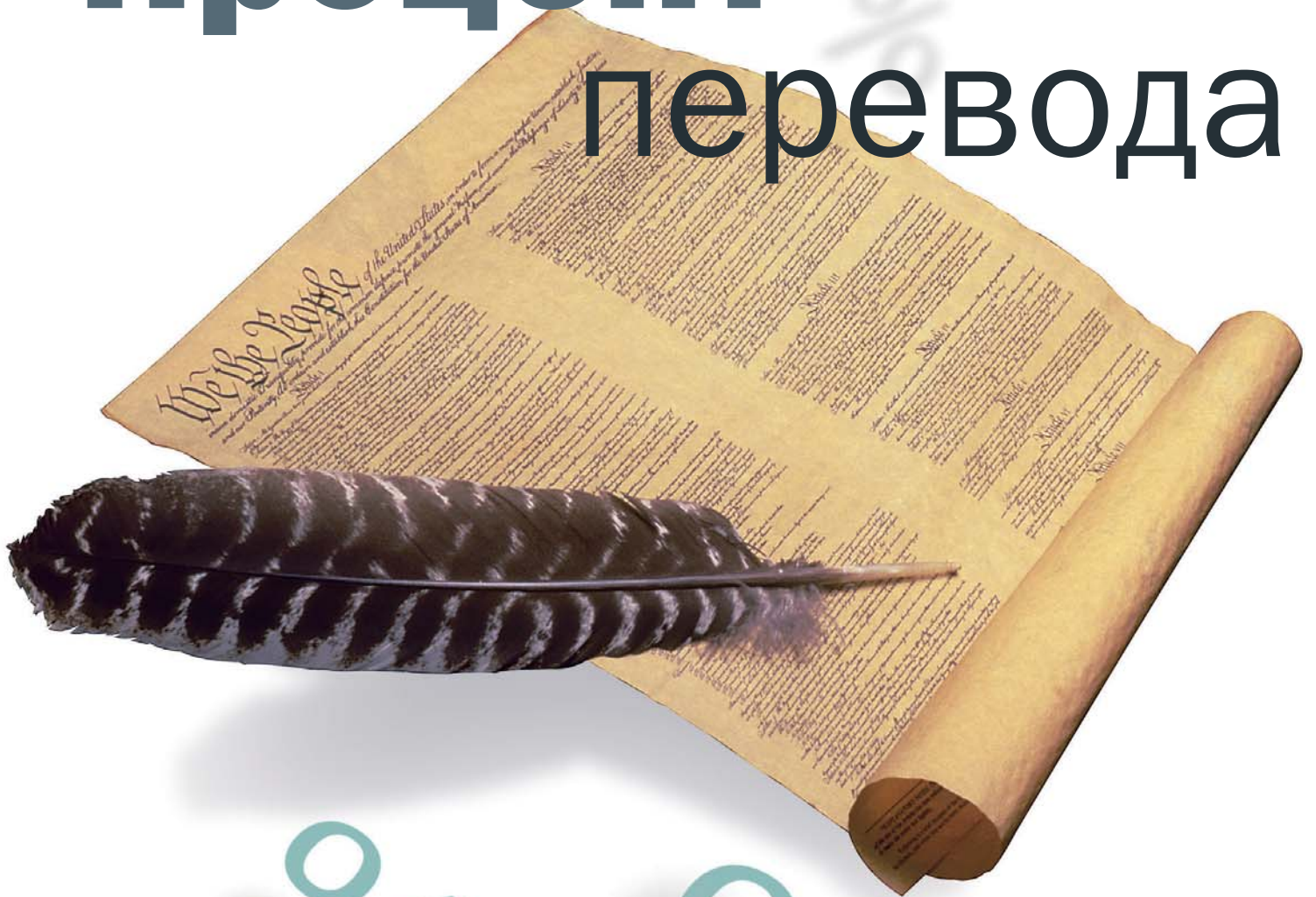


Процент перевода



Обзор программ-переводчиков

В последнее время все чаще можно услышать мнение, что машинный перевод — это такая же утопия, как и вечный двигатель. Красивая мечта, не более того. Однако это вряд ли соответствует истине, ведь тогда получается, что разработчики программного обеспечения продают нам то, чего нет и никогда не будет.

