
Как улучшить качество машинного перевода?

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.

Разговорный и публицистический стиль, где много специфических оборотов, но большинство слов используется в прямом смысле, пригодны для ознакомительного перевода, однако для получения грамотного выходного текста потребуется ручная правка. А вот поэзия и художественная литература, к сожалению, совершенно непригодны для машинного перевода.

Смысл текста, построенного на иносказательных выражениях, при машинном переводе искажается и недоступен даже для ознакомления.

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 словарь, 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 общепотребительных 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 потребуются 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.