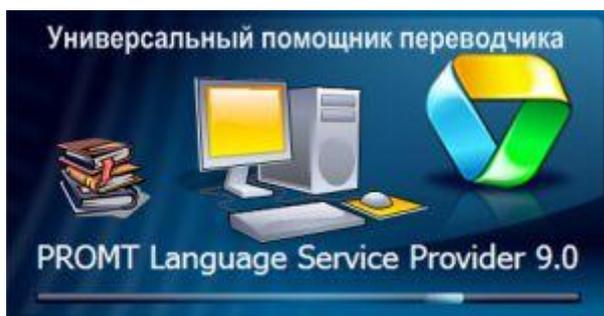


Вспомнить все! Компьютер в помощь переводчику. Часть 2

31.12.2010 / [Программы](#) / Максим Тигулев



В первой части статьи мы вспомнили историю машинного перевода, рассмотрели три составные части успешного перевода, а также изучили системы помощи переводчику. Давайте применим полученные знания на практике, сделав машинный перевод пробного английского текста с помощью программы [PROMT Language Service Provider](#)

[9.0 \(LSP\)](#), а затем, попробуем настроить результат под собственные требования и улучшить результат перевода, применив различные методы оптимизации, доступные в данной программе.

Итак, представим, что мы работаем в некой компании и нам регулярно приходят письма от иностранного партнера - компании MagisterCard. Входящие письма необходимо регистрировать, переводить на русский язык для руководства и использования в работе с клиентами. Вот пример письма, с которым мы будем работать:

MagisterCard Member Letter #25

Disclaimer: This e-mail is intended only for the MagisterCard Member. It contains information that is confidential and protected from disclosure. Any review, dissemination or use of this transmission or its contents by unauthorized persons is strictly prohibited.

To: Business development department

Subject: MagisterCard Experiences and offers

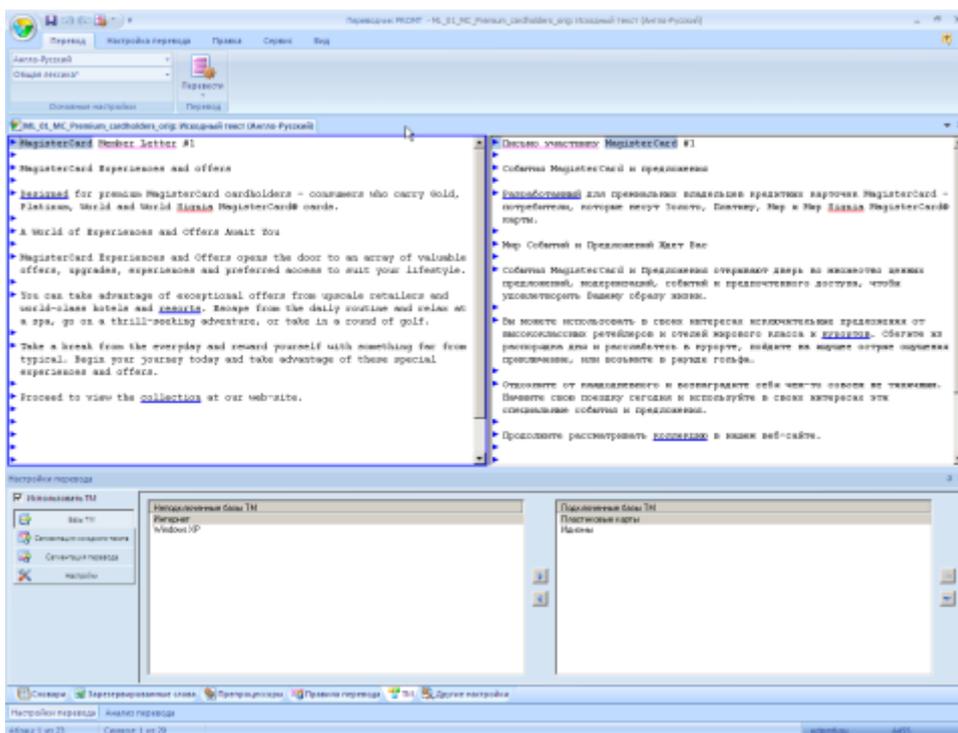
This offer is designed for premium customers with Gold, Platinum and World Signia MagisterCard® cards.

Dear customers!

MagisterCard Experiences and Offers offer to you an array of valuable offers. Take a break on your job, make a pause in your business, relax at our wonderful hotels, visit our outstanding spa or overcome your partner in a round of golf. Escape from the everyday troubles, have fun at our bars and enjoy at our discotheques.