Полтора года назад мы впервые обратились к проблеме машинного перевода в статье «Экзамен по английскому» (Chip №12 2001), сопоставив четыре системы, поддерживающие русский язык: PROMT, Socrat, Retrans Vista и Pragma. Ни одна из них не смогла тогда адекватно перевести небольшой английский текст о мифологии.

Разработчики, однако, не топтались на месте, одна версия переводчиков сменяла другую, регулярно пополнялись словарные базы. Кроме того, на рынок начали выходить знаменитые зарубежные системы Systran и LogoMedia. Поэтому мы решили повторить состязание машинных переводчиков, но уже в новом, расширенном составе и с усложненными правилами. На переэкзаменовке им придется перевести не один, а три текста: фрагмент публицистической статьи, взятой с сайта американского журнала Newsweek, страницу из описания эмулятора Wine и рецензии программ-антивирусов с одного из англоязычных download-серверов.

При выборе тестовых заданий мы исходили из того, что именно в этих трех случаях пользователи, не владеющие английским языком, чаще всего обращаются за помощью к компьютерным переводчикам.

@prompt Professional

Популярная программа PROMT производства одноименной компании из Санкт-Петербурга регулярно обновляется. Буквально каждый год на рынке появляется новая версия. В 2002 году широкой общественности была представлена PROMT XT, а в нынешнем — @prompt Professional.

С каждой новой версией переводчик, обрстая полезными и не очень полезными функциями, становится все объемнее. @prompt полной инсталляции заняла около 200 Мбайт дискового пространства, львиная доля которого ушла на инсталляцию словаря Electronic Dictionary. От него можно и отказаться, но, как говорится, нет худа без добра: при окончательной доработке текста переводов теперь не нужно прибегать к помощи электронных словарей других разработчиков.

Из технологических новшеств PROMT последних версий отметим внедрение

технологии «ассоциированной памяти» (AM), позволяющей улучшить качество перевода массы сходных между собой документов. Понятно, что за основу AM питерские разработчики взяли популярную технологию Translation Memory. Машина в этом случае переводит только то, что ей до сих пор не было известно. Ранее же переведенные конструкции, встретившиеся в новом тексте, просто извлекаются ею из сформированных баз данных, которые, соответственно, каждый раз пополняются. В @prompt любой текст или даже отдельный его фрагмент можно добавить в AM буквально двумя кликами мыши («AM → Add Selected Text»).

Другой эксклюзивной функцией переводчика, безусловно, является встраивание его в виде плагина в Adobe Acrobat Reader 4 и 5 версий. Активизация этой опции осуществляется с помощью программы PROMT Plug-In for Acrobat. Данный плагин, как и другие компоненты @prompt, позволяет перевести текст с элементами интерактивности, с его помощью можно задать нужную те-

матику, выбрать подходящие словари и зарезервировать отдельные слова и словосочетания.

Во многом благодаря именно такой тщательной предварительной настройке и мощным встроенным тематическим словарям PROMT сумела-таки обойти победительницу прошлого соревнования, Retrans Vista, и занять первое место.

В первом задании, фрагменте статьи из журнала Newsweek, @prompt абсолютно точно сумела воспроизвести русские эквиваленты имен и фамилий политических деятелей США и Великобритании. Причем не только Джорджа Буша и Тони Блэра, но даже директора ЦРУ Джорджа Теннета. И это несмотря на то, что @prompt неверно определила тему статьи как «Телекоммуникации».

