

Совместное применение SDL Trados Studio 2011 и PROMT LSP 9.5 в технологическом процессе перевода

Выгодский Олег, Аргонавт

olegv@argonaut.spb.ru

Структура документа:

- 1. Общее описание рабочего процесса**
- 2. Применяемые программные продукты**
- 3. Подробное описание рабочего процесса**

Общее описание рабочего процесса

Главные задачи: необходимо перевести файл(ы) формата **SDLXLIFF** с помощью пакета **SDL Trados Studio 2011**. При этом, все «неизвестные сегменты», которых нет в памяти ТМ, должны быть переведены системой **PROMT LSP 9.5**, которая должна быть настроена надлежащим образом. Форматирование (теги) машинно переведенных «неизвестных сегментов» должно быть сохранено. Напрямую эти продукты между собой не связаны.

Следовательно, необходимо найти способ встраивания указанных выше продуктов в технологический процесс перевода.

Основная идея: из документа, подлежащего переводу, извлекаются «неизвестные сегменты» (в формате **TMX**). Они переводятся с помощью модуля пакетного перевода **PROMT**, а затем импортируются в отдельную базу ТМ в **SDL Trados Studio 2011**.

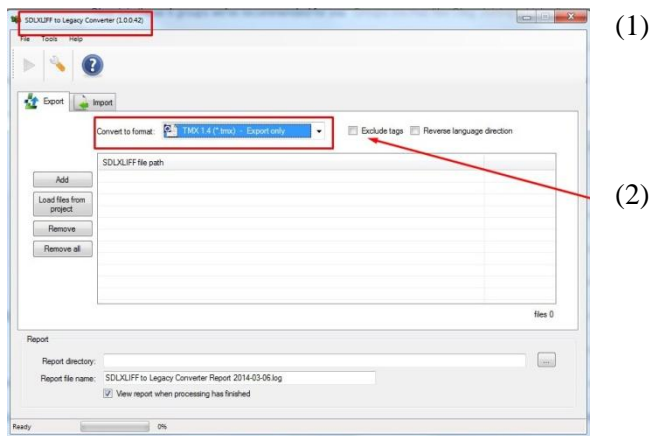
Примечание: описанную ниже процедуру также можно использовать вместе с продуктами **SDL Trados Studio 2009 и 2014**, а также других систем **PROMT**, начиная с версии **9.5** (например, **PROMT Professional 9.5**).

Применяемые программные продукты

1. SDLXLIFF to Legacy Converter
2. Olifant – Okapi Translation Memory Editor
3. PROMT LSP 9.5
4. SDL Trados Studio 2011

Подробное описание рабочего процесса

SDLXLIFF to Legacy Converter (1.0.0.42)

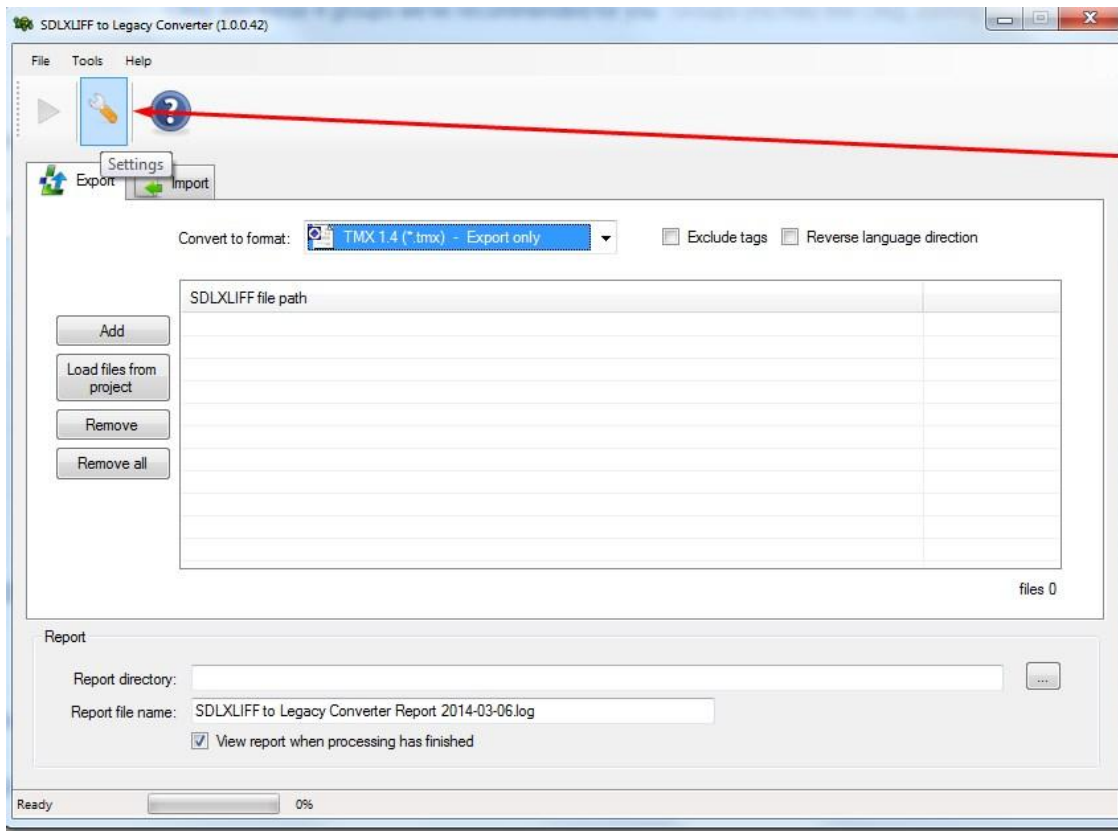


1. Запустите **SDLXLIFF to Legacy Converter** для извлечения «неизвестных сегментов» в формате **TMX**. Выберите опцию конвертации в формат **TMX**.
2. Не ставьте галочку в поле **Exclude tags**. На данном этапе нужно получить тегированный **TMX**-файл, который подлежит машинному переводу с помощью **PROMT**-а.

Примечание 1: если вы применяете пакет **SDLXLIFF to Legacy Converter** более ранней версии (например, 1.0.0.27), то при выборе опции конвертации в формат **TMX**, скорее всего, будут извлечены не все «неизвестные сегменты». Поэтому, вам будет нужно использовать опцию конвертации в формат **TTX**, а затем анализировать результирующий **TTX**-файл(ы) с помощью **SDL Trados 2007**. В настоящей презентации описывается применение упомянутого выше пакета версии 1.0.0.42 (см. рис. выше).

