

Перевод в киберпространстве

Теперь, когда Интернет становится все более важным источником информации, появилась необходимость расширить доступность документов для широкого круга людей на их родных языках.

Жан-Пьер Клотье, автор колонки новостей «Кибернетические хроники», сказал в 1999 году: «Во всемирной паутине слишком мало переводов важнейших текстов и эссе с английского на другие языки, а также и наоборот».

И он не одинок, вызывая к большей доступности в Интернете информации на других языках, кроме английского. «Когда же мы сможем прочесть перевод испаноговорящих мыслителей и общественных деятелей? — спрашивает Мария Виктория Маринетти, профессор делового испанского языка и переводчик. — Общение на разных языках необходимо. Я бы даже сказала обязательно, потому что информация в Интернете предназначается всему миру. Так почему же мы не можем получать ее на родном языке или на языке, на котором нам хочется читать ее? Разве глобальный доступ к информации и ограниченность выбора языков не противоречат друг другу?»

Не следует забывать, что люди, выходящие в Интернете, зачастую говорят только на одном языке. Например, Мириам Меллман работает в San Francisco Chronicle, очень популярной в США газете. Она говорит только по-английски. «Хотелось бы, чтобы для таких ленивых людей как я сделали программы быстрого перевода, — говорит она. — Если я когда-нибудь решусь выучить иностранный язык, это будет только один язык из огромного множества. Перевод мог бы сделать общение намного проще».

И даже люди, говорящие на нескольких языках, высказывают подобное же пожелание. Жерар Фурестье — создатель Rubrique a Vas, сайта, предоставляющего базы данных для старшеклассников и студентов, рассказывает: «Я — француз. Я учил немецкий, английский и арабский, но по-прежнему очень далек от того, чтобы познать все законы всемирной паутины. Жаль, если только самые большие и мощные группы будут представлены в сети. Что же касается перевода, там еще непечатый край работы».

Пьер-Ноэль Фавенек, специалист France Telecom R&D, говорит о том, что исследования должны проводиться не только в области автоматизированного перевода на основные языки, но и на все остальные тоже.

Автоматизированный перевод

Само собой разумеется, что по качеству исполнения автоматизированный перевод не идет ни в какое сравнение с работами переводчика-профессионала. Однако профессиональный перевод требует много времени и зачастую дорого стоит.

Уже существующие программы перевода вполне пригодны для использования, они переводят моментально и очень дешево, или вовсе бесплатно. Web-страница или небольшой текст могут быть переведены за считанные секунды, возможны различные комбинации языков. Программа берет исходный текст и автоматически генерирует соответствующий ему на желаемом языке, используя особые правила перевода грамматических структур.

Европейская ассоциация машинного перевода (ЕАМТ) заявляет следующее: «На сегодняшний день уже есть несколько систем, обеспечивающие не блестящий, но приемлемый перевод, который может использоваться в определенных случаях, обычно в области технической документации. Мало того, программы перевода, созданные в помощь переводчикам-людям, становятся все более популярными в среде профессионального перевода».

В 1998 году компания Globalink, которая специализировалась на программах перевода и переводческих услугах, опубликовала на своем сайте историю автоматизированного перевода. (Этот сайт больше не существует. Компания Globalink была куплена Lernaut & Hauspie, которая в свою очередь была куплена Scansoft.) В этих материалах рассказывалось, что концепция машинного перевода и обработки естественных языков впервые появилась в 1930-х годах одновременно с компьютерной обработкой данных. Во время второй мировой войны благодаря достижениям криптографии и работам по расшифровке военных кодов появились первые программируемые компьютеры.

После войны в зародившемся секторе информационных технологий не угасал интерес к переводам и анализу текстов на естественных языках. В 1950-е годы были проведены исследования по художественному переводу. Это был пословный перевод, который не учитывал языковые правила. Проект по русскому языку, начатый в 1950 году Университетом Джорджтауна, был первой попыткой создать пригодную к употреблению систему автоматизированного перевода. Опубликованная в 1965 году работа Нозля Чомски, озаглавленная «Аспекты теории синтаксиса», отразила быстрый прогресс теоретической лингвистики.

Автор предложил новые определения фонологии, морфологии, синтаксиса и семантики живых языков. Однако в 1966 году в Америке была дана отрицательная официальная оценка системам автоматизированного перевода. В результате этого на 10 лет было прекращено финансирование и эксперименты в этой области.