Второе задание, перевод страницы из руководства пользователя эмулятора Wine, системы — лидеры обзора выполнили примерно одинаково. Сложности у них возникли разве что с его названием. Название Wine, естественно, все по умолчанию перевели как «вино». »



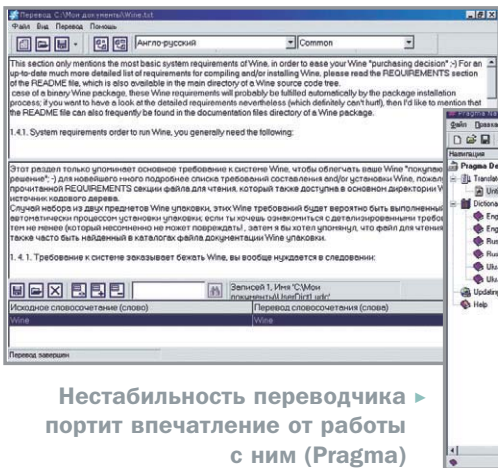
Беспристрастный судья

Один из способов добиться объективных оценок при тестировании переводчиков — использование программы Corvet, разработанной компанией «ПРОМТ». Она выносит суждение о качестве того или иного перевода практически полностью в автоматическом режиме. От пользователя требуется только предложить ей для анализа три текста: исходный на английском языке («File → Open → Source Text»), его машинный перевод («Automated Translation») и «правильный» эквивалент на русском. После того как вы кликните по кнопке «Evaluate» или нажмете клавишу F7, Corvet их сравнит и выдаст в процентах коэффициент несоответствия. Чем выше он будет, тем хуже машинный перевод. Но, откровенно говоря, мы не стали целиком и полностью доверять мнению этой программы, потому что невозможно подобрать или создать буквальный и при этом безукоризненный эквивалент английского текста. Из-за несоответствия грамматического строя различных языков перевод, сделанный человеком, получается почти

всегда хоть немного, но вольным. В свою очередь, системы машинного перевода, включая @prompt, нередко предлагают два и более варианта для одной словоформы, увеличивая тем самым вероятность погрешности.

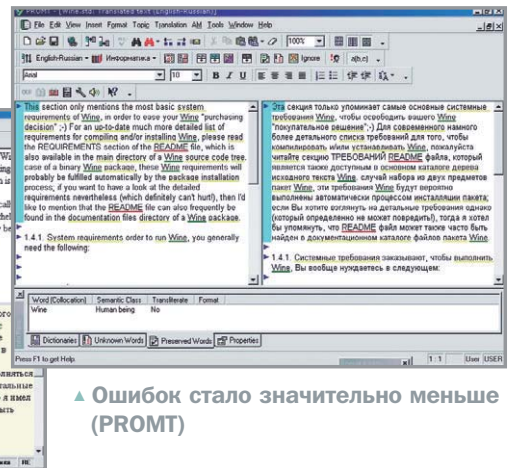
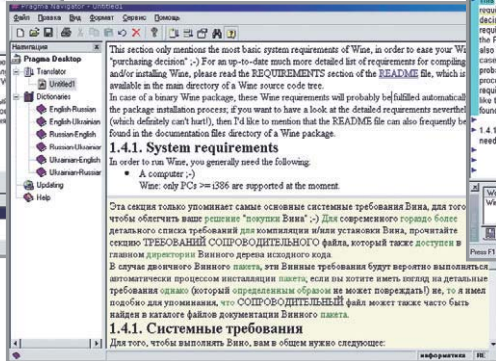
Использовать Corvet мы решили только для сравнения двух главных на сегодня конкурентов, @prompt и Retrans Vista, предложив им перевести рекламные проспекты друг друга, которые были приложены к дистрибутивам. У кого коэффициент несоответствия на «чужом» тексте относительно варианта конкурента будет меньше, тот и победил.

Сразу скажу, что выиграла опять-таки @prompt. У нее коэффициент оказался 60%, в то время как у Retrans Vista — 69%. На всякий случай мы предложили им перевести текст пояснения к одной программе, имеющий русский вариант. Результат: у @prompt — 93% несоответствия, у Retrans Vista — 94%. Столько же, кстати, у идущей третьей в нашем состязании украинской системы Pragma и, как ни странно, у Systran.