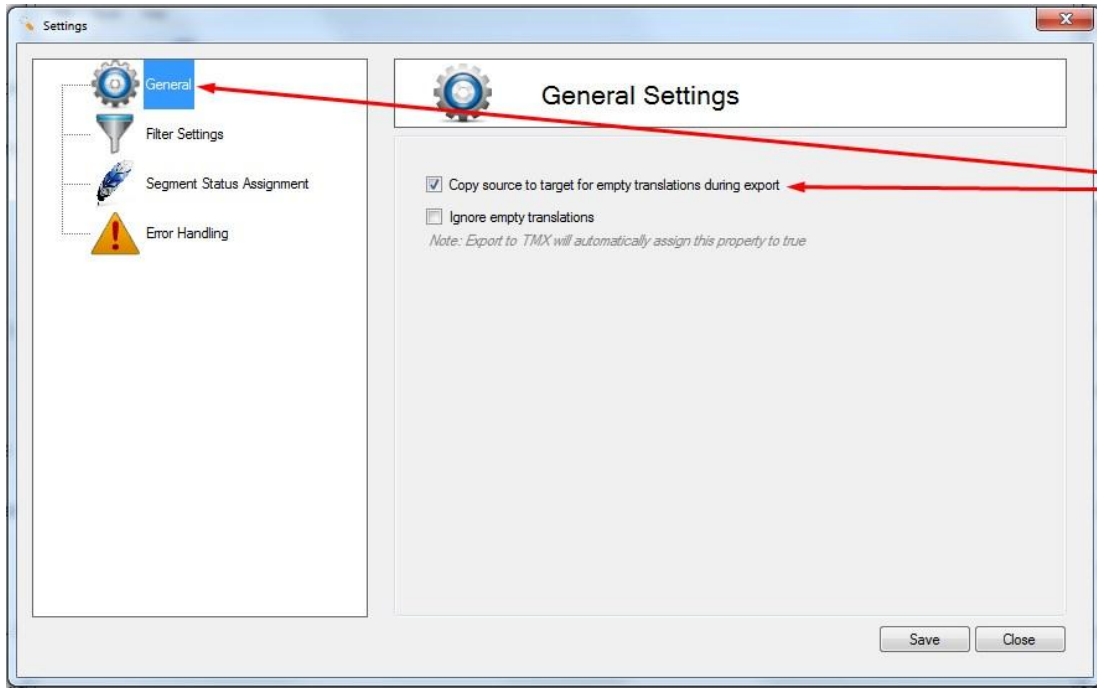
Примечание 2: процедуру анализа **TTX**-файла(ов) с помощью **SDL Trados 2007** и экспорта «неизвестных сегментов» в формате **TMX 1.4b** см. в документе «Совместное применение SDL Trados 2007 и PROMT LSP 9.5 в технологическом процессе перевода» (http://argonaut.spb.ru/files/Trados_2007_Promt.pdf).

SDLXLIFF to Legacy Converter



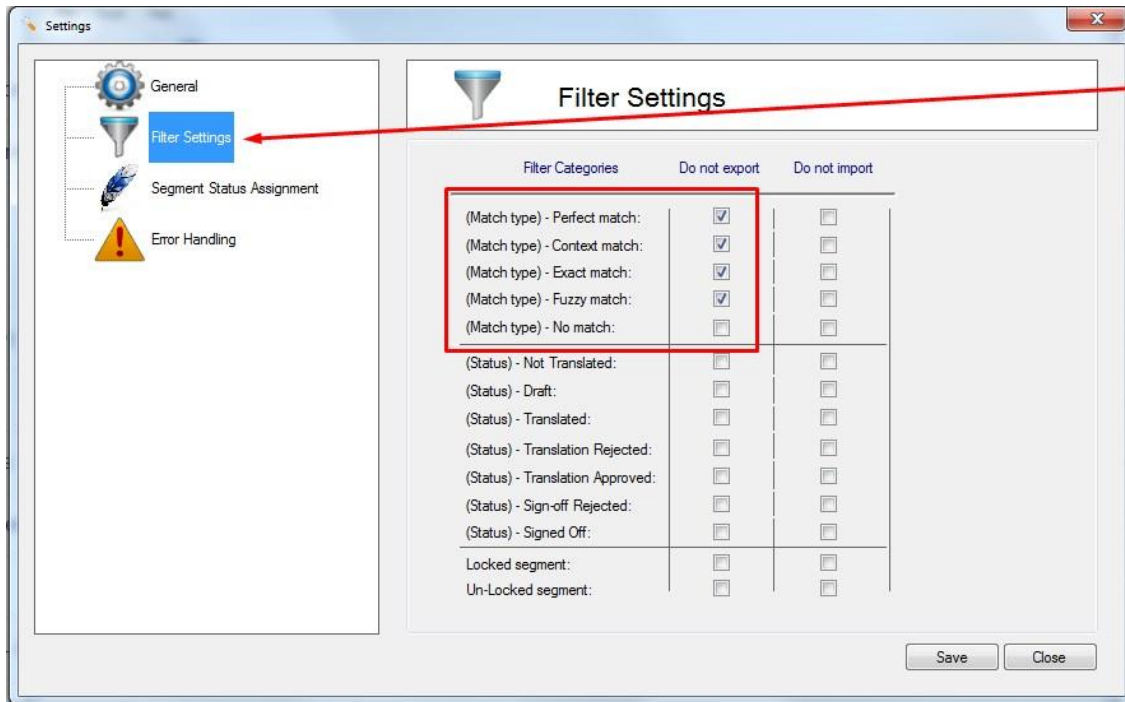
Нажмите на кнопку **Settings** для настройки параметров.

SDLXLIFF to Legacy Converter



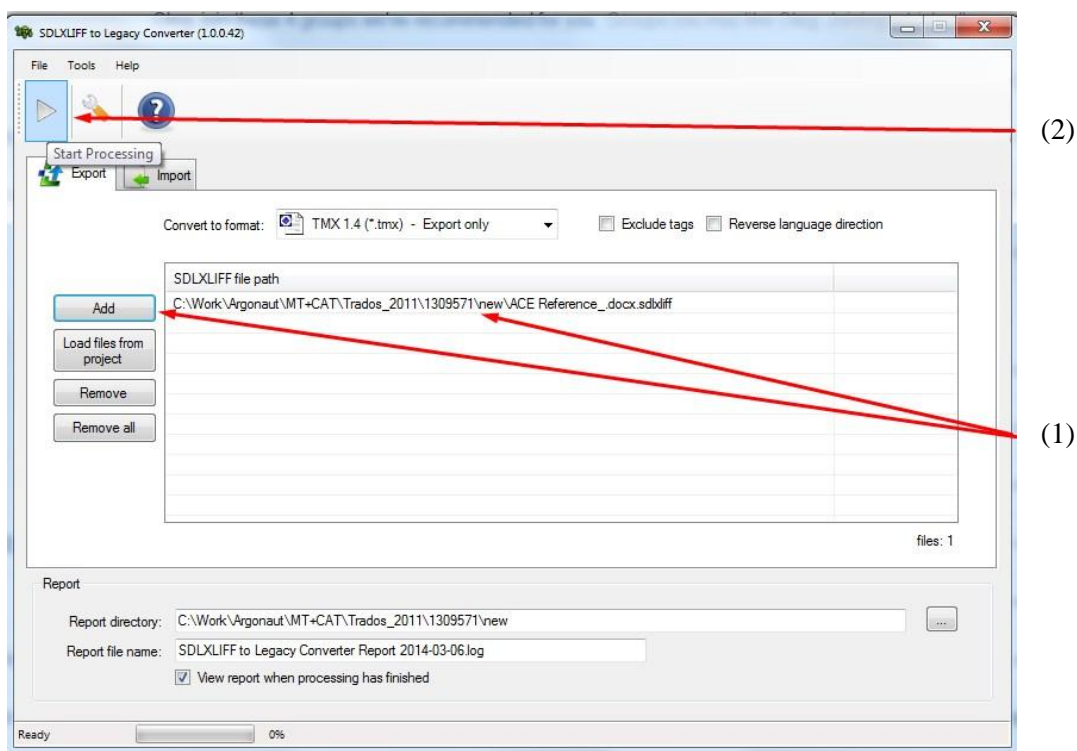
В области **General** установите флажок для копирования исходных сегментов в целевые.

SDLXLIFF to Legacy Converter



В области **Filter Settings** установите флажки в верхней группе, кроме «неизвестных сегментов» (**No match**). Это позволит экспортировать только «неизвестные сегменты».