Best wishes, MagisterCard

Перевод документов мы будем делать с помощью программы PROMT LSP 9.0, которая предлагает переводчику набор разных уровней подхода к переводам текстов, что позволяет переводчику переводить тексты с требуемой точностью и скоростью.



Экран главного окна перевода

Условно можно выделить такие уровни:

- Первый уровень - машинный перевод. Программа осуществляет связный перевод текста, используя определенные лингвистические алгоритмы. Сначала система анализирует структурные элементы входного предложения, затем преобразует его в соответствии со структурой перевода и синтезирует окончательный вариант. Для перевода используется встроенный словарь.
- Второй уровень - подключение специализированных словарей по теме перевода (финансы, юриспруденция, промышленность и другие).
- Третий уровень - пользовательский словарь, содержит особые варианты перевода, например термины, принятые внутри компании.
- Четвертый уровень - подключение баз данных памяти перевода, как встроенных, так и собственных. В программе PROMT LSP 9.0 уже встроены базы идиом ("Every cloud has a silver lining" - "Нет худа без добра"), привычной терминологии Windows ("Click Network and Internet Connections" - "Щелкните по ссылке Сеть и подключения к Интернету"), а также терминологии сети Интернет ("Add to favourites" - "Добавить в избранное").

Давайте переведем текст письма, не изменяя настроек PROMT, а затем улучшим результат перевода, применив различные методы оптимизации, доступные в программе. Вот результат перевода с типовыми настройками программы PROMT:

Письмо члена MagisterCard #25

Правовая оговорка: эта электронная почта предназначена только для члена MagisterCard. Это содержит информацию, которая является конфиденциальной и защищена от раскрытия. Любой обзор, распространение или использование этой передачи или ее содержания неправомерными людьми строго запрещены.

К: отдел Коммерческого развития

Предмет: События MagisterCard и Предложения

Это предложение разработано для премиальных клиентов с Золотом, Платиной и Миром Signia MagisterCard® карты.

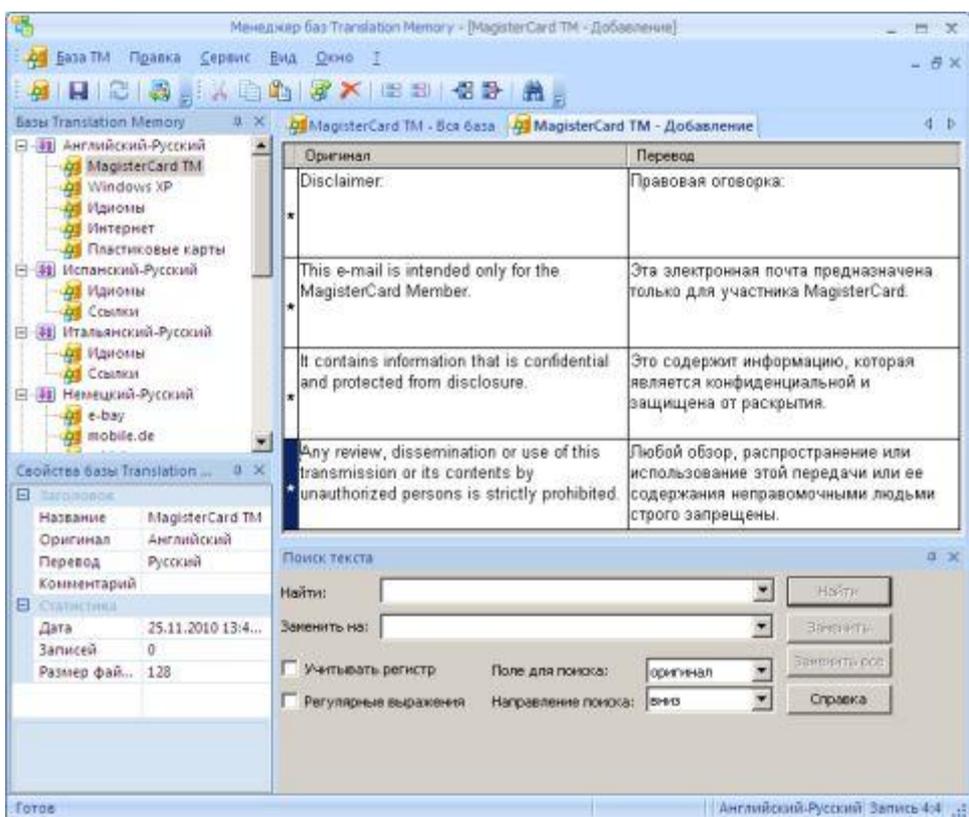
Дорогие клиенты!