Нестабильность переводчика портит впечатление от работы с ним (Pragma)

Попробуйте интерактивный режим перевода (Retrans Vista)



Ошибок стало значительно меньше (PRoMT)

» Другое дело, что в одних случаях эту ошибку можно было без труда исправить, а в других нет. В @promt, например, достаточно было выделить мышкой «Wine» и переместить его в поле зарезервированных слов (Reserved Words).

В третьем тексте с описанием бесплатных антивирусов нам пришлось отметить значительно больше не подлежащих переводу словоформ. И все равно ошибок в терминах даже при подключении словаря по информатике было немало. Так, например, компьютерных червей («worms») @promt упорно именуется «тире».

Retrans Vista 1.0

Отстав буквально на полшага от @promt, второе место заняла московская система Retrans Vista. Секрет успеха детища маленькой команды из Vista Technologies прост: в основе лежит оригинальная и, на мой взгляд, наиболее перспективная теория семантического, преимущественно фразеологического перевода профессора Г. Белоногова. Согласно ей, основной единицей языка является не слово, а понятие, которое может быть выражено как отдельным словом, так и целым словосочетанием. Для получения адекватного перевода достаточно набрать некую критическую массу подобных понятий. Миллионов 15–20, как предполагает Г. Белоногов.

В настоящее время каждое из направлений перевода, русско-английское и англо-русское, в Retrans Vista включает примерно по полтора миллиона смысловых единиц такого рода.

Может быть, именно поэтому качество получаемых на выходе текстов пока

еще не радует. Особенно много ошибок Retrans Vista сделала при переводе статьи из Newsweek. Президента США Буша, к примеру, эта программа назвала «кустом» (Systran, впрочем, и вовсе «кустиком»). Когда же мы добавили в ее словарь правильное соответствие «Bush = Буш», почему-то в первом же предложении программа использовала женские окончания для этой словоформы. Вместо «Буш с вызовом настояя» получилось «Буша с вызовом настояя».

В общем, не быть Retrans Vista в этот раз среди лидеров, если бы не предусмотренный в ней интерактивный «Перевод с редактированием промежуточных результатов». В этом режиме программа предварительно показывает в виде таблицы исходные английские слова и словосочетания и, соответственно, их русские эквиваленты. Кликнув правой кнопкой мыши по любому из них, вы сможете выбрать наиболее подходящий вариант из имеющихся или предложить свой собственный («Добавить перевод»).

Особенно эффективен интерактивный режим при переводе небольших технических текстов. Та же аббревиатура Wine исправляется мгновенно. При автоматическом же режиме пришлось бы добавить новый вариант перевода «Wine = Wine» и сохранить его в пользовательском словаре. К сожалению, зарезервировать слова в Retrans Vista нельзя.

Злую шутку с «преимущественно фразеологическим» переводчиком в ряде случаев сыграл и его собственный метод. Например, в тексте о бесплатных антивирусах Retrans Vista не справилась

с термином «Home Edition». В принципе, не смогли этого сделать и конкуренты, но только Retrans Vista умудрилась перевести данное словосочетание как «издание газеты для распространения в городе». А все потому, что она не умеет самостоятельно определять тематику текста, и подсказать ей нельзя, поскольку предусмотрено только пять готовых вариантов.

Графическая оболочка Retrans Vista — не верх совершенства. Неслучайно сами разработчики называют ее Tran Demo. С другой стороны, к ней можно и не обращаться, ведь Retrans Vista встраивается в качестве модуля в Internet Explorer и Microsoft Word, включая новейшую его версию из состава MS Office 2003 Server. Кроме того, к вашим услугам утилита быстрого перевода небольших текстов, скопированных в Буфер обмена, ClipTran.