SDLXLIFF to Legacy Converter



1. Нажмите на кнопку **Add** для выбора файла, из которого следует экспортировать «неизвестные сегменты».
 2. Нажмите на эту кнопку для запуска процесса экспорта. Экспортированный файл будет создан в том же каталоге, где находится анализируемый файл (SDLXLIFF), и будет иметь расширение ***.sdlxiff.tmx**.
- Примечание:** если вы выбрали несколько анализируемых файлов (SDLXLIFF), то будет создано по одному экспортированному TMX-файлу для каждого исходного файла.

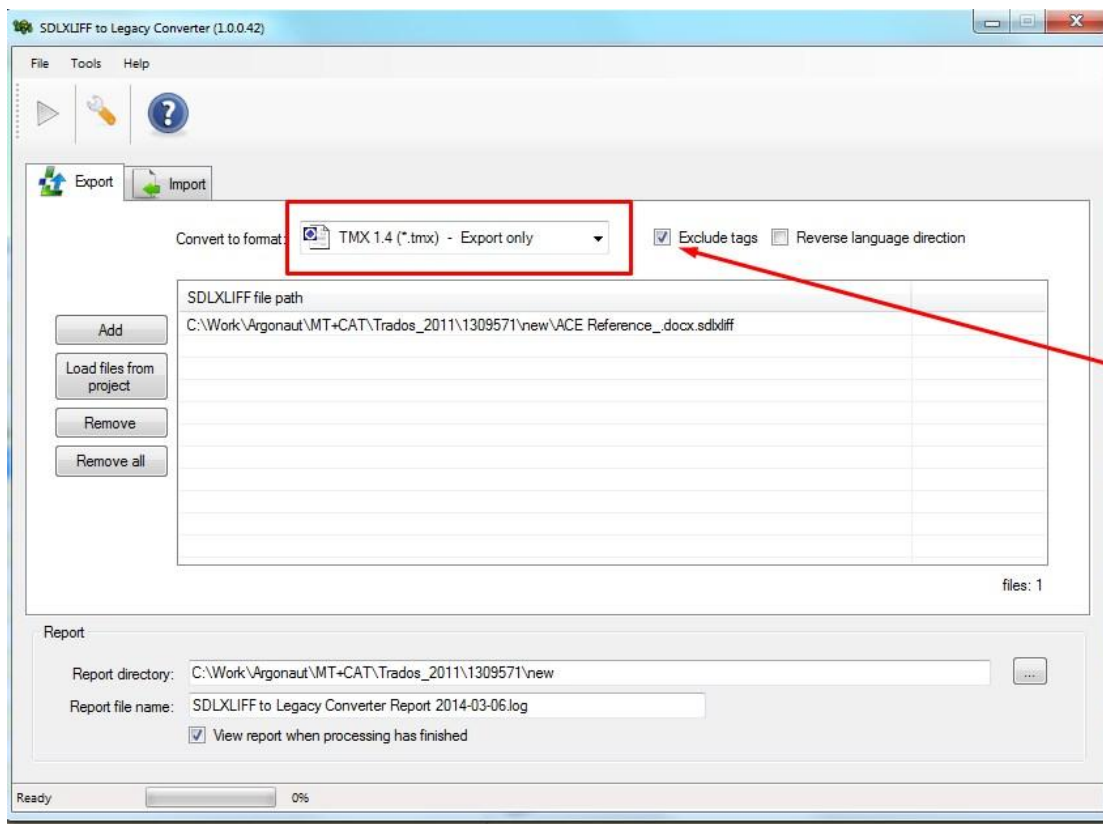
SDLXLIFF to Legacy Converter

```

SDLXLIFF to Legacy Converter Report 2014-03-06 — Блокнот
Файл  Правка  Формат  Вид  Справка
SDLXLIFF to Legacy Converter (1.0.0.42)
Report Name      : SDLXLIFF to Legacy Converter Report 2014-03-06.log
Processing Type  : Export
Conversion Type  : TMX 1.4 (*.tmx) - Export only
Start Processing: 06.03.2014 12:52:14
-----
Total files processed: 1
-----
Original file   : C:\work\Argonaut\MT+CAT\Trados_2011\1309571\new\ACE Reference_.docx.sdlxliff
Converted file  : C:\work\Argonaut\MT+CAT\Trados_2011\1309571\new\ACE Reference_.docx.sdlxliff.tmx
Source Language : en-GB - English (United Kingdom)
Target Language : ru-RU - Russian (Russia)
Total segments
Read            : 1793
Exported       : 1516
Ignored        : 277
-----
End Processing: 06.03.2014 12:52:27
-----
    
```

Выше показан отчет об экспорте «неизвестных сегментов» в формате **TMX** из анализируемого исходного файла. Следует напомнить, что это будет тегированный TMX-файл, который создается в том же каталоге, что и исходный файл (SDLXLIFF). Создайте для данного TMX-файла **отдельный каталог** и переместите его туда.

SDLXLIFF to Legacy Converter



Ранее мы экспортировали «неизвестные сегменты» в **тегированный** TMX-файл. В настоящее время пакет PROMT не позволяет открывать такие файлы и маскировать форматную информацию, включая теги. Поэтому, описанную выше процедуру экспорта в формат TMX необходимо повторить, но установить флажок в поле **Exclude tags**. Это позволит удалить все теги в экспортируемых сегментах. Данный «нетегированный» файл потребуется нам для проверки машинного перевода и настройки ресурсов PROMT-а.

Перед повторным выполнением процедуры экспорта перенесите уже экспортированный тегированный TMX-файл в **отдельный каталог** (в противном случае, вы затрете уже созданный TMX-файл).

Проверьте настройки и снова запустите процесс конвертации. В результате, будет создан **нетегированный** TMX-файл.

Запустить PROMT? Пока НЕТ!

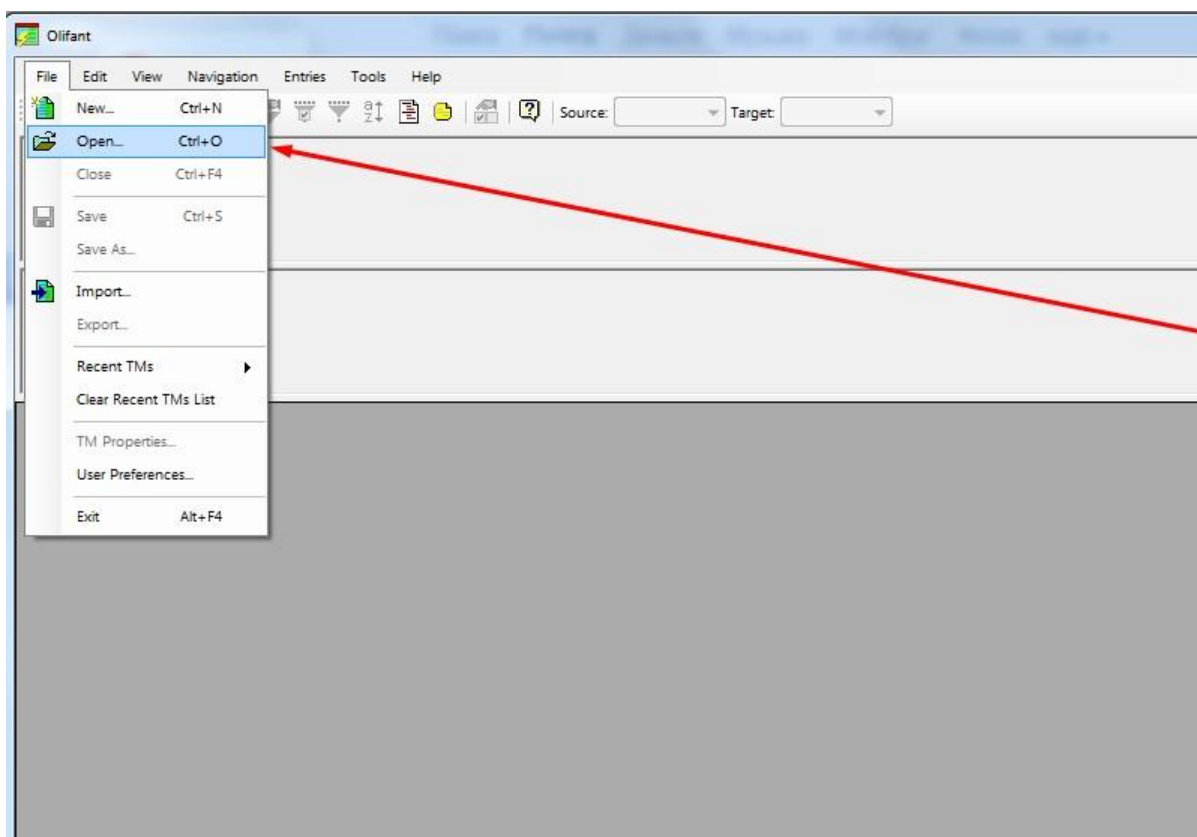
Не запускайте вслепую машинный перевод! Убедитесь, что машинный перевод настроен.