События MagisterCard и Предложения предлагают Вам множество ценных предложений. Отдохните на своей работе, сделайте паузу в своем бизнесе, расслабьтесь в наших замечательных отелях, посетите наш выдающийся курорт или преодолите своего партнера в раунде гольфа. Сбегите из каждодневных проблем, весело проведите время в наших барах и наслаждайтесь в наших дискотеках.

Наилучшие пожелания, MagisterCard

Ну что же, машинный перевод дает вполне приемлемый результат, но улучшить кое-что в данном переводе возможно.

- Зарезервируем с помощью соответствующей функции название карточной программы World Signia, чтобы название машина не переводила. Для этого выделим в оригинале текста это название и нажмем сочетание клавиш Ctrl-F8, а затем - ОК. Теперь название World Signia не будет переводиться на русский язык.
- Мне никогда не нравился стандартный перевод слова Member, поэтому добавим слово Member в пользовательский словарь (по умолчанию он называется "Мой словарь", но можно создать и другой, эту возможность мы рассмотрим ниже). Для этого выделим слово и нажмем клавишу F8, затем в появившемся окне нажмем кнопку "Создать". В таблице "Создать словарную статью" напротив слова укажем желаемый перевод - "участник" и нажмем кнопку "Сохранить". Теперь слово везде переводится, как надо нам.
- Абзац, содержащий правовую оговорку (Disclaimer), будет одинаковым во всех входящих письмах, поэтому исправим получившийся перевод согласно стандартам нашей компании и добавим его в базу памяти переводов. Для этого выделим абзац и нажмем клавишу F10, откроется диалог "Базы Translation Memory". Создадим новую базу, нажав кнопку "Новая". В поле "Название" введем название "MagisterCard TM" и нажмем ОК. Выберем вновь созданную базу. После чего откроется окно менеджера баз Translation Memory. В окне менеджера мы увидим уже сегментированный по предложениям текст абзаца. Так как абзац нам нужно сохранить полностью, объединим сегменты сначала оригинала, а потом и перевода, нажав кнопку на панели инструментов или выбрав пункт меню "Правка/Объединение и разделение записей". В правой части окна переведем абзац согласно требованиям юридического отдела и сохраним изменения в базе. Для использования созданной базы TM необходимо в нижней части главного экрана перевода выбрать закладку "TM", поставить галочку напротив пункта "Использовать TM", установить курсор на название нашей базы и подключить ее, нажав стрелку вправо. Название базы появится в колонке "Подключенные базы TM".