Pragma 2.0

Если успех @promt и Retrans Vista трудно было предсказать, то удачное выступление прежнего аутсайдера, украинской системы Pragma, стало приятной неожиданностью. Возможно, она заняла бы и более высокое место, если бы была представлена полной версией. Демонстрационный же вариант (который через две недели начинает напоминать о себе робким сообщением о желательности регистрации) работал не всегда стабильно.

Нельзя в демоверсии и ничего добавить в пользовательский словарь. В результате словосочетание «in order to run Wine» («для того чтобы запустить Wine») Pragma перевела весьма ориги-

» нально: «заказывают, чтобы двинуть Вино». Исправить эту ошибку нам не удалось даже с помощью режима полусловного перевода «Ассистент» («Сервис —> Ассистент»). При его активизации Pragma наглухо зависла.

С другой стороны, количество непереведенных или переведенных неверно слов в других заданиях оказалось у украинского переводчика даже меньше, чем у той же Retrans Vista. Чтобы еще немного повысить качество перевода, можно попробовать снять режим автоматического определения темы текста в настройках и выбрать ее самостоятельно из 95 предлагаемых вариантов.

Но и без этого трактовки Pragma были порой даже удачнее, чем у победителя @prompt. Например, фразу «free complete ICSA certified antivirus package for home noncommercial use» Pragma перевела следующим образом: «свободный полный удостоверенный пакет антивируса ICSA для домашнего некоммерческого использования». Ошиблась она только в интерпретации слова «complete» как прилагательного «полный», а не наречия «полностью». А вот у @prompt перевод получился далеко не таким гладким: «свободный законченный ICSA удостоверенный пакет для домашнего некоммерческого использования».

Выгодно отличается Pragma от конкурентов и количеством (более 20) поддерживаемых форматов файлов. В этом отношении она опять-таки сумела превзойти @prompt. Как и в первой версии, ее можно встроить не только в Microsoft Word и Internet Explorer, но и в Проводник Windows, почтовый клиент Outlook Express, редактор WordPad и даже в обычный Блокнот. При этом системные требования у Pragma существенно ниже, чем у @prompt и Retrans Vista.

Systran 4.0 Premium

Провал старейшины систем машинного перевода, американской программы-переводчика Systran, оказался для нас неприятным сюрпризом. В ее пользу говорили несомненный авторитет среди лингвистов всего мира и очень высокая цена. Можно сказать, что это главная сенсация проведенного соревнования со знаком минус. Мы ожи-

дали, что именно Systran составит основную конкуренцию @prompt и Retrans Vista.

К сожалению, Systran допустила больше ошибок, чем конкуренты, и оставила намного больше непереведенных слов. Только по тексту из Newsweek мы насчитали 11 нераспознанных ей словоформ. В то время как у @prompt и Retrans Vista их было всего по три.

Возможно, это связано с тем, что англо-русское направление недавно появилось в коллекции Systran. Возможно, разработчики не уделяют ему должного внимания. Остается надеяться, что в последующих версиях если не все, то

хотя бы самые важные недочеты будут устранены. Для начала Systran неплохо было бы отучиться от дурной привычки придумывать слова и формы, которых в русском языке нет и быть не может, чтобы она не писала больше «великобританский» вместо «британский», «счесн» вместо «найден», «зарегистрирование» вместо «регистрация» и т. п.

В компьютерных текстах перевод у Systran получился чуть лучше, но и в них без откровенных ляпов вроде «смогите download» все же не обошлось.