PROMT может переводить TMX-файлы в **пакетном режиме**. Но (текущие версии) не позволяет открывать такие файлы во встроенном редакторе и скрывать форматную информацию. Поэтому, экспортированный **нетегированный** TMX-файл следует преобразовать в обычный текстовый файл – с помощью пакета **Olifant**. Такой файл нам нужен только для настройки ресурсов системы PROMT.

Впоследствии (после настройки PROMT-а) в пакетном режиме будет нужно обработать исходный тегированный TMX-файл.

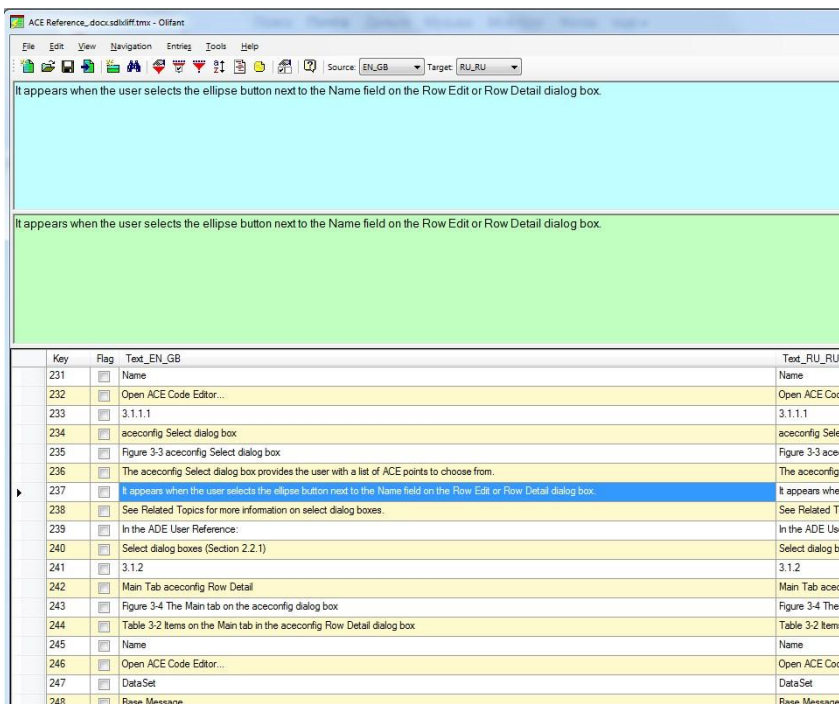
Итак, откройте **Olifant**.

Olifant



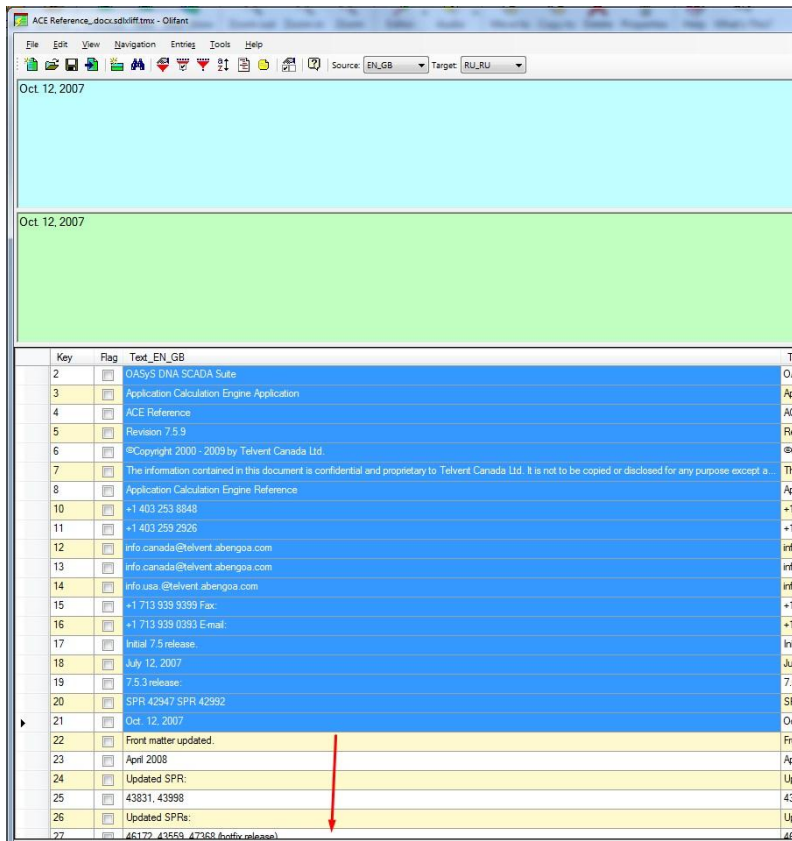
Нажмите на кнопку **Open** и выберите экспортированный **нетегированный** TMX-файл.

