

# Мастер-класс PROMT – повышаем качество машинного перевода

Автоматизация любой сферы деятельности, где задействован человек, часто связана с необходимостью применения специальных инструментов и настроек, позволяющих улучшить или сделать более стабильным ее результат. В случае с машинным переводом текста такие операции нужно проводить всегда, когда речь идет о переводах стандартизированной документации – таким образом можно получить существенный выигрыш по времени. Посмотрим, какие средства используются для повышения качества переводных текстов в самой известной подобной системе [PROMT](#).

## Подготовка к работе

Обычно все настройки для улучшения качества перевода проводятся уже на конкретных примерах текстов, которые были получены в ходе работы **PROMT**. Поэтому все описываемые процедуры будут рассматриваться уже на переведенном тексте, качество которого мы будем последовательно улучшать. В контексте промышленного использования машинного переводчика с этой целью обычно используют выровненные тексты, то есть, переведенные построчно и по абзацам фрагменты. Они сохраняются в так называемые базы Translation Memory, которые можно подключать к машинному переводчику и получать перевод с сохранением терминологии и стилистики. Кроме того, при переводах может задействоваться пользовательский словарь, в который будет включена необходимая лексика.

## Настройки словаря

При совершении перевода в **PROMT** в правой части окна пользователь видит переведенный текст, размеченный специальным форматированием. Программа выделяет слова, оставшиеся без перевода, а также слова, имеющие несколько вариантов перевода. Естественно, что в неадаптированных под общую лексику текстах (например, в специализированной литературе) доля таких фрагментов будет высокой и итоговый текст вряд ли можно будет считать удобным для дальнейшей работы. Поэтому нам потребуются настроить перевод этих слов.

Начнем с незнакомых слов. Они по умолчанию подчеркиваются красной линией. Нам нужно выделить такое слово и добавить его в пользовательский словарь (предполагается, что мы знаем его перевод). Если таких слов в тексте очень много, то можно переместиться на вкладку Анализ перевода и выбрать там раздел Незнакомые слова, после чего произвести добавление нужных слов. Кроме того, в **PROMT** поддерживается и «перетаскивание» слов в нижнюю часть окна, где отображается список задействованных словарей перевода, и добавление по горячей клавише F8. Далее от пользователя потребуется заполнить появившуюся форму – указать часть речи, склонение, возможные исключения из правил, перевод и так далее (о том, как это сделать, можно прочитать в материале про создание «шаблонов» перевода в **PROMT**).

Помимо добавления слов в пользовательский словарь можно воспользоваться штатной функцией поиска его перевода в онлайн-словарях (используется известная в среде филологов и лингвистов база «Мультитран») и в существующих словарях **PROMT**, установленных в системе, но неподключенных в данный момент. Для этого нужно выделить слово (или поместить курсор перед ним) и на вкладке Настройка перевода в группе Словарь нажать на кнопку Найти в словарях, или же воспользоваться клавиатурным сочетанием Ctrl+Shift+Z. Переводы появятся на боковой вкладке Словари **PROMT**, где отобразится список переведенных слов с указанием источника. В загруженной вкладке работает поиск по тексту, сортировка и отключение словарей для отображения. Результаты поиска можно копировать, а в случае с онлайн-словарями – добавлять в пользовательский словарь.

Заметим, что в систему **PROMT** можно подключать словари от других производителей. В частности поддерживаются словари Lingvo от ABBYY, Multitex от Paragon Software и Contex от «Информатик».

## Настройка базы переводов

В **PROMT**, как было сказано выше, поддерживается технология параллельных текстов, которая используется для формирования баз переводов Translation Memory. Фактически это представляет собой метод сличения оригинального текста с предоставляемым в данный момент на предмет нахождения в нем устойчивых конструкций, выражений и так далее. Так, например, можно переводить различные инструкции, шаблоны договоров, анкеты и так далее.

При переводе текстов в **PROMT** можно использовать как встроенные (например, базу поговорок и идиом английского языка), так и собственные базы Translation Memory. Соответственно, если при переводе задействуется несколько баз, то в итоговый текст будет добавлен тот вариант, который имеет наибольший процент совпадений с оригиналом или из той базы ТМ, которая имеет больший приоритет (определяется пользователем в настройках).

