париж
(Франция)

Как рынок машинного перевода соотносится с рынком традиционного перевода?

Традиционный рынок переводов устойчив и представляет процесс, выполняемый человеком, вроде письма. Совсем иной рынок МП (машинного перевода). Предлагаемые здесь переводческие услуги и приложения основаны на интернет-технологии и выполняются в реальном времени. Целесообразно применять МП для многоязычных публикаций, когда нужно перевести большой объем информации, причем пользователь контролирует исходный материал. 99% текущих приложений на рынке МП используется для «выделения сути», когда ежедневно переводятся миллионы страниц на таких порталах, как Yahoo! и Google.

МП, или программа для перевода с языка на язык, представляет собой автоматический процесс, который дает возможность пользователю:

- понимать контент на иностранном языке в переводе на родной язык в реальном времени, причем бесплатно;
- публиковать контент на разных языках в реальном времени путем «контроля» исходного текста.

Понимают ли люди возможности автоматических систем? Склонны ли они переоценивать или недооценивать их?

Ценность МП для целей аннотирования очевидна. Доказательством служат миллионы переводимых ежедневно страниц.

Ценность применения МП для многоязычных публикаций тоже очевидна, но здесь нередко недооценивают или упускают из вида значительные инвестиции, необходимые для того, чтобы «контролировать» или «структурировать» исходный текст. Чтобы осуществить это, весь производственный процесс издания должен быть с самого начала ориентирован на многоязычие. На практике компании создают контент (например, технические описания изделий, материалы по маркетингу, информацию по техническому сопровождению, базы знаний и т.д.) на одном исходном языке, обычно английском, и рассматривают локализацию контента лишь как дополнительный шаг. Из-за бюджетных ограничений объем адаптированного текста бывает гораздо меньше, чем исходного. Кроме того, сохранение адаптированных вариантов контента – дело непростое, поскольку оно требует больших расходов, медленное (процесс, выполняемый человеком) и не позволяет полностью решить проблемы непоследовательности терминологии.

Считаете ли вы МП просто способом рационализации ручного перевода или скорее новым типом перевода?

МП – это и рабочий инструмент для ручного перевода, и новый тип перевода. Переводчикам-людям нужно взять на вооружение МП, потому что он повышает производительность труда пользователя и экономит время. Программное обеспечение пока что не нашло широкого применения, потому что эргономика и пользовательский интерфейс не были приспособлены специально для переводчиков-людей. В компании SYSTRAN мы много сделали для того, чтобы создать подобные инструменты, и сегодня МП может применяться в условиях традиционных переводческих услуг для повышения производительности труда.

6. Заключение

Оперативные, надежные и дешевые способы общения, передачи и обработки данных – это опора современного информационного общества. То, что справедливо для обычных информационно-коммуникационных технологий, справедливо и для технических систем, переводящих с одного языка на другой, как в целях коммуникации, так и в целях доступа к информации. Для Европейского Союза с его двадцатью официальными языками и еще множеством разговорных языков быстрый, надежный и дешевый перевод совершенно необходим, и переводческая технология должна рассматриваться как стратегически важная для ЕС.

В других частях мира услуги автоматического перевода тоже нужны, но в иных условиях. Никто не решит наши проблемы; будет лучше, если мы решим их сами. И у нас есть для этого средства.

Ввиду того, что в Европе существует потребность в машинном переводе и переводе устной речи, для компаний, намеревающихся предложить переводческие технологии и услуги, сейчас имеется благоприятная рыночная конъюнктура. Несколько сильных европейских научно-исследовательских групп могут

выступать поставщиками переводческих технологий. За пределами европейского рынка спрос на машинный перевод будет большим в Японии, Китае и Корее, а также в Индии.

Переводческая отрасль быстро растет, но простое повышение производительности сегодняшнего ручного перевода – это далеко не весь потенциал машинного перевода. Предполагается, что он развернется в сегментах рынка, которые сейчас едва существуют и имеют настолько низкий коэффициент прибыльности, что никогда не будут представлять интереса для традиционного перевода. Точно так же, как невозможно было экстраполировать применение SMS-сообщений на основании масштаба использования телеграмм, нам будет трудно прогнозировать объем перевода в будущем. Он будет очень большим и поможет раскрыть ту часть нашего экономического потенциала, которая все еще блокируется языковыми барьерами.

Все эти инновации требуют времени, но они уже появились и будут оказывать громадное влияние. Тем временем нужно установить такой курс на будущее, который будет благоприятным для нас. И формировать его предстоит нам.