Не слишком удобен и интерфейс основной программы пакета — Translation Project Manager. Для того чтобы заре-



Особенности перевода с немецкого

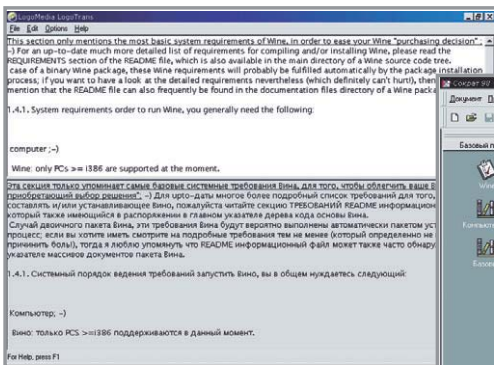
Взаимопонимание с Европой

Немецкоязычная информация является одним из столпов мирового бизнеса. В частности, головные офисы компаний Ahead, BMW, Volkswagen, Opel, Deimler Benz, Siemens и т. д. находятся в Германии, и немецкий язык является родным для их специалистов. Основные трудности при машинном переводе немецкого текста состоят в следующем:

- ▶ Развитое словоизменение. Количество словоформ, конечно, меньше, чем в русском, но все равно довольно велико. Положение осложняет еще и то, что существует довольно много слов, образование форм которых является исключением из правил, типичный пример тому — спряжение сильных глаголов.
- ▶ Большое количество диалектов. Особенно сильно отличаются от литературного немецкого баварский, швейцарский и тирольский диалекты. И при этом на них активно издаются книги и периодические издания, существуют диалектальный театр, средства массовой информации, а также довольно много ресурсов в Интернете.
- ▶ Активное употребление аббревиатур и сложносокращенных слов вроде U-Bahn (метро), S-Bahn (городская электричка), U-Boot (подводная лодка).
- ▶ Широкое распространение очень сложных составных слов. Именно эта особенность доставляет наибольшие сложности, так как требует определенного интеллекта по их разбору.

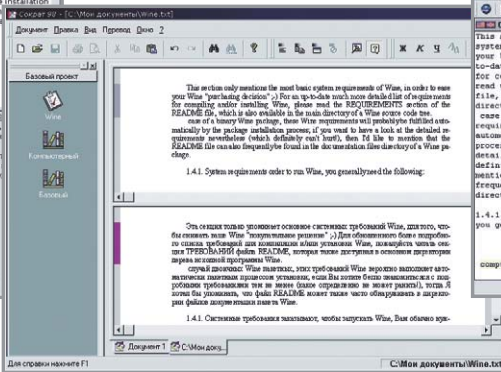
Борьба с этими трудностями длилась долго, в результате чего немецкий язык поддерживает не так много программ, а русско-немецкую пару и вовсе две — Socrat и PROMT. Работа первых версий этих программ вызвала множество нареканий. Однако время шло, и переводчики совершенствовались. Результаты тестирования немецко-русского перевода последних версий Socrat и PROMT позволили сделать следующие выводы:

- ▶ Процесс перевода немецкого текста идет в среднем в полтора раза медленнее, чем английского.
- ▶ Многие диалектизмы и языковые особенности разных стран не нашли своего отражения в базах программ.
- ▶ Не всегда корректно обрабатываются глаголы с отделяемыми приставками и возвратными частицами, что ведет к появлению ошибок, искажающих исходный смысл документа.
- ▶ Общий смысл документа на литературном немецком языке, переведенного с помощью как Socrat, так и PROMT, понять можно. Следует, правда, отметить, что вторая программа переводит более качественно. Естественно, если в тексте нет специфической лексики, игры слов, а авторы не злоупотребляют диалектизмами.
- ▶ Немецкое направление перевода в обеих программах несколько уступает по качеству английскому, и получаемый текст нуждается в серьезном редактировании.



Мы ожидали от этой системы > большего (Socrat 98 Professional)

«Для тех, кому важна скорость (LogoMedia)



▲ Неплохой переводчик из прошлого (Systran 4.0 Premium)

» зривировать в нем не требующие перевода слова, необходимо сначала отметить в настройках опцию «Do not Translate Capitalized Words» и только потом выбрать их с помощью контекстного меню в документе («Set DNT block»).

Помимо Translation Project Manager, в состав набора входят также удобная панель быстрого запуска Clipboard Taskbar, утилита пакетного перевода MultiTranslate, редактор словарей Dictionary Manager и резидентная программа Systran Server. Кроме того, переводчик интегрируется в Microsoft Word.