Точно таким же образом, как и в словарях, при работе с **PROMT** можно использовать данные, находящиеся в неподключенных и незадействованных в переводе базах Translation Memory. Для этого нужно выделить сегмент текста или поместить курсор в начало абзаца, после чего на панели Настройка перевода выбрать пункт Найти в Translation Memory или задействовать клавиатурную комбинацию Ctrl+G. В результате появится вкладка, на которой будут находиться переведенные фрагменты текста, имеющие определенный уровень совпадения (по умолчанию равен 100%, можно снизить). Программа показывает несовпадения (выделяет их цветом) и позволяет ограничивать область поиска уже в найденных переводах через указание процента совпадений при поиске.

Очевидно, что в самом редакторе **PROMT** можно создавать базы переводов прямо на текущих текстах. Для этого потребуется выделить необходимый фрагмент текста или воспользоваться встроенной разбивкой его на сегменты (производится по знакам препинания: точка, вопросительный/восклицательный знаки, знак абзаца, двоеточие), после чего нажать на кнопку Добавить в Translation Memory или нажать на клавиатуре F10. В результате этого у пользователя возникнет диалоговое окно мастера добавления в базу ТМ, где можно указать соответствующее место хранения сегментов и другие параметры.

## Многовариантные слова

В большинстве существующих мировых языков встречаются так называемые многовариантные слова, перевод которых может быть контекстуально-зависимым (омонимы), обусловлен правилами грамматики (например, модальность глагола) и другими свойствами языка. Все подобные слова выделяются в тексте перевода подчеркиванием цветом, зависящим от типа словаря, в котором они были обнаружены (общем, пользовательском или специализированном). Пользователь может задать вид отображения этого фрагмента в тексте, а также политику показа всех или некоторых вариантов перевода у многовариантных слов. По умолчанию **PROMT** выбирает первое по порядку значение слова в словаре, а все остальные отображает следом в скобках через разделитель.

Пользователь может применить нужный ему перевод слова следующим образом – или выделив слово и левой кнопкой мыши выбрав соответствующий пункт из всплывающей панели перевода (ее можно отключить), или кликнув правой кнопкой мыши и вызвав контекстное меню с вариантами. Заметим, что в **PROMT** по умолчанию включен (его можно отключить) режим автоматического редактировать словарных статей и, если пользователь выбирает перевод из второстепенных вариантов, то они становятся первым по приоритету. Статьи заменяются не в основном и не специализированном словаре, а только в пользовательском (он становится автоматически подключенным к системе и хранит все измененные статьи).