Olifant



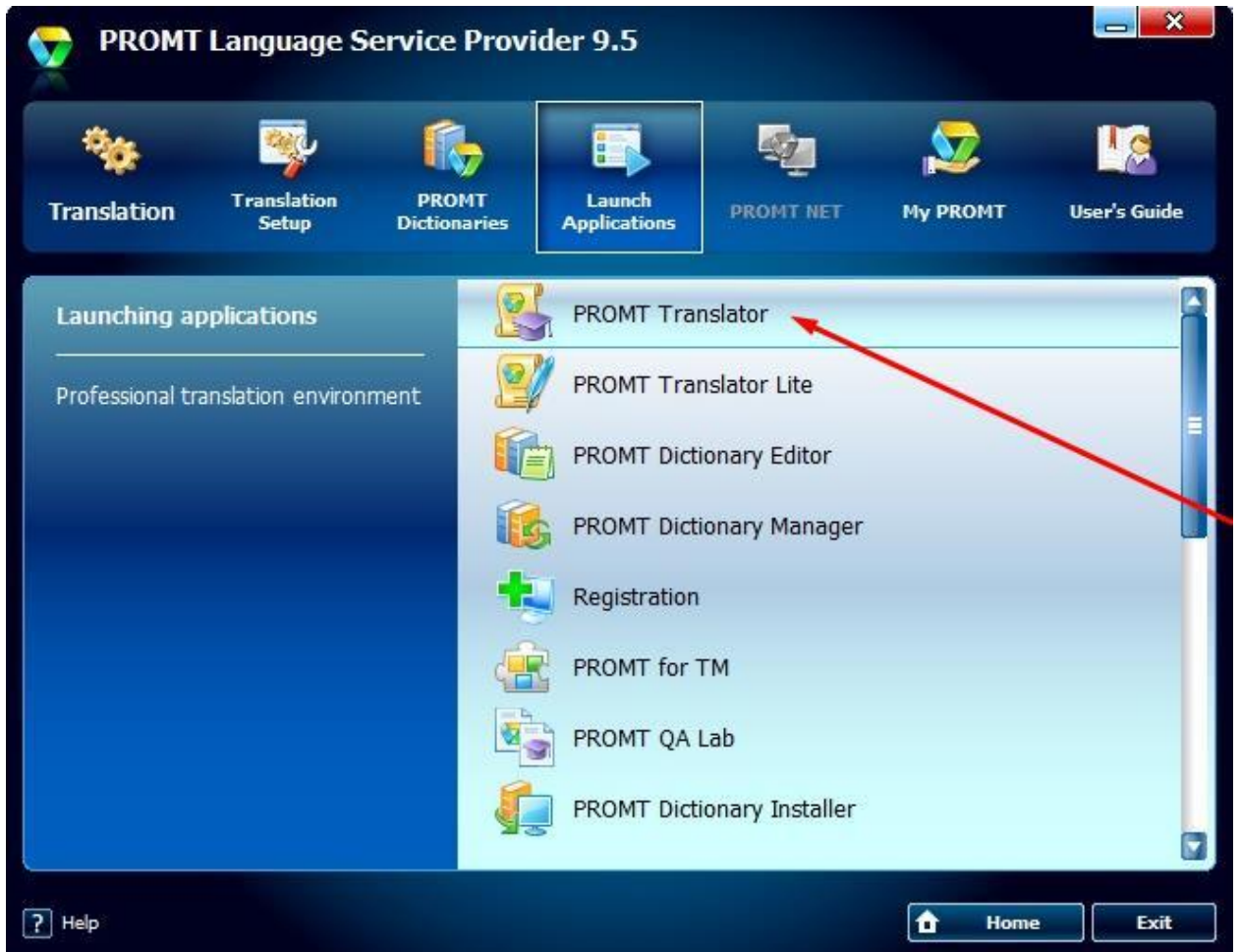
В открытом TMX-файле теги отсутствуют. Следует напомнить, что теги на данном этапе нам не нужны, так как нам требуется «чистый» текст для загрузки в PROMT.

Olifant



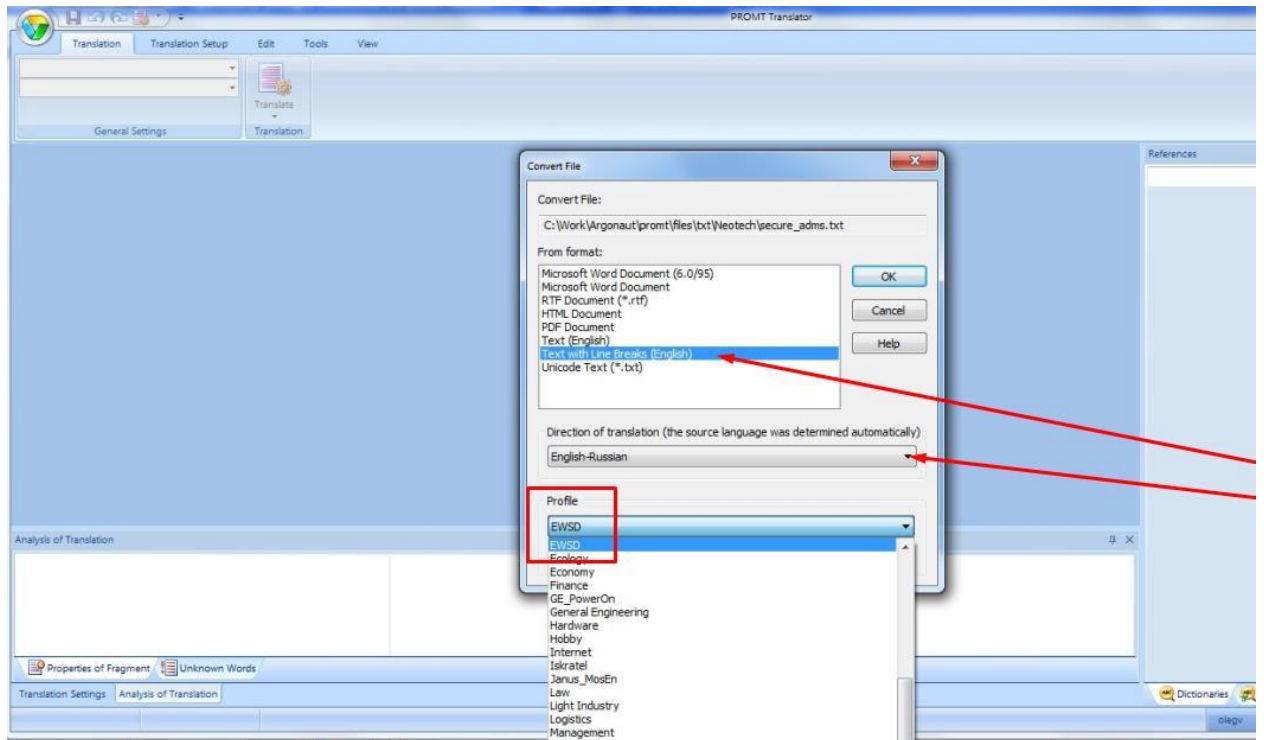
Выделите весь столбец полностью и сохраните в текстовом файле (либо скопируйте в буфер обмена).

PROMT



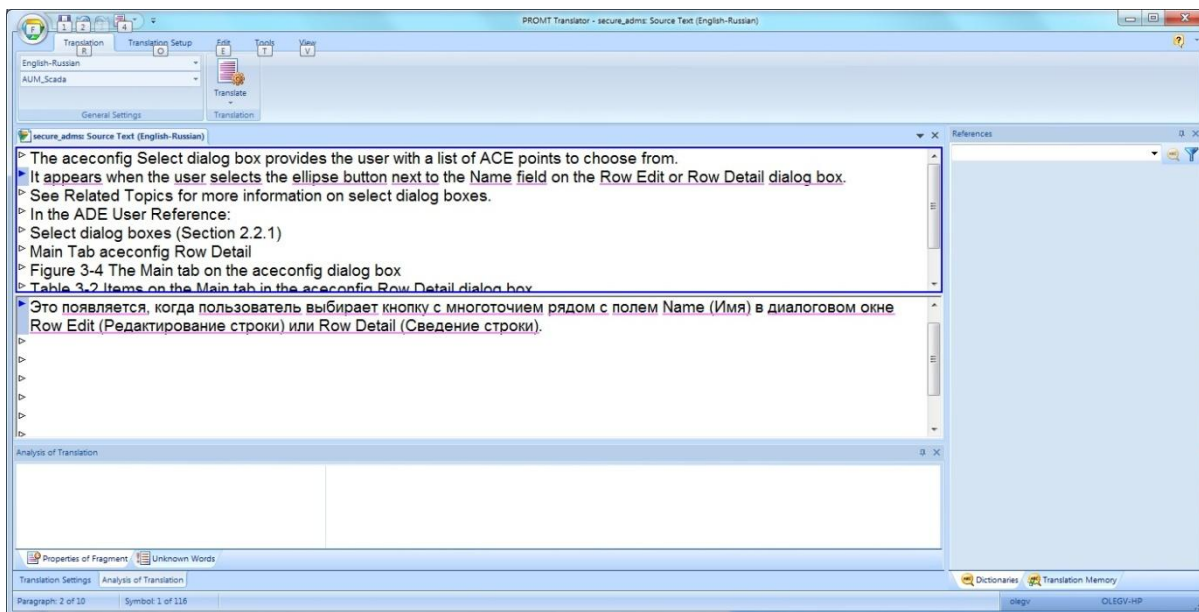
Запустите **PROMT Translator**.