7. Использованные материалы и дополнительная информация

7.1. Библиография

- [Chr] C. Christensen: *The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail.* Harvard Business School Press, Boston, Mass., 1977.
- [CRa] C. Christensen, M. Raynor: *The Innovator's Solution. Creating and Sustaining Successful Growth.* Harvard Business School Press, Boston, Mass., 2003.
- [Cry] D. Crystal: *English as a Global Language.* Cambridge University Press. 2nd edition, 2003 (1st ed. in 1997).
- [CSA] R. Beninato, D. DePalma: *Ranking of Top 20 Translation Companies.* Common Sense Advisory, Inc., June 2005. Can be downloaded from the Common Sense Advisory website <http://www.commonsenseadvisory.com> under http://www.commonsenseadvisory.com/members/res CGI.php/050701_QT_top_20.php
- [DGT] The Directorate-General for Translation of the European Commission (DGT). http://europa.eu.int/comm/dgs/translation/index_en.htm
- [EB1] Europeans and languages. Eurobarometer 54 Special. INRA Report, 60 pages, Feb. 2001. Отчет можно скачать с сайта: http://europa.eu.int/comm/education/policies/lang/languages/barolang_en.pdf о нем идет речь в http://europa.eu.int/comm/education/policies/lang/languages/index_en.html. Просьба принять к сведению, что эти данные относятся к прежним 15 странам ЕС.
- [EB5] Europeans and languages. Eurobarometer 63.4, September 2005. Обследование, проведенное в 25 государствах-членах ЕС, в странах, готовящихся к вступлению в ЕС (Болгария и Румыния), в странах-кандидатах (Хорватия и Турция) и среди турок-киприотов. Отчет можно скачать с сайта: http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_237.en.pdf о нем идет речь в <http://europa.eu.int/languages/en/document/80/20>
- [ELP] Languages of Europe. На официальном веб-сайте Комиссии Европейского Сообщества, касающемся языков, на которых говорят в ЕС. http://europa.eu.int/comm/education/policies/lang/languages/index_en.html
- [Gra] D. Graddol: *The Future of English? A guide to forecasting the popularity of the English language in the 21st century.* (A report commissioned by The British Council). The English Company (UK), 64 p., 2000 (first published 1997), ISBN 0-86355-356-7.

- [Gri] B. Grimes: Ethnologue Language Database. <http://www.sil.org/ethnologue/> «Этнолог» – каталог более чем 6700 языков, на которых говорят в 228 странах.
- [Gor] R. G. Gordon (ed.): Ethnologue: Languages of the World. Fifteenth edition. Dallas, Tex., 2005, SIL International. Online version: <http://www.ethnologue.com/>
- [Ney] H. Ney: One Decade of Statistical Machine Translation: 1996–2005. In Proceedings of the MT Summit X, pp. i-12 - i-17, Phuket, Thailand, September 2005. Документы можно скачать с сайта <http://www-i6.informatik.rwthachen.de/web/Publications/index.html>
- [TMC] Translating for a multilingual community. Directorate-General for Translation of the European Commission (DGT), April 2005, 19 pages. Брошюру в формате pdf можно скачать с сайта: http://europa.eu.int/comm/dgs/translation/bookshelf/brochure_en.pdf
- [TTW] Translation tools and workflow. Directorate-General for Translation of the European Commission (DGT), April 2005, 25 pages. Брошюру в формате pdf можно скачать с сайта: http://europa.eu.int/comm/dgs/translation/bookshelf/tools_and_workflow_en.pdf

7.2. Для дальнейшего ознакомления

Common Sense Advisory <http://www.commonsenseadvisory.com/>

Europa languages portal <http://europa.eu.int/languages/en/home>

LISA – Localization Industry Standards Organization <http://www.lisa.org/>

Multilingual Computing, Inc. <http://www.multilingual.com/>

EAMT – European Association for Machine Translation <http://www.eamt.org/>

ELRA – European Language Resources Association <http://www.elra.info/>

ELSNET – European Network in Language and Speech <http://www.elsnet.org/>

GALA – The Globalization and Localization Association
<http://www.gala-global.org/>

7.3. Таблицы

Таблица 1. Торговая конъюнктура для международной торговли внутри ЕС в отношении реальных товаров и информации (информационных услуг).

Таблица 2. Количество говорящих на основных международных языках как на родном (в миллионах человек) согласно двум источникам: модель «engco» (English Company) [Gra] и сравнительные цифры из «Этнолога» ([Gri]; см. [Gra]).

Таблица 3. Уровень «мирового влияния» 12 главнейших языков согласно модели «engco» (см. таблицу 2). 100-балльный показатель соответствует позиции английского языка на 1995 год [Gra].

