УДК 681.3: 37.022

А.И.ИМАНАЛИЕВА
E-mail: ksucta@elcat.kg.

АВТОМАТИЧЕСКИЙ (МАШИННЫЙ) ПЕРЕВОД ТЕКСТА

Бул макала машина аркылуу которууну изилдөөгө арналат.

Эта статья посвящена исследованию машинного перевода текста и его дальнейшего развития.

This article is devoted to peculiarities of machine translation of the text, translation and its development.

Среди многих сложнейших проблем, изучаемым современным языкознанием, важное место занимают перевод и переводческая деятельность. Перевод имеет долгую историю. Своими корнями он восходит к тем далеким временам, когда праязык начал распадаться на отдельные языки и возникла необходимость в людях, знавших несколько языков и способных выступать в роли посредников при общении представителей разных языковых общин.

Переводом называется процесс и результат создания на основе исходного текста на одном языке равноценного ему в коммуникативном отношении текста на другом языке. При этом коммуникативная равноценность, или эквивалентность, понимается как такое качество текста перевода, которое позволяет ему выступать в процессе общения носителей разных языков в качестве полноправной замены исходного текста в сфере действия языка перевода

В последнее время знание иностранных языков может понадобиться не только в путешествии или на приеме гостей из-за рубежа, не только на работе, но и в собственном доме, например, при просмотре голливудских кинолент, при чтении инструкции по использованию зарубежных товаров или Web-страниц. Таким образом, оказывается, что даже не покидая родных стен, мы нуждаемся в услугах переводчика. Но сейчас необходимую помощь нам вполне может оказать домашний компьютер.

Системы машинного перевода (МП) давно перестали быть новинкой. Они постепенно выходят из младенческого возраста и вместо бессвязного детского лепета начинают изъясняться на вполне понятном, "человеческом" языке. До последнего времени такие программы были не только очень дороги, уступая в цене разве что мощным графическим и издательским системам, но и весьма сложны и капризны в работе. И вот появились первые переводчики, пригодные для использования на домашнем ПК. Стремительные потоки информационного обмена между высокоразвитыми промышленными странами, лавина научно-технической документации, поступающая от производителей товаров и современных технологий, требуют совершенно нового подхода к проблеме перевода технической литературы. Выход один: максимально

автоматизировать процесс, оставив человеку его творческую редакционную часть. В этом помогает система машинного перевода. Ее параметры должны удовлетворять четырем основным требованиям:

- оперативность;
- гибкость;
- скорость;
- точность.

Оперативность машинных систем – это возможность постоянного пополнения словарного запаса и создания новых тематических словарей. В этом параметре они значительно опережают привычные типографские издания различных словарей.

Гибкость – это возможность *"грубой настройки"* на конкретную предметную область (для этой цели служат специализированные словари) и *"тонкой настройки"* на конкретный текст, книгу или группу документов (модифицируемые пользовательские словари).

Скорость – возможность автоматического ввода и обработки текстовой информации с бумажных носителей. Только одна система оптического ввода текстов *(OCR-System)* ежедневно заменяет более десяти классных машинисток.

Точность – стилистически и грамматически правильная адекватная передача смысла исходного текста на язык перевода. Это наиболее *"уязвимое"* место систем машинного перевода. Однако столь явное улучшение качества перевода в поздних версиях систем машинного перевода, как, например, PROMT, вселяет уверенность, что вскоре компьютер полностью примет на себя всю рутинную часть перевода. Во-первых, всем ясно, что чем больше словарь, тем лучше перевод, значит, первая проблема – проблема создания больших словарей для систем.

Во-вторых, ясно, что система должна переводить такие предложения: «Привет, как дела?». Значит, еще одна проблема – научить систему распознавать устойчивые обороты.

В-третьих, понятно, что предложение для перевода пишется по определенным правилам, по определенным правилам переводится, а значит, есть еще одна проблема: записать все эти правила в виде программы. Вот, собственно, и все.

Самое интересное, что эти проблемы действительно являются основными при разработке систем МП, другое дело, что методы их решения известны далеко не всем и отнюдь не так просты, как может показаться.

Системы МП семейства PROMT (PROgrammer's Machine Translation) – очень хороший объект, чтобы продемонстрировать, каким образом эти проблемы могут решаться эффективно.