PROMT



Выберите формат файла (просто текст) и профиль.

PROMT



Откройте в редакторе PROMT текстовый файл или создайте новый файл и скопируйте в него текст из буфера обмена.

Настройте необходимые ресурсы PROMT-а (словари, список зарезервированных слов, правила перевода) и убедитесь, что PROMT обеспечивает нужный машинный перевод на выходе. Редактировать машинный перевод сейчас не нужно.

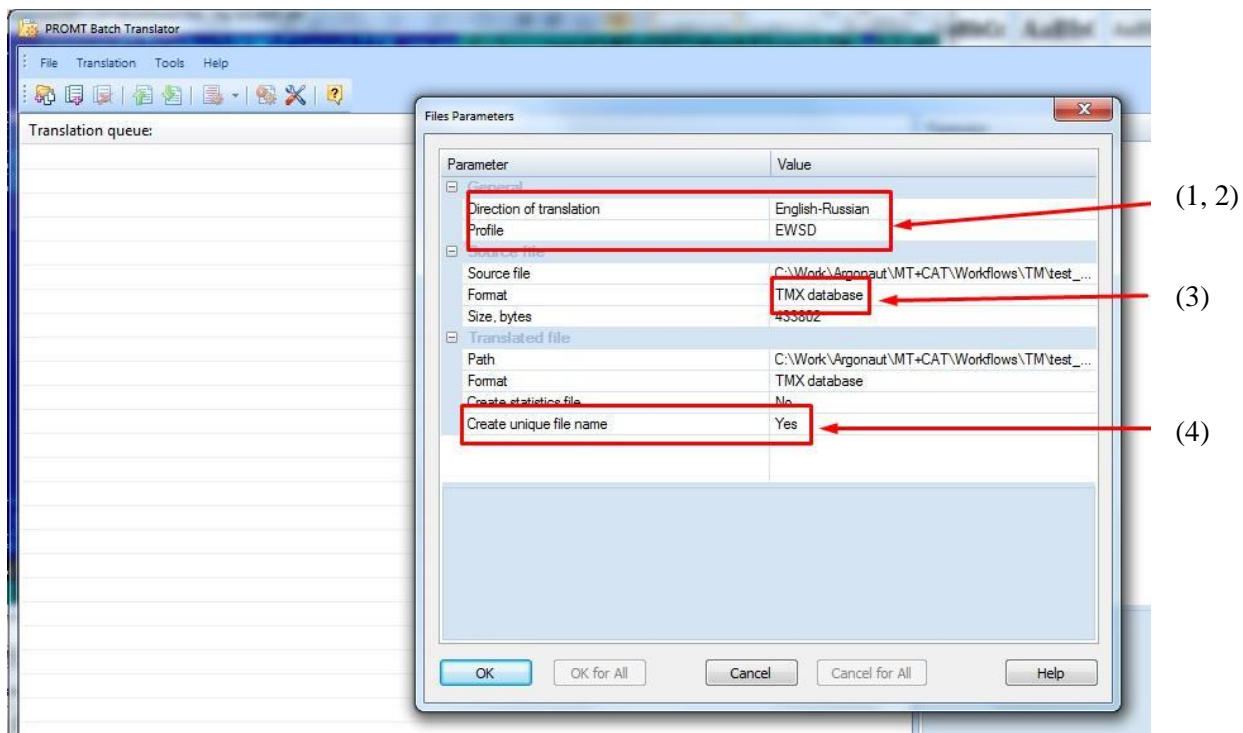
Примечание: дополнительную информацию о разработке терминологии, особенностях словарей системы машинного перевода и возможностях настройки системы PROMT см. в презентации «Машинный перевод как инструмент переводчика» (http://argonaut.spb.ru/files/TFR11_OlegV.pdf).

PROMT



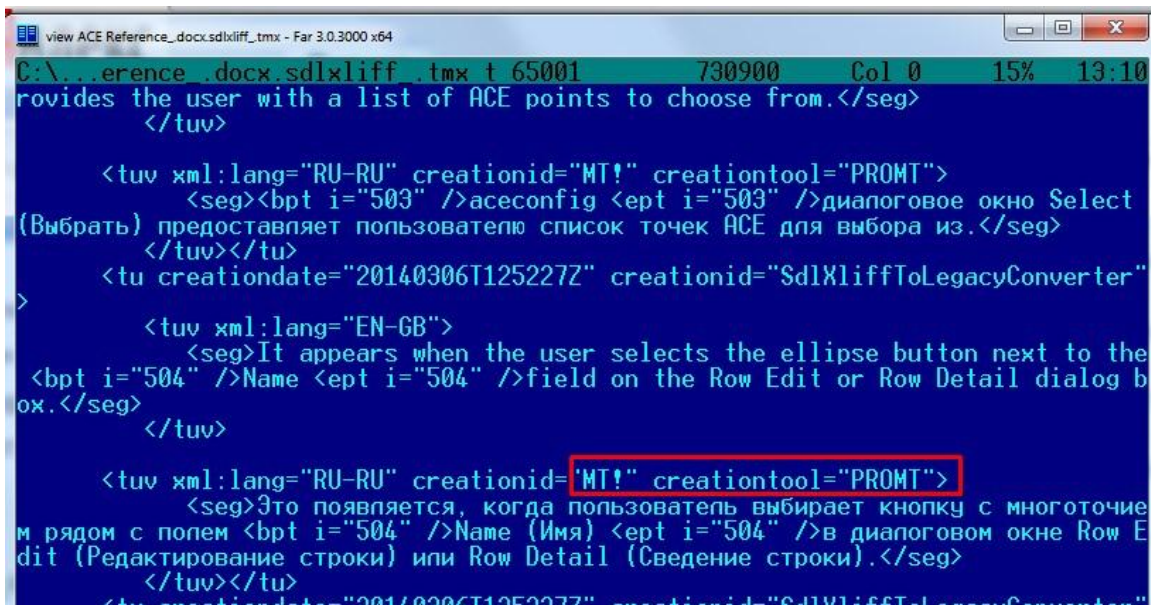
Если вы уверены, что система машинного перевода настроена надлежащим образом, запустите модуль пакетного перевода PROMT (**PROMT Batch Translator**).

PROMT



1. Выберите направление перевода.
2. Выберите профиль.
3. Выберите исходный **тегированный** файл (TMX), который был экспортирован из исходного файла с помощью пакета **SDLXLIFF to Legacy Converter**.
4. Выберите опцию создания уникального имени файла (чтобы не затереть исходный файл).
5. Нажмите на кнопку ОК.
6. Запустите процесс перевода (нажмите на F9).

