

Machine Translation Post-Editing Interfaces

1. Introduction
2. Online machine translation engines
 - a) Google Translate
 - b) Systranet
 - c) Windows Live Translator
 - d) SDL freetranslation
 - e) PROMT
 - f) Language Weaver
 - g) Conclusions post-editing in online MT engines
3. Desktop machine translation engines
 - a) Systran Premium 6
 - b) LINGUATEC Personal Translator 2008 Professional
 - c) Comparative survey
 - d) Conclusions post-editing in desktop MT engines
4. Translation memory editor: Olifant (Okapi tools)
5. Conclusions for an outline of the PaCo interface

1. Introduction

Machine translation is coming of age, and especially the hybrid (rule based/SMT) systems seem to have a brilliant future.

We will examine and assess the interfaces of 6 online engines: some purely SMT, some purely rule-based and some hybrid. The interfaces are quite rudimentary and do not allow for much user interaction, because of various reasons:

- they are free of charge
- direct user intervention for SMT is nonexistent
- direct user intervention for rule-based systems is possible, but the user has to be well-versed in grammatical categories

We will also explore the interfaces for 2 desktop MT engines: they allow for more interaction and post-editing but are purely rule-based.

PaCo intends to combine the best of both worlds: better translation results of statistical engines and post-editing possibilities of rule-based systems. We hope that this hybrid approach based on parsing of statistical raw data will allow us to ensure “more natural” translations which, in the short term, can be optimised by extensive post-editing functionality.

Koen De Smet
Business Manager Oneliner bvba

2. Online machine translation engines

a. Google Translate: <http://translate.google.com/>

[Web](#) [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more ▼](#)

Google
Translate BETA

Home [Text and Web](#) [Translated Search](#) [Dictionary](#) [Tools](#)

Translate text or webpage

Enter text or a webpage URL.

 > [swap](#)

Search other languages


Enter a search phrase in your own language to find information in other languages.
[Translate and search](#)

Translation tools

Get translation with a single click and make your webpage instantly available in other languages.
[Get translation tools now](#)

Languages available for translation

| | |
|------------|------------|
| Arabic | Italian |
| Bulgarian | Japanese |
| Catalan | Korean |
| Chinese | Latvian |
| Croatian | Lithuanian |
| Czech | Norwegian |
| Danish | Polish |
| Dutch | Portuguese |
| English | Romanian |
| Filipino | Russian |
| Finnish | Serbian |
| French | Slovak |
| German | Slovenian |
| Greek | Spanish |
| Hebrew | Swedish |
| Hindi | Ukrainian |
| Indonesian | Vietnamese |

 [Google Home](#) - [About Google Translate](#)
©2009 Google

[Web](#) [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more ▼](#)

Google
Translate BETA

Home [Text and Web](#) [Translated Search](#) [Dictionary](#) [Tools](#)

Translate text or webpage

Enter text or a webpage URL.

 > [swap](#)

| | |
|--|--|
| Number of language combinations | 34 * 34 |
| Automatic source language detection | yes |
| Language swap (reverse source and target language) | yes |
| Text length limitation | yes (none documented, but limited to 30K) |
| Dictionary search | yes |
| Text formats | Html, txt; more recently a feature to enter a link to an online module was added (all office formats supported) for converting the document into html, the result of which is translated HTML. |
| URL translation | Yes |
| Interaction | No interaction at all, but in its future Translation Centre, Google will offer to the translators' community a virtual translation environment with machine translation support. The translators will be able to edit the machine output, and Google will use their adaptations in order to optimise and fine-tune its translation engine; this will be the first systematic use of user feedback. |

b. Systranet: <http://www.systranet.com/>

The screenshot shows the SYSTRANet website interface. At the top left is the SYSTRANet logo and the text "Your free online translator". At the top right, it says "You are logged in as kdesr". Below the header is a navigation bar with tabs: "Text", "Web Page", "Dictionary", "RSS", "File", and "My Dictionary". The "Text" tab is selected. Below the navigation bar, there are two dropdown menus for language selection, currently set to "Dutch" and "English", with a double-headed arrow between them. To the right of these menus is an "Options:" section with several icons: a calendar, a document, an envelope, a printer, and a speech bubble. The main area of the page is a large, empty white box for text input. At the bottom right of this area is a black button labeled "Translate".