И только в конце 1970-х годов возобновились серьезные экспериментальные работы. В то же время продолжалось развитие информационных и языковых технологий. В этот период разрабатывались системы перехода с языка на язык и делались первые попытки продаж автоматизированных систем перевода. Такие компании как Systran и Metal увидели жизнеспособность и полезность данного сектора рынка. Они стали предлагать продукты автоматизированного перевода и услуги через центральный сервер.

Но по-прежнему оставалось немало проблем: высокая стоимость разработок, ограниченность предложения по языкам, недоступность

подобных систем для среднего потребителя и сложности при переходе на новые стадии разработки.

В 1999 и 2000 годах распространение Интернета и возникновение электронной торговли создали устойчивый рынок. Три компании, Softissimo, Systran и Lernaut & Hauspie, выпустили продукты для широкой публики, профессионалов и производителей. Softissimo выпустила серию программ перевода Reverso, а также многоязычные программы письменной речи, электронные словари и лингвистические обучающие программы. Reverso, например, используется France Telecom в их машине поиска Voila.

Systran выпустил программу для перевода, которой в основном пользуется поисковая машина AltaVista. А Lernaut & Hauspie (после того как их купила компания ScanSoft) предлагает продукты и услуги по записи, переводу, усилению звука, модуляции голоса и по промышленной документации.

В марте 2001 года компания IBM вышла на этот постоянно расширяющийся рынок. Она выпустила лучший продукт для профессионалов — WebSphere Translation Server. Он представляет собой пакет программ, которые осуществляют моментальный перевод Web-страниц, электронной почты и чатов на китайский, английский, французский, немецкий, итальянский, японский, корейский и испанский языки. Программы обрабатывают 500 слов в секунду, при необходимости можно добавить специализированную лексику.

В июне 2001 года вновь сформированная компания Champollion Wordfast, которая предлагает услуги по разработке программ перевода, локализации и управлению многоязычной информацией, выпустила продукт Wordfast. Это — пакет программ, который оснащен набором лексики в реальном времени, контролем правописания и совместим с пакетами программ IBM Translation Server, TMX и Trados. Упрощенную версию Wordfast и руководство для пользователя на 16 языках можно загрузить бесплатно.

Многие общественные организации тоже принимают участие в исследованиях и разработке программ автоматизированного перевода. Ниже приведены три примера: одна организация в англоговорящей об-

щине, вторая во франкоговорящей общине и третья в многонациональной общине.

Группа естественных языков совместно с Институтом теории информации Университета Южной Калифорнии (USC/ISI) работает в нескольких областях исследований естественного языка, таких как автоматизированный перевод, краткое изложение текстов, подача команды на нескольких языках, разработка концептуальной систематики (онтологии), создание текста, компиляция огромных, многоязычных глоссариев и мультимедийная коммуникация.

Во Франции в Гренобле при Институте информатики и прикладной математики создана комплексная группа, состоящая из специалистов в информационных технологиях и лингвистов, известная как GETA. Эта группа занимается исследованиями теоретических, методологических и практических аспектов компьютерного перевода, а также многоязычными информационными технологиями в целом. GETA вносит свою лепту и в создание Всемирного языка сетевого общения (UNL) — цифрового метаязыка — для кодировки запоминающих устройств, исследований и передачи многоязычной информации независимо от языка источника.

Этот метаязык разработан Международной программой UNL, в ней участвуют различные языковые общины. Программа работает при Институте современных исследований Университета ООН под эгидой Фонда всемирного цифрового языка для сетевого общения (UNDL).

Взгляд в будущее

При эффективной работе программы автоматизированного перевода результат появляется сразу вслед за моментальной расшифровкой лингвистических, культурных и национальных кодов.

Как говорит Рэнди Хоблер, специалист по маркетингу программ и услуг по переводу в Интернете: «Мы очень быстро подойдем к моменту, когда хороший перевод текста и слова станет настолько естественным, что превратится в часть платформ или даже чипы».

Сейчас, когда развитие Интернета достигло крейсерской скорости, когда перевод соответствует оригиналу на 98 %, и когда различные комбинации языков покрыли большую часть рынка, только языковой прозрачности и возможности общаться на любых языках, будет недостаточно для тех, кто торгует этой технологией.