После PROMT Batch Translator – проверка двуязычного TMX-файла

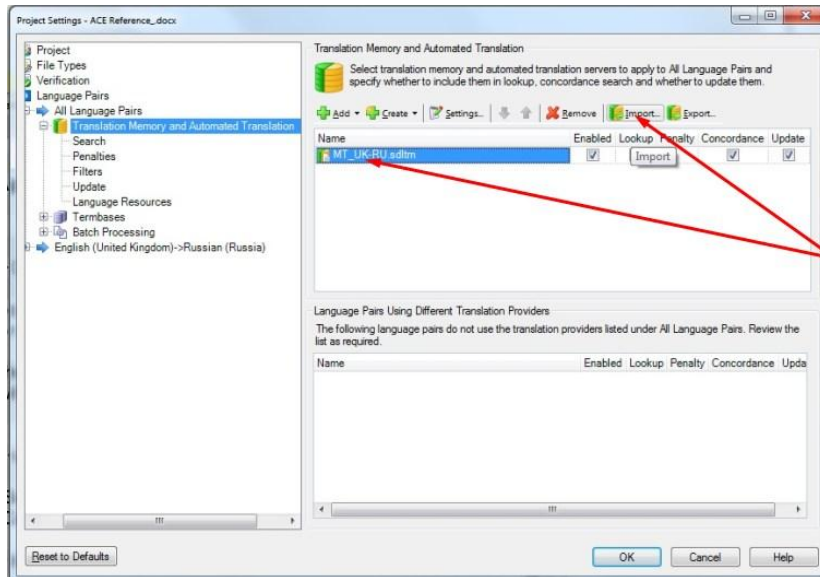


```
view ACE Reference_docx.sdlxliff_tm - Far 3.0.3000 x64
C:\...erence_docx.sdlxliff_tm t 65001 730900 Col 0 15% 13:10
rovides the user with a list of ACE points to choose from.</seg>
</tuv>
<tuv xml:lang="RU-RU" creationid="MT!" creationtool="PROMT">
  <seg><bpt i="503" />acesconfig <ept i="503" />диалоговое окно Select
  (Выбрать) предоставляет пользователю список точек ACE для выбора из.</seg>
</tuv></tu>
<tu creationdate="20140306T125227Z" creationid="SdlXliffToLegacyConverter"
>
  <tuv xml:lang="EN-GB">
    <seg>It appears when the user selects the ellipse button next to the
    <bpt i="504" />Name <ept i="504" />field on the Row Edit or Row Detail dialog b
    ox.</seg>
  </tuv>
  <tuv xml:lang="RU-RU" creationid="MT!" creationtool="PROMT">
    <seg>Это появляется, когда пользователь выбирает кнопку с многоточие
    м рядом с полем <bpt i="504" />Name (Имя) <ept i="504" />в диалоговом окне Row E
    dit (Редактирование строки) или Row Detail (Сведение строки).</seg>
  </tuv></tu>
  <tu creationdate="20140306T125227Z" creationid="SdlXliffToLegacyConverter"
```

(1)

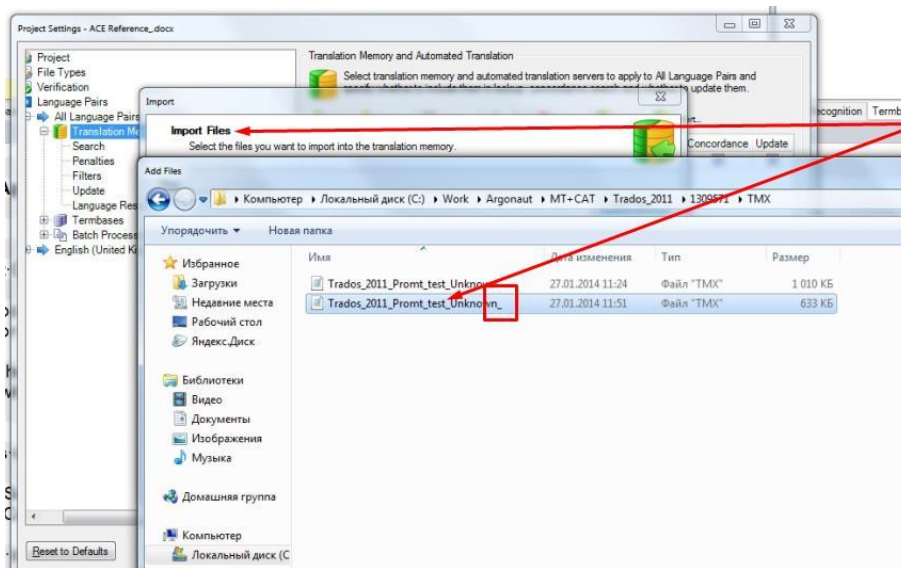
- (1) PROMT создал русскоязычные сегменты с идентификатором **MT!** Следует отметить, что PROMT сохранил теги в сегментах.

SDL Trados Studio



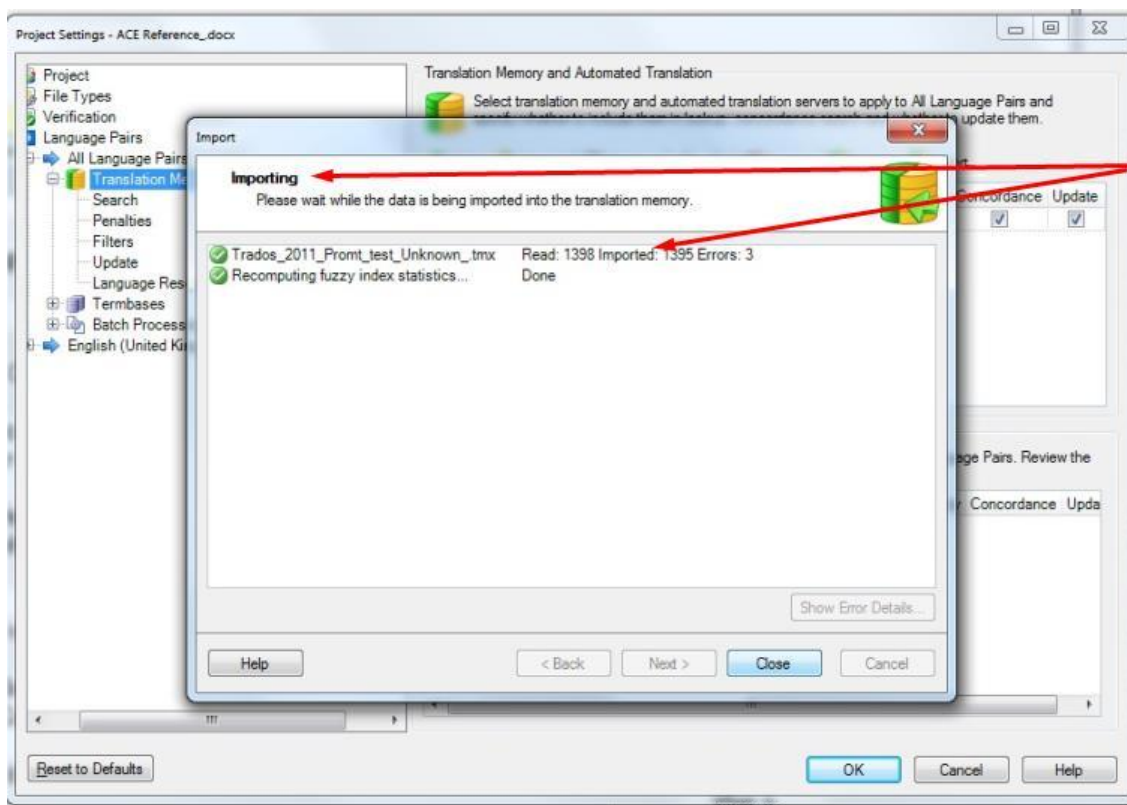
1. Откройте в **SDL Trados Studio** исходный файл (SDLXLIFF) в редакторе.
2. Создайте или откройте (временную) память ТМ для импорта тегированного двуязычного ТМХ-файла (созданного PROMT-ом).