The screenshot shows the SYSTRANet website interface with the "My Dictionary" tab selected. The header and navigation bar are the same as in the previous screenshot. Below the navigation bar, there are two dropdown menus for language selection, currently set to "Dutch" and "English", with a double-headed arrow between them. Below these menus is a section for dictionary settings. It includes a "Dutch" label with a question mark icon and an empty input field. Next to it is a "DNT" checkbox, which is unchecked, followed by an "English" label and another empty input field. To the right of these is a "Part of Speech" label and a dropdown menu currently set to "Detect". There is also a small icon to the right of the dropdown menu. The main area of the page is a large, empty white box for text input.

| | |
|--|---|
| Number of language combinations | 52 |
| Automatic source language detection | no |
| Language swap (reverse source and target language) | no |
| Text length limitation | 100k keystrokes |
| Dictionary search | yes |
| Text formats | doc, docx, rtf, html, htm, txt |
| URL translation | yes |
| Interaction | My dictionary: interactive tool which the user can use to input his own word translations with part of speech and DNT (Do Not Translate) "negative" list. Choose alternative meanings matching the context. |

c. Windows Live Translator: <http://www.windowsslivetranslator.com/>

The screenshot shows the Windows Live Translator website. At the top, there is a blue header with the Windows Live logo and navigation links for Start, Hotmail, Spaces, OneCare, and MSN. A search bar with a 'Zoeken' button is also present. Below the header, the page title is 'Translator' and a subtitle reads 'Gratis online vertalingen waarmee het web echt wereldwijd wordt'. A link 'Toevoegen aan Favorieten' is visible in the top right.

The main content area is divided into two columns. The left column is titled 'Voer een tekstfragment in (tot 500 woorden)' and contains a text area with the instruction: 'Klik hier om de tekst die moet worden vertaald te typen of te plakken. Zorg ervoor dat de juiste talencombinatie is geselecteerd en klik op de knop Vertalen.' The right column is titled 'Vertaalde tekst' and contains the translated text: 'Automatisch vertalen helpt om de strekking van een tekst in een vreemde taal te begrijpen maar is niet geschikt om een professionele vertaler te vervangen als een foutloze en goed lopende tekst is vereist.'

Below the text areas, there is a link 'Alles wissen' and a note 'Vertaling door Microsoft® Translator'. At the bottom of the main area, there is a section 'Of vertaal een webpagina' with a text input field containing 'http://', a language dropdown menu set to 'Engels - Nederlands', and a 'Tekst vertalen' button.

On the left side of this section, there are several links: 'Nieuw: Assistent voor Engels als tweede taal', 'Translator-knop voor Windows Live Toolbar', 'Translator voor webpagina's toevoegen aan eigen website', and '+ Veelgestelde vragen over Translator'. On the right side, there is an 'Introductie' section featuring a blue robot icon labeled 'TBot' and the text: 'Vertaalbot voor Windows Live Messenger. Voeg mtbot@hotmail.com toe als contactpersoon en nodig TBot uit voor Messenger-gesprekken.'

The footer contains copyright information '© 2009 Microsoft', links for 'Privacyverklaring' and 'Gebruiksrechtsovereenkomst', and support links for 'Help voor Windows Live Translator', 'Help', 'Accounts services', and 'Feedback'.

| | |
|--|--|
| Number of language combinations | 28 |
| Automatic source language detection | no |
| Language swap (reverse source and target language) | no |
| Text length limitation | 500 words |
| Dictionary search | no |
| Text formats | html, txt |
| URL translation | yes |
| Interaction | You may either submit a one-click star rating on the Translator home page , or use the Feedback link in the footer of the home page and the Bilingual Viewer. If non-offensive text has resulted in an offensive translation by mistake, you may also report this by clicking on the link "Report offensive translations" which, like the star rating option, is available on the Translator home page after a translation has been submitted. |

d. SDL freetranslation: <http://www.freetranslation.com/>

Click2Translate.com SDL

SDL | FreeTranslation.com English

Home | About | Shop | More Services | Help & FAQ | Blog | Platinum Club | Your Account SHARE