LogoMedia

Вторая участница состязания из США, программа LogoMedia, выступила ни-

чуть не лучше первой, также не сумев перевести 11 слов в тексте из Newsweek.

Впрочем, трудно было ожидать от нее другого результата. Интерфейс главной программы этого пакета, LogoTrans, выглядит просто миниатюрным по сравнению с @prompt. Судя по всему, LogoMedia предназначена для быстрого, а не точного перевода.

При открытии текстового файла в нижнем окне мгновенно появляется перевод. При желании можно подключить 26 дополнительных англо-русских или, соответственно, столько же русско-английских словарей. Их тематический состав напоминает «промотовский». Для активизации словарей перейдите в меню «Shared Translation

Options» («Edit → Show List Of Dictionaries») и там щелчком по кнопке «Load» откройте диалоговое окно выбора.

Грамотное подключение дополнительных словарных баз позволит вам существенно повысить качество перевода. Вот, скажем, оборот «according to a White House source» без дополнительных словарей LogoMedia переводит так: «в соответствии с Белым источником дома». Согласитесь, получается полная бессмыслица. С подключенным же словарем Sociology все встает на свои места: «в соответствии с источником Белого дома».

В отличие от @prompt в LogoMedia нельзя зарезервировать отдельные слова и конструкции или хотя бы пополнить »

	@prompt	Retrans Vista
Сайт разработчика	www.prompt.ru	www.advantis.ru
Условия распространения	commercial	shareware
Цена, \$	259	от 15
Язык интерфейса	англ.*	рус.
Дополнительные словари (англо-русские)	16**	4
Предварительная настройка	Резервирование слов, выбор тематики текста, технология «ассоциированной памяти»	Интерактивный режим перевода, словарь пользователя
Количество непереведенных слов в тестовом задании	3	3
Время доработки текста первого тестового задания, мин.	8	11
Место на HDD после инсталляции, Мбайт	185	150
Объем занимаемой оперативной памяти, Мбайт	7,4	3,8
Открываемые для чтения форматы файлов	RTF, TXT, DOC, HTML, DOT	TXT, RTF
Интеграция в приложения	Internet Explorer, Microsoft Word, Adobe Acrobat Reader	Internet Explorer, Microsoft Word
Дополнительные возможности	Словарь Electronic Dictionary, пакетный переводчик File Translator, озвучивание текста***	Утилита быстрого перевода ClipTran
Плюсы	Широкий спектр дополнительных словарей, применение технологии «ассоциированной памяти», возможность встраивания в программу Adobe Acrobat Reader	Фразеологический метод обработки текста, наличие интерактивного режима работы
Минусы	Сравнительно высокие системные требования	Всего пять готовых тем перевода, неудобная графическая оболочка

* В тестируемой версии ** Полный список словарей см. на сайте разработчиков *** При наличии голосовых библиотек и движков **** В полной версии

» ими пользовательский словарь. Поэтому компьютерные, двуязычные по определению тексты эта программа перевела не лучшим образом. Забавных несуразниц было немало. «Вино готовится для других водителей показа графики», — так LogoMedia перевела фразу: «Wine is prepared for other graphics display drivers» («Wine подходит для других драйверов графических дисплеев»).

Настолько низкое качество перевода вряд ли компенсирует то обстоятельство, что LogoMedia может встраиваться в Microsoft Word и включает в себя программы быстрого и пакетного перевода (TransIt и FileTrans).

Socrat 98 Professional

Программа Socrat Professional Санкт-петербургской компании «Арсеналь» выступила в нашем состязании вне конкурса. Все-таки с 1998 года она не развивается, и никаких конкретных планов относительно ее дальнейшей судьбы у разработчиков в настоящее время нет.