SDL Trados Studio



Выберите файл с уникальным именем и импортируйте его во временную память ТМ. Эта память будет содержать только двязычные сегменты, созданные PROMT-ом.

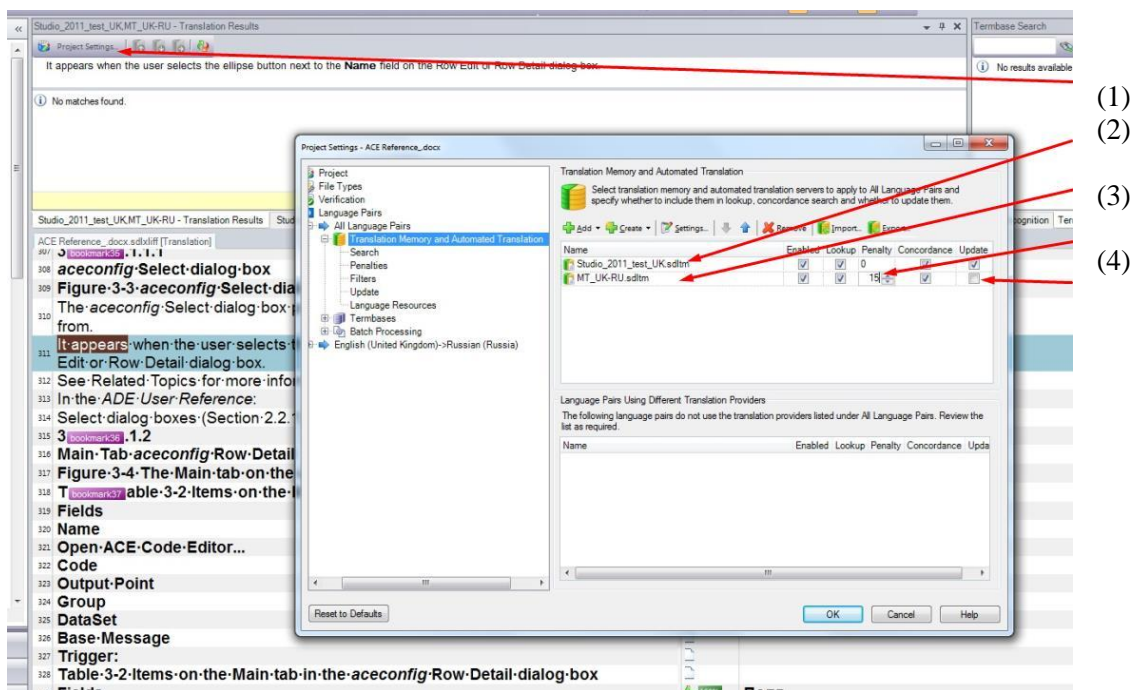
SDL Trados Studio



Операция импорта завершена.

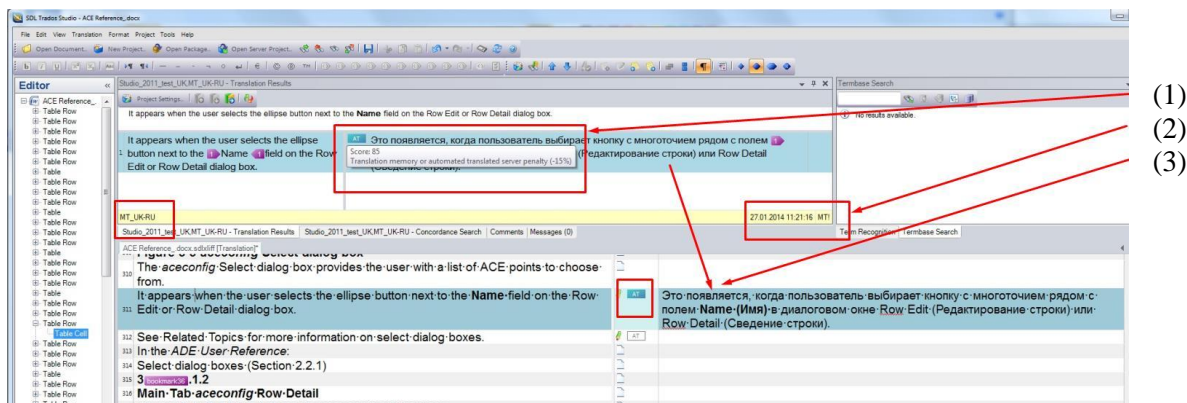
Примечание: на рисунке видно, что при импорте трех машинно переведенных сегментов произошли ошибки. В данный момент достоверной информации о возможных причинах ошибок у меня нет. Такие ошибки не являются критическими. Они просто означают, что три сегмента не были импортированы в конкретную память ТМ, и вам будет нужно перевести эти сегменты вручную.

SDL Trados Studio



1. Выберите **Project Settings**.
2. Нажмите на кнопку **Add** для открытия или создания рабочей памяти ТМ. Переместите эту память наверх и установите флажок **Update** (в ходе постредактирования данная память будет обновляться).
3. Установите штрафы для временной памяти, которая содержит машинно переведенные сегменты (например, **15%**).
4. Снимите для этой памяти флажок **Update** (данная память нужна только на этапе постредактирования, и обновлять ее не нужно; по окончании постредактирования ее содержимое можно удалить).

SDL Trados Studio



1. Начните постредктировать исходный файл. Если в рабочей памяти ТМ нет сегментов с уровнем совпадения выше, чем для машинного перевода (то есть, **выше 85%**), то в целевой сегмент подставляется машинно переведенный вариант из временной памяти ТМ – с **пенальти 15%**).
2. PROMT назначил всем подобным сегментам идентификатор **MT!**
3. На рисунке видно, что в данном сегменте сохранены теги.

Заключение

1. В результате выполненной процедуры временная память ТМ содержит все двуязычные сегменты для переводимого документа, которые ранее относились к категории «неизвестные» (**No match**). Сегменты, которые на этапе анализа файла в базе ТМ отсутствовали, были созданы PROMT-ом. Таким образом, пакеты **SDL Trados Studio 2011** и **PROMT LSP 9.5** встроены в технологический процесс перевода.
2. Данная процедура не является автоматической. По моему опыту работы это не является существенным недостатком, так как существует возможность предварительной проверки машинного перевода и, при необходимости, донастройки ресурсов PROMT-а для обеспечения нужного качества машинного перевода.

Дополнения

Наш опыт работы показывает, что наличие в отдельной памяти ТМ машинно переведенных сегментов предоставляет ряд преимуществ:

- с помощью функции **Concordance** вы можете искать текст не только в ранее отредактированных сегментах (**High matches** и **Low matches**), которые уже имеются в памяти ТМ, но и в сегментах, которые еще не отредактированы. Это упрощает и ускоряет поиск релевантной информации в базах ТМ (в дополнение к поиску в оригинальном документе(ах)),

- после подстановки машинно переведенного сегмента из памяти ТМ (с нужной лексикой и элементами стиля) вы экономите значительное время (в среднем, 40-50%) при пост-редактировании и изменении целевого текста (вам не нужно вручную набирать весь целевой текст с нуля, ИЛИ вам нужно гораздо меньше копировать нужный текст из исходного предложения в целевое),

- благодаря предварительному анализу содержимого переводимого документа, разработке релевантной терминологии и настройке системы машинного перевода, вы будете тратить меньше времени на поиск нужных терминов (они уже имеются в машинно переведенных сегментах).

Благодарственное слово

Ниже указаны мои коллеги, которые предоставили полезные комментарии и рекомендации. Уважаемые друзья! Спасибо вам большое:

Кис Хайналка (info@kiss2language.com)

Сергей Савельев (saveliev.v.sergey@gmail.com)