Для качественного перевода очень важно, чтобы практически все слова исходного текста легко было найти и в словаре системы. А те из них, которых в нем нет, переносятся в текст непереведенными уже на выходе из системы, и их впоследствии переводят вручную при редактировании результатов перевода. Такие слова могут повлиять на качество перевода предложения. Дело в том, что для определения, к какой части речи относится рассматриваемое

слово, система производит анализ всего предложения в целом. При этом имитируется мыслительная деятельность человека (такую систему принято называть системой с элементами искусственного интеллекта). Если значение хотя бы одного слова в предложении не определено, то это может исказить анализ всего предложения, а иногда и результаты всего перевода.

Методы организации больших баз данных достаточно хорошо разработаны, но для перевода не менее, а может быть, и более важно правильно структурировать информацию, которая приписывается элементу базы, правильно выбрать этот самый элемент. Сколько, например, записей в словаре должно соответствовать обыкновенному русскому слову "программа"? Например, существительные в русском языке изменяются по падежам и по числам, то есть для одного существительного может существовать до 12 разных форм, а для глаголов и прилагательных, как правило, существует еще большее количество различных форм (более тридцати). Следовательно, чтобы переводить предложения, содержащие слова "программу", "программе", "программы" и т.д., хорошо было бы иметь способ соотнесения словарной статьи из автоматического словаря для слова "программа" со многими словарными статтьями, или словарь, который позволяет распознать много слов из текста.

Поэтому для описания и входного и выходного языка в системе должен существовать некоторый формальный метод описания морфологии, на котором основывается выбор единицы словаря.

В системах семейства PROMT разработано практически уникальное по полноте морфологическое описание для всех языков, с которыми системы умеют обращаться. Оно содержит 800 типов словоизменений для русского языка, более 300 типов как для немецкого, так и для французского языка, и даже для английского, который не принадлежит к флективным языкам, выделено более 250 типов словоизменений. Множество окончаний для каждого языка хранится в виде древесных структур, что обеспечивает не только эффективный способ хранения, но и эффективный алгоритм морфологического анализа.

Кроме того, используемая модель морфологии позволила разработать экспертную систему для пользователя – создателя словаря. Эта система фактически автоматизирует процедуру выделения основы и определения типа словоизменения при вводе новых словарных статей.

Однако разработка описания морфологии позволяет решить только проблему того, что является заголовком словарной статьи, по которому происходит идентификация единицы текста и единицы словаря. Но ведь идентификация слова из текста со словарной статьей происходит не ради идентификации, как это требуется в электронных словарях, она необходима для выполнения программой собственно процедур перевода. Какая же нужна информация в словарной статье и как должны быть описаны правила перевода для того, чтобы программа переводила?

С развитием МП как области прикладной лингвистики появилось множество лингвистических работ, предлагавших структуру описания свойств живого слова в словарной статье машинного словаря. При этом совершенно отдельно появлялись исследования,

описывающие, например, "структуру именной группы" или "способы выражения прямого дополнения для глаголов говорения".

Например, на основе признака "принадлежность к части речи" описывалась грамматика такого типа:

• именная группа – это существительное;

• именная группа – это прилагательное + именная группа;

• глагольная группа – это глагол + именная группа;

• предложение – это именная группа + глагольная группа.

Понятно, что некоторая часть предложений естественного языка описывается такой грамматикой, но эта часть очень незначительна, и на ее основе нельзя правильно анализировать и переводить хоть сколько-нибудь реальный текст. Но зато можно использовать эффективные методы построения преобразователя по заданной грамматике или, на худой конец, написать программу, которая путем перебора построит деревья зависимостей для ограниченного множества предложений.

Хотелось бы надеяться, что эти сведения позволят потенциальным пользователям систем перевода понять, что создание системы МП – задача не такая уж простая и, что называется, наукоемкая. А следовательно, количество действительно пригодных к использованию систем перевода, которое может появляться в единицу времени, принципиально ограничено.

В любом случае, стилистические и грамматические огрехи машинного перевода компенсируются потрясающей скоростью получения его чернового варианта.