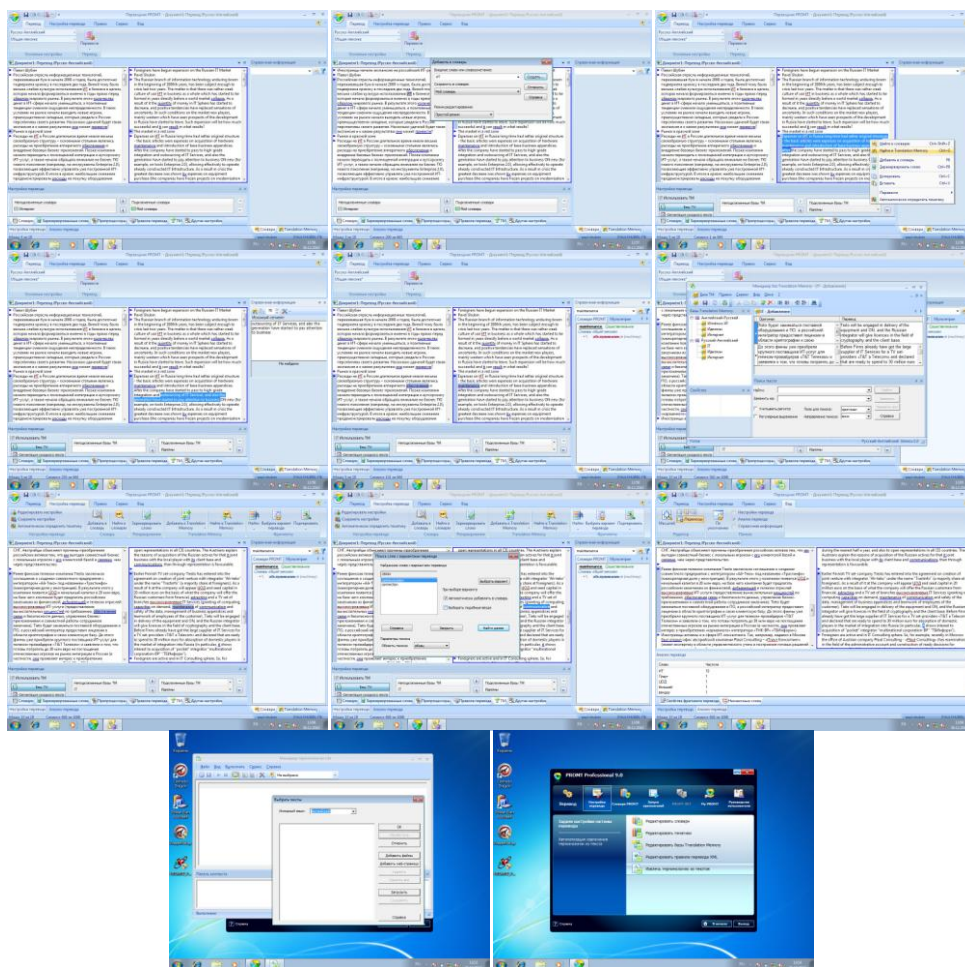
В **PROMT** предусмотрен удобный инструмент последовательного поиска слов с вариантами перевода. Это значит, что в случае, если программа на протяжении всего текста переводила одно и то же слово одним образом, но при этом предлагала варианты, то пользователю не нужно вручную просматривать весь итоговый текст и заменять все вхождения перевода на новые варианты. Подобный механизм несколько отличается от обычного поиска с автозаменой. Чтобы им воспользоваться, нужно на вкладке Настройка перевода выбрать группу Фрагменты и нажать на кнопку Выбрать вариант перевода, после чего в появившемся диалоговом окне Поиск слов с вариантами перевода указать область поиска, политику редактирования словарных статей и автозамену всех подобных слов в тексте. В итоге будет запущен поиск по тексту, который будет останавливаться на найденных словах и предлагать нужные варианты. Заметим, что варианты перевода доступны только для абзацев, которые не редактировались после перевода вручную.

## Управление терминологией

При работе с документами по специализированной тематике важно учитывать использование в них различных терминов. В **PROMT** есть автоматизированное средство для извлечения их из текстов и дальнейшего включения в словари. Оно называется «Менеджер терминологии Lite» и предназначено для выявления так называемых «терминологических кандидатов», то есть слов и словосочетаний, встречающихся в текстах с определенной частотой. Анализ могут быть подвергнуты файлы в форматах TXT, HTML, MS Office (Word, Excel, PowerPoint) 95-2007, а также веб-страницы. Пользователь может добавлять их поодиночке или группой, а также сохранять этот список для дальнейшего использования.

Инструмент извлечения терминологии имеет настройки обработки. Во-первых, он может различать регистр разных слов, то есть, слов с заглавной и прописной букв, аббревиатур, написанных одинаково с общеупотребительными словами; во-вторых, исключать из поля анализа повторяющиеся предложения; в-третьих, нормализовывать (приводить к исходной форме) все словоформы. Результатом работы Менеджера терминологии является список терминологических кандидатов, для которых отображается частота вхождения в тексте, а также возможные переводы в словарях **PROMT** (если они есть). Кроме того, приложение позволяет увидеть контекст использования этого термина в виде специальной панели (терминологический кандидат выделен красным цветом, для окружающего контекста доступен перевод).

Найденные терминологические кандидаты могут быть добавлены в словари пользователя. Для этого нужно или выполнить двойной клик по этому слову, или правой кнопкой мыши выбрать соответствующий пункт меню, или нажать Ctrl+G. Процедура добавления аналогична добавлению слова из текста перевода.



## Заключение

Мы рассмотрели основные возможности улучшения качества перевода в системе **PROMT**. В целом, система позволяет выполнить тонкую настройку этого процесса на разных уровнях сложности, что дает возможность получить выверенный и корректный перевод.

*Павел Шубин*