Таблица 4. Отрасли знания, в которых немецкие ученые считают английский своим рабочим языком [Gra].

Таблица 5. Двадцать официальных языков Европейского Союза и их сокращенные обозначения [ELP]. Эрский (ирландский) язык станет 21-м официальным языком ЕС с 1 января 2007 г.

Таблица 6. Владение иностранными языками в ЕС: Доля граждан ЕС, говорящих на соответствующем языке в ЕС (как на родном / как на втором или иностранном / на обоих). Обследование было проведено в 25 странах–членах ЕС, а также в странах, готовящихся к вступлению в ЕС (Болгарии и Румынии), странах-кандидатах (Хорватии и Турции) и среди турок-киприотов. – Источник: [EB5].

Таблица 7. Классификация 20 ведущих поставщиков переводческих услуг по прибыли за 2004 г. (В цифрах учтены два важных слияния 2005 г.: поглощение компании Bowne Global Solutions компанией Lionbridge и поглощение компании TRADOS компанией SDL.)

Таблица 8. Примеры разрушительных технологических инноваций вместе с соответствующей привычной технологией. – Источник: [Chr], [CRa].

7.4. Рисунки

Рис. 1. Доля мировой книжной продукции на каждом языке. Английский – самый распространенный иностранный язык в книгоиздании: книги на английском языке выходят более чем в 60 странах мира [Gra].

Рис. 2. Языки мира. Каждая точка соответствует первичному местонахождению живого языка, учтенного в «Этнологе».

Рис. 3. Поддерживающие и разрушительные технологические изменения. – Источник: [CRa].

Рис. 4. Бюджет Министерства обороны США на технологии перевода с языка на язык. Смета расходов Управления перспективных исследовательских программ в области обороны (DARPA) на 2005 финансовый год на исследования, разработки, испытания и оценку по всем оборонным программам. – Источник: DARPA.

Рис. 5. Функциональная схема перевода «речь в речь» (SST).

7.5. Список сокращений

AFNLP	Asian Federation of Natural Language Processing (Азиатская федерация по обработке естественного языка)
ASR	automatic speech recognition (автоматическое распознавание речи)
BLEU	статистическая мера качества перевода, находящаяся в хорошем соответствии с суждениями человека; чем выше показатель, тем лучше
CEC	Commission of the European Communities (Комиссия европейских сообществ)
DARPA	Defense Advanced Research Projects Agency (Управление перспективных исследовательских программ в области обороны)
DGT	Directorate General for Translation (Главное управление по вопросам перевода)
DTMF	dual tone multiple frequency (двухтональный многочастотный набор)
ELRA	European Language Resources Association (Европейская ассоциация языковых ресурсов)
EU	European Union (Европейский Союз)
GALE	global autonomous language exploitation (глобальное автономное использование языка)
HLT	human language technology / human language technologies (переводческие технологии)
HT	human translation (ручной перевод, перевод, выполненный человеком)
IP	здесь: integrated project (интегрированный проект) (в других контекстах: интернет-протокол, интеллектуальная собственность)
IR	information retrieval (информационный поиск)
IT	information technology (информационная технология)
IVR	interactive voice response (интерактивный речевой ответ)
LISA	Localization Industry Standards Association (Ассоциация стандартов индустрии локализации)
MT	machine translation (машинный перевод)
NLP	natural language processing (обработка естественного языка)
ROI	return on investment (отдача на инвестиции)
SLT	spoken language translation (перевод устной речи)
SME	small or medium-sized enterprise (малые и средние предприятия)
SMS	short message service (служба коротких сообщений)
SST	speech-to-speech translation (перевод из речи в речь)
TC-STAR	Technology and Corpora for Speech-to-speech Translation (Технология и совокупность текстов для перевода из речи в речь)
TM	translation memory (переводческая память)
TTS	text-to-speech (из текста в речь)

Finito di stampare
nel mese di maggio 2006
dalla Tipolitografia TEMI - Trento

Переводческие технологии для Европы
Руководитель проекта: Джанни Лаццари

Редакторы *Е.И. Кузьмин, Е.В. Плыс*
Корректор *Е.Г. Сербина*

Компьютерная верстка *И.М. Горюнов*
Ответственные за выпуск *С.Д. Бакейкин*

Издатель:

Межрегиональный центр библиотечного сотрудничества (МЦБС)
105066, г. Москва, 1-й Басманный пер., д.м 2а, стр. 1
Тел.: (495) 267 33 34, факс: (495) 657 96 20
www.mcbs.ru

Подписано в печать 12.05.08
Формат 60 x 90/16
Печ. л. 4
Заказ №
Тираж 1000 экз.

Отпечатано в типографии