Применение машинного перевода без настройки на тематику служит предметом многочисленных бродящих по Интернету шуток. Один из примеров известен, это текст «Гуртовщики Мыши» (перевод компьютерной документации программой *Poliglossum* на основе медицинского, коммерческого и юридического словарей); из кратких – фраза *«My cat has given birth to four kittens, two yellow, one white and one black»,* которую переводчик компании ПРОМТ превращает в *«Моя кошка родила четырех котят, два желтых цвета, одного белого и одного афроамериканца».* Главной причиной того, почему программа перевела именно так, было то, что после слова *black* нужно было добавить *kitten*, тогда программа переведет правильно: *«Моя кошка родила четырех котят, двух желтых, одного белого и одного черного котенка».*

И у нас появляется вопрос, как же можно улучшить качество перевода? Чтобы ответить на этот вопрос, узнаем для начала несколько практических советов. Наверняка многие уже имеют опыт "общения" с системами машинного перевода. Кое-кто сумел уловить правильный подход к работе с этими программами и эффективно использует их. Другие, наоборот, после первого же сеанса работы испытали разочарование, оценив качество полученного текста. Но не стоит отчаиваться. Существуют способы улучшения результатов машинного перевода, доступные каждому пользователю. О некоторых из них мы сейчас и поговорим.

1. Исход работы в значительной мере решается еще до ее начала

Прежде чем приступить к переводу, обязательно определите две вещи: во-первых, для каких целей предполагается использовать его результаты, а во-вторых, что представляет собой исходный текст. Назначение перевода играет первостепенную роль при оценке его качества. В самом деле, один и тот же результат можно считать отличным, если нужно просто узнать, о чем идет речь в оригинальной статье, и совершенно непригодным, если нужно получить текст для публикации в книге или журнале. Но иногда даже самый "грубый" перевод оказывается приемлемым, если в нем имеется достаточно информации, по которой специалист в соответствующей предметной области может легко восстановить содержание текста. С другой стороны, определив, к какому стилю речи принадлежит исходный текст, нетрудно оценить его пригодность для машинного перевода, а значит, и предугадать результат. Чем больше в тексте иносказательных оборотов, метафор, чем свободнее стиль, тем хуже справится компьютер с его переводом. Лучше других обрабатываются научные, технические и образовательные тексты, которым присуще строгое изложение материала. Если своевременно пополнять специальные словари новыми терминами, то можно получать полностью связный перевод текстов, требующий минимальной стилистической доработки.

2. Бойтесь опечаток!

Очень часто причиной неправильного перевода являются опечатки в оригинале. В особенности это касается отсканированных и распознанных текстов. Слова с орфографическими ошибками в большинстве случаев помечаются системой как незнакомые, поскольку в исковерканном виде они в словарях отсутствуют. Сложнее, если опечатка превращает одно слово в другое, которое также существует в иностранном языке, – программа переведет его, но смысл текста будет искажен. Но самыми серьезными "подводными камнями" являются ошибки в пунктуации. Одна неправильно поставленная запятая способна серьезно исказить перевод предложения. Поэтому перед переводом как можно тщательнее проверьте исходный текст.

3. Хороший словарь – половина успеха

Обязательно найдите и подключите специальные словари по тематике переводимого текста. Если в точности такой тематики нет, определите наиболее подходящую комбинацию имеющихся у вас словарей и, кроме того, обязательно создавайте свои. Идеально, конечно, иметь для каждого текста свой словарик, но оптимальным, с точки зрения продуктивности работы, является разбиение наиболее распространенных тем на подпункты. Например, в рамках компьютерной тематики можно создать словари "Офисные программы", "Графика", "Сети", "Internet и WWW" и т.д.

4. Строим "пирамиду"

Если к системе подключено несколько словарей, то успех перевода во многом зависит от того, в каком порядке программа ищет в них текущее слово. Поэтому организуйте иерархию словарей в порядке от частного к общему. Самый высокий приоритет должен иметь словарь, созданный для текущего текста, затем – тематические (в порядке расширения предметной области), а самый низкий уровень остается за словарем общеупотребительных понятий. Так, при

переводе текста о программе Adobe Photoshop лучше всего на самом высоком уровне поставить словарь "Photoshop" (созданный вами специально для этого текста), затем – "Компьютерная графика", "Информатика", и в самом конце списка – общий словарь. Поскольку объем узкоспециализированных пользовательских словарей, как правило, невелик, а общего, наоборот, огромен, то полученная в результате "конструкция" в чем-то походит на пирамиду.

5. Не надо спешить!

Никогда не переводите сразу весь текст. В нем всегда найдется как минимум одно-два слова, отсутствующих в словарях, и превеликое множество таких, которые система переводит неправильно. Покончив с подключением словарей и определением других опций перевода, для начала переведите небольшой фрагмент в начале текста, например первый абзац. Найдите в этом фрагменте слова, переведенные неправильно, и внесите их в словарь самого высокого уровня. Переведите фрагмент заново. Если результат вас удовлетворит, переходите к следующему абзацу. Практика показывает, что для точной настройки системы необходимо перевести таким образом четверть, а иногда даже треть материала и только после этого запускать автоматическую обработку всего текста.