Следующим достижением будет «межкультурная и транснациональная прозрачность». Когда другие аспекты человеческого общения, коммерции и взаимодействия, а не только язык, выйдут на сцену.

Например, жесты и мимика у разных народов имеют разное значение. В США, когда мы соединяем большой и указательный пальцы в виде буквы О, это означает «прекрасно», а в Аргентине этот же жест имеет непристойное значение.

Когда получат большее развитие многоязычные мультимедийные видеоконференции, нужно будет корректировать и жестикуляцию в изображении. Microsoft, лаборатория Массачусетского института технологий, и другие компании работают над распознаванием мимики и идентификацией биометрических характеристик по выражению лица.

Жаль, если американский бизнесмен, проведя блестящую презентацию в режиме видеоконференции для аргентинских бизнесменов, которая прекрасно переводится на аргентинский вариант испанского языка, вдруг сделает этот жест. Компьютеры смогут вмешаться и исправить в изображении его оплошность.

(Все цитаты в этой главе взяты из серии публикаций Entretiens, изданных Университетом Торонто под названием Net des etudes francaises — <http://www.etudes.francaises.net>).

Полезные ссылки

Перевод в киберпространстве

Американский институт национальных стандартов (ANSI)
<http://www.ansi.org>

ASCII (Американский стандартный код для международного обмена информацией) <http://www.asciitable.com/>

Языки, находящиеся в опасности
<http://www.yourdictionary.com/elr/>

Европейская ассоциация машинного перевода (EAMT)
<http://www.lim.nl/eamt/>

Языки европейских меньшинств
<http://www.smo.uhi.ac.uk/saoghal/mion-chanain/en/>

Объединение автоматизированного перевода (GETA)
<http://www-clips.imag.fr/geta/>

Исследования IBM <http://www.research.ibm.com/>

Медиа лаборатория <http://www.media.cit.edu/>

Группа естественных языков ISI
<http://www.isi.edu/natural-language/nlp-at-isi.html>

Opera <http://www.opera.com/>

ScanSoft <http://www.scansoft.com/>

Softissimo <http://www.softissimo.com/>

Systran <http://www.systransoft.com/>

TMX <http://www.tmx.de/>

Trados <http://www.trados.com/>

UNDL <http://www.undl.org/>

Unicode <http://www.unicode.org/>

Windows (Microsoft) <http://www.microsoft.com/windows/>

Wordfast <http://www.wordfast.org/>
yourDictionary.com <http://www.yourdictionary.com/>

Глобальная Интернет-статистика по языкам

<http://www.gleach.com/globstats/>

Многоязычие и цифровые документы

<http://www.info.unicaen.fr/bnum/jelec/Solaris/d06/61/61oupy.html>
<http://www.afla-redi.org/gic/monterrey.asp>
<http://www.iecat.net/inici.htm>
<http://www.iec.es/recerca/jornades/ciberespai/inici.htm>
<http://www.star-ag.ch/eng/aktuelles/links.html>

Статистика международных переводов

http://www.unesco.org/culture/xtrans/html_eng/index1.shtml

Культурные политики и многоязычие

<http://www.tlfq.ulaval.ca/axl/index.shtml>
<http://www.unesco.org/most/ln2lin.htm>

Обучение языкам

<http://www.frenchteachers.org/technology/cyberspace.htm>
<http://www.becta.org.uk/>

Термины ИКТ в Вашем местном языке

<http://www.becta.org.uk/technology/sayit/index.html>
(Албанский, арабский, бенгальский, вьетнамский, греческий, гуджарати, испанский, итальянский, йоруба, кантонезийский, корейский, курдский, немецкий, пашто, португальский, пунжаби, русский, сербский, сомали, суахили, тамильский, турецкий, украинский, урду, уэльш, фарси, хинди, японский)

Помощь при работе с местными языками в Интернете

Индия: <http://www.indianlanguages.com/>

Интернационализация формата HTML для создания Интернет-страниц на местном языке

<http://www.webreference.com/dlab/books/html/39-0.html>

Шрифты в киберпространстве

<http://www.sil.org/computing/fonts/>

Списки кодов ISP по странам

<http://thelist.internet.com/countrycode.html>