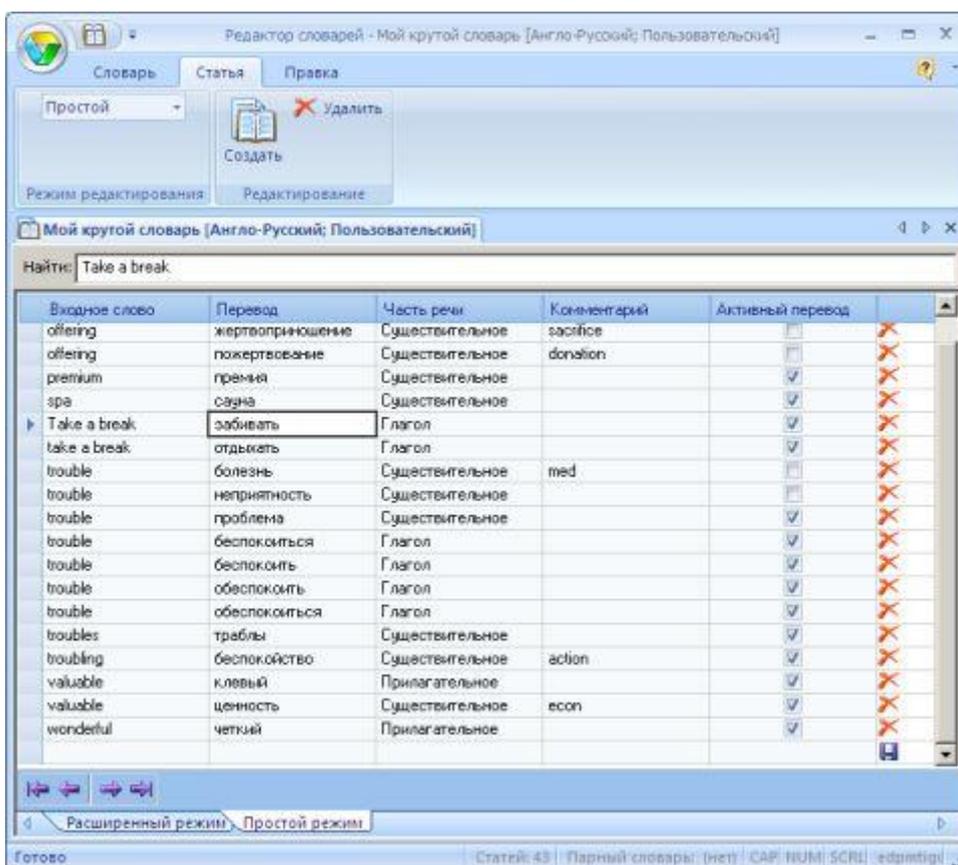


Окно менеджера баз Translation Memory

После выполнения перечисленных действий мы получим достаточно точный перевод, нуждающийся в минимальной правке.

Все - письмо можно рассылать ВИП-клиентам.

А теперь вообразим, что среди наших клиентов есть представители неформальных течений и к ним надо обратиться на их языке. Почему бы и нет? Помните, у "Альфа-банка" была реклама под девизом "С каждым клиентом мы найдем общий язык", в которой крутой парень приходил в отделение и просил "Залить кеш в понтовый пластик"? Давайте настроим PROMT LSP 9.0 так, чтобы он автоматически переводил текст письма на молодежный сленг. В этом нам помогут пользовательский словарь и сайт ["СловоНово"](#).



Окно редактора словарей

В нижней части главного экрана перевода выберем закладку "Словари", щелкнем в левом столбце правой кнопкой мыши и выберем пункт "Создать новый", в окне введем имя словаря, пусть будет "Мой крутой словарь". Перед использованием словарь надо подключить, для чего надо щелкнуть на его имени в левой колонке, а затем кликнуть на зеленую стрелку, показывающую вправо. Теперь словарь можно наполнять. Сделаем это прямо из оригинального текста - выделим слово "offers", нажмем кнопку F8, затем "Создать", в окне "Создать словарную статью" введем нашу версию перевода слова "offers" - "фишки", а затем - "Сохранить". Добавим остальные слова и переводы, более полно отражающие мировоззрение нашей целевой аудитории. Снова запустим текст на перевод. Получим следующий текст:

Письмо участника MagisterCard #25

Правовая оговорка: это сообщение электронной почты предназначено только для участника MagisterCard. Оно содержит информацию, которая является конфиденциальной и защищена от раскрытия. Любой обзор, распространение или использование этого сообщения или его содержания неправомочными лицами строго запрещены.

К: отдел Коммерческого развития

Предмет: Ивенты MagisterCard и фишки

Это предложение разработано для реальных пацанов с Золотом, Платиной и World Signia карты MagisterCard®.

Эй, пацаны!

Ивенты MagisterCard и Фишки предлагают Вам тучу клевых фишек. Забейте на своей работе, сделайте паузу в своем бизнесе, расслабьтесь в наших четких отелях, посетите

*нашу кайфовую сауну или сделайте своего партнера в раунде бильярда. Свалите из каждодневных траблов, оттянитесь в наших барах и тащитесь в наших дискотеках.
Чмоки, MagisterCard*

Этот шуточный пример проиллюстрировал, как можно, всего лишь подключив специальный словарь, из одного и того же исходного текста создать совершенно другое письмо.

Подобный метод применим также для специальной терминологии, принятой внутри компании.

Подводя итог проведенному исследованию, отметим, что благодаря хитроумным алгоритмам проблема машинного перевода решается с приемлемой степенью точности, однако полностью заменить человека машине пока не удастся. Но, как это часто бывает, если совместить аналитическую мощь компьютера и творческий подход человека (придать компьютер в помощь человеку), то результаты будут отличные.

В завершение разберемся с особенностями работы переводчика Google. Как мы поняли, банки памяти перевода с одного языка на другой обычно составляют и редактируют люди. Схема работы системы перевода Google - принципиально отлична. Набор правил и схем перевода формируется на основе статического анализа большого объема текстов, составляя огромную базу памяти перевода. То есть получается, что если какая-то фраза в статистически значимой части текстов переводится определенным образом, то она считается правильной и помещается в базу памяти перевода. При этом база памяти перевода еще и периодически обновляется на основании новых статистических данных. Так что разгадка "ошибки" переводчика Google - проста: в наборе проанализированных Google текстов достаточно часто встречалось сочетание "калина - Mazda" и Google поместил это сочетание в свою базу памяти перевода. Кстати, Google довольно оперативно реагирует на найденные пользователями "ошибки", поэтому на момент публикации статьи эта "ошибка" уже исправлена.