6. Различайте общее и частное!

Прежде чем внести новое слово в пользовательский словарь, определите, будет ли оно переводиться в тексте данным образом только один-два раза или постоянно. Искусство правильной работы с системой машинного перевода состоит в том, чтобы различать общее и частное. В идеальном случае от вас потребуется знание иностранного языка и предметной области текста. В словарь вносите только систематически встречающиеся варианты перевода, единичные же исправляйте вручную. В противном случае слово по всему тексту будет переведено неправильно.

Соблюдение приведенных выше простых правил обычно позволяет существенно повысить качество переведенных компьютером текстов. Нельзя гарантировать, что они всегда окажутся абсолютно правильными и стилистически грамотными. Однако, вероятнее всего, свою первую задачу – понять смысл текста – вы решите.

В ходе данной работы мы выяснили, что машинный перевод – это эффективное средство для просмотра и поиска информации на иностранном языке, и именно эта функция является главной при работе в Internet. Далее, в результате настройки на предметную область и интеграции с другими программами обработки документов средство машинного перевода позволяет автоматизировать получение перевода. И, наконец, это уникальный гуманитарный инструмент, позволяющий преодолевать проблемы общения в системах, работающих на разных языках. И, пожалуй, самый главный, поистине революционный для прикладной лингвистики вывод состоит в том, что многие разработчики осознали: при создании программы машинного перевода, что кроме хорошо реализованной лингвистики необходима достойная программная реализация.

Несомненно, средства машинного перевода никогда не смогут улавливать все смысловые нюансы оригинального текста. Различия в синтаксисе и семантике, особенно между западными и

восточными языками, – скажем, английским и китайским – слишком велики для этого. Даже сторонники машинного перевода признают, что он способен в лучшем случае передать основную суть документа.

Но нельзя забывать, что ошибки часто случаются и у обычных переводчиков, и наивно ожидать от системы машинного перевода, что она способна исправить ошибки оригинала и выдаст грамотный перевод. Безусловно, тезис о том, что мозг человека совершенней компьютера, не требует никаких доказательств. Хотя, честно сказать, после общения на выставках с некоторыми пользователями в этом появляются сомнения.

В заключение хотелось бы подчеркнуть, что программа-переводчик – это, прежде всего, инструмент, который позволяет решить проблемы перевода или повысить эффективность труда переводчика только в том случае, если он используется грамотно.

СПИСОК ЛИТЕРАТУРЫ

1. Аристов Н.В. Основы перевода. – М.: Изд-во литер. на иностр. языках, 1959.
2. Бархударов Л.С. Язык и перевод. – М.: Международные отношения, 1975.
3. Ванников Ю.В. Языковая сложность текста как фактор трудности перевода: Методическое пособие. – М.: Всесоюзный центр переводов, 1988.
4. Винокуров А.А., Чуканов В.О. Новый метод оценки машинного перевода // Информационные технологии и системы. Hardware Software Security. Тенденции и перспективы: Сборник статей. – М.: Международная академия информатизации, 1997.
5. Система перевода текста PROMT Internet. Руководство пользователя. – СПб.: Фирма "ПРОМТ", 1999.
6. Скороходько Э.Ф. Вопросы перевода английской технической литературы. – К.: Изд-во Киевского университета, 1963.
7. Федоров А.В. Введение в теорию перевода. – М.: Изд-во литер. на иностр. языках, 1958.
8. Федоров А.В. Основы общей теории перевода. – М.: Высшая шк, 1968.
9. Флорин Сидер. Муки переводческие. – М.: Высш. школа, 1983.
10. Циммерман М.Г., Веденеева К.З. Русско-английский научно-технический словарь переводчика. – М.: Наука, 1991.
11. Черняховская Л.А. Перевод и смысловая структура. – М.: Международные отношения, 1976.
12. Читалина Н.А. Учитесь переводить (Лексические проблемы перевода). – М.: Международные отношения, 1975.
13. http://www.itc.kiev.ua/itc/dpk/archive/1999/02/software.shtml
14. http://www.promt.ru
15. http://www.socrat.ru

16. http://www.translate.ru