Очень жаль, если этот проект будет заморожен, ведь профессиональный «Сократ», в отличие от более известного широкой публике Socrat Personal, позволяет подключать дополнительные словари. Два из них, «Компьютерный» и «Бизнес», имеются даже в его свободно рас-

пространяемой демонстрационной версии, в которой нельзя сохранить перевод.

Как ни странно, результаты у Socrat Professional оказались не намного хуже, чем у новейших версий @prompt, Retrans Vista и Pragma, и даже лучше, чем у System и LogoMedia. И это несмотря на то, что в демоверсии невозможно резервировать слова.

Отдельные же конструкции Socrat Professional распознала точнее всех современных конкурентов, включая лидера состязания. Например, фразу «You will obtain the free license key by E-mail after the registration...» она перевела практически безошибочно: «Вы получите свободный лицензионный ключ Эл. почтой после регистрации». Непонятно только, почему Socrat дает сокращение для слова «электронной», да еще с заглавной буквы. @prompt этой ошибки не допустила, однако предложила при этом не один, а два варианта толкования слова «key»: правильный — «ключ» и неверный в данном случае — «клавиша».

Ошибки у Socrat, конечно, тоже были, но не такие грубые, как можно было бы ожидать от программы пятилетней давности. Например, она не сумела перевести достаточно сложное выражение «right click on the object allows» («щелчок правой кнопкой по объекту позволяет»).

В трактовке Socrat это звучит так: «щелчок права в объекте позволяет». Здесь неверно определено как часть речи слово «right». Существительным в этом значении оно никак не может быть.

К недостаткам Socrat Professional можно отнести и то, что программа никуда не встраивается и существует только в своей оболочке. Зато она может переводить не только TXT- и RTF-файлы, но и DOC в версии Microsoft Office 6.0-7.0, и даже LX некогда знаменитого «Лексикона».

Отличники и хорошисты

Итак, дальше других за полтора года продвинулась вперед PROMT — безусловно, лучший продукт на сегодня. Во всяком случае, ошибок стало значительно меньше, чем прежде. Буквально в затылок ей дышат многочисленные конкуренты, прежде всего — пришедшая к финишу второй Retrans Vista. С учетом особого интерактивного режима перевода она является оптимальным выбором, по крайней мере, для тех пользователей, которым необходимо одновременно быстро и качественно переводить тексты с английского на русский и особенно с русского на английский. Второе, кстати, получается у систем машинного перевода лучше.

■ ■ ■ Александр Евдокимов,
Филипп Рудых

Pragma	Systran	LogoMedia	Socrat Professional
http://web.lucky.net/~trident/index.html	www.systransoft.com	www.logomedia.net	www.ars.ru
shareware	commercial	shareware	shareware
180	759	199,95	100
рус., англ., укр.	англ.	англ.	рус.
—	5	26	11****
Словарь пользователя, пословный режим перевода «Ассистент»	Резервирование слов	—	Резервирование слов, пользовательский словарь
5	11	11	9
12	17	15	11
15	72	115	4
0,9	1,5	1,5	5,1
HTML, TXT, RTF, DOC, WPD, XLS, ANS, ASC, WPS, WK 1,3,4 и др.	RTF, TXT, DOC	TXT	RTF, TXT, DOC (Word 6.0-7.0), LX, WR
Internet Explorer, Microsoft Word, Word Pad, Проводник, Блокнот	Microsoft Word	Microsoft Word	—
Панель быстрого запуска Clipboard Taskbar, пакетный переводчик MultiTranslate	—	Программа быстрого перевода TransIt, пакетный переводчик FileTrans	—
Большой набор дополнительных словарей, интегрируется во множество приложений, поддерживает более 20 форматов файлов	Удобная панель быстрого запуска	Мобильный переводчик, много дополнительных словарей	Недорогая программа, понимает LX-файлы
Не всегда стабильно работает, может конфликтовать с другим ПО	Самый дорогой переводчик. Оставляет много непереведенных слов, создает несуществующие формы русских слов	Нет предварительной настройки, корректно поддерживается только TXT-формат	—