 **SDL FreeTranslation.com** is a [free translation website](#) allowing you to [translate web pages](#) and text into different languages. For [professional, human translations](#) see [Click2Translate.com](#)

Free Text Translation | Free Web Translation | Professional Translation

English to Spanish **Advanced Translation Tools**

Type or paste some text here.

a e i o u Other

Human Translation **FREE Translation**

Need Professional, Human Translation?
Click2Translate is the world's leading professional online translation site, offering high quality language translations.

ING Autoverzekering
't is waar, de prijs is ongelooflijk.
ING

New Chat Translator

Translate while chatting online with Chat Translator
Instantly translate your online conversations with SDL Chat Translator.

Fully integrated with AOL AIM 6.8, you can chat to your buddy and translate into their language, and they can chat to you in yours.

NEW!

[Find out more](#)

Desktop Translator

Instantly translate with Desktop Translator
Quickly translate whole Word documents, emails and highlighted text seamlessly with [Desktop Translator](#).

All text is translated using specialized dictionaries, and any formatting within the document is kept.

[Find out more](#)

More Translation Gadgets

SDL Desktop Translator
Translates whole documents & emails from Word & Outlook

SDL Clipboard Translator
Copy text from anywhere, paste translated text back.

More Translation Gadgets
Other online and downloadable translation tools for your websites or blogs

[Corporate Solutions](#) | [Professional Translation](#) | [Document Translation](#) | [Website Translation](#) | [Localization Services](#) | [Translation Software](#) | [Currency Converter](#) | [More...](#)

Want to earn money from your website? [Sign up to become an affiliate.](#)

[About FreeTranslation.com](#) | [Site Map](#) | [About SDL](#) | [Site Ownership](#)

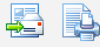



| | |
|--|--|
| Number of language combinations | 13 (from and into English) |
| Automatic source language detection | no |
| Language swap (reverse source and target language) | no |
| Text length limitation | 750 words |
| Dictionary search | Only in paid platinum Specialized Dictionaries Improve the quality of your "Business", "Current Affairs" and "IT" translations by using our regularly updated specialized dictionaries. SDL monitors the usage of these specialized dictionaries for untranslated words and makes regular updates. |
| Text formats | Txt (copy/paste only; paid desktop translator available) |
| URL translation | yes |
| Interaction | Options only Advanced Translation Tools Some languages have different ways of addressing people depending on the conversation situation. For instance, you would use the informal way of speaking in a family situation and the formal way in a business situation. |

e. PROMT: <http://www.promt.com/>

Text Translator | Website Translator | Desktop Translators | [m.online-translator.com for mobile](#)

[< Simple mode](#) | [Advanced mode](#)

Show variants of translation
 Transliterate unknown words

Translation | **ABC** Spelling |  Result |  Dictionary |  Keyboard |  Help

Source text | **Translated text**


Type or paste your text here

Translation Software

[Compare translators](#)
[Download translator](#)


Online Store

@promt Expert 8.0



\$ 535
[Buy now >](#)

English to Russian translation
General lexicon

 Translate

| | |
|--|---|
| Number of language combinations | 24 |
| Automatic source language detection | no |
| Language swap (reverse source and target language) | no |
| Text length limitation | 3,000 keystrokes (10,000 for registered users) |
| Dictionary search | Yes |
| Text formats | txt (copy/paste only) |
| URL translation | yes |
| Interaction | <p>The “Translation options” group includes the following settings:</p> <p>“Show variants of translation” (if a word has several meanings, alternative translations are displayed in brackets)</p> <p>“Transliterate unknown words” (words not found in the dictionaries are transliterated from Cyrillic into Latin script or vice versa).</p> <p>Spell-checking before translation may improve the quality of the final result by eliminating spelling errors in the source text. The results of the spellcheck are displayed, along with suggestions to replace unknown or misspelled words.</p> |

f. Language Weaver: www.languageweaver.com

No free online module available; access requested to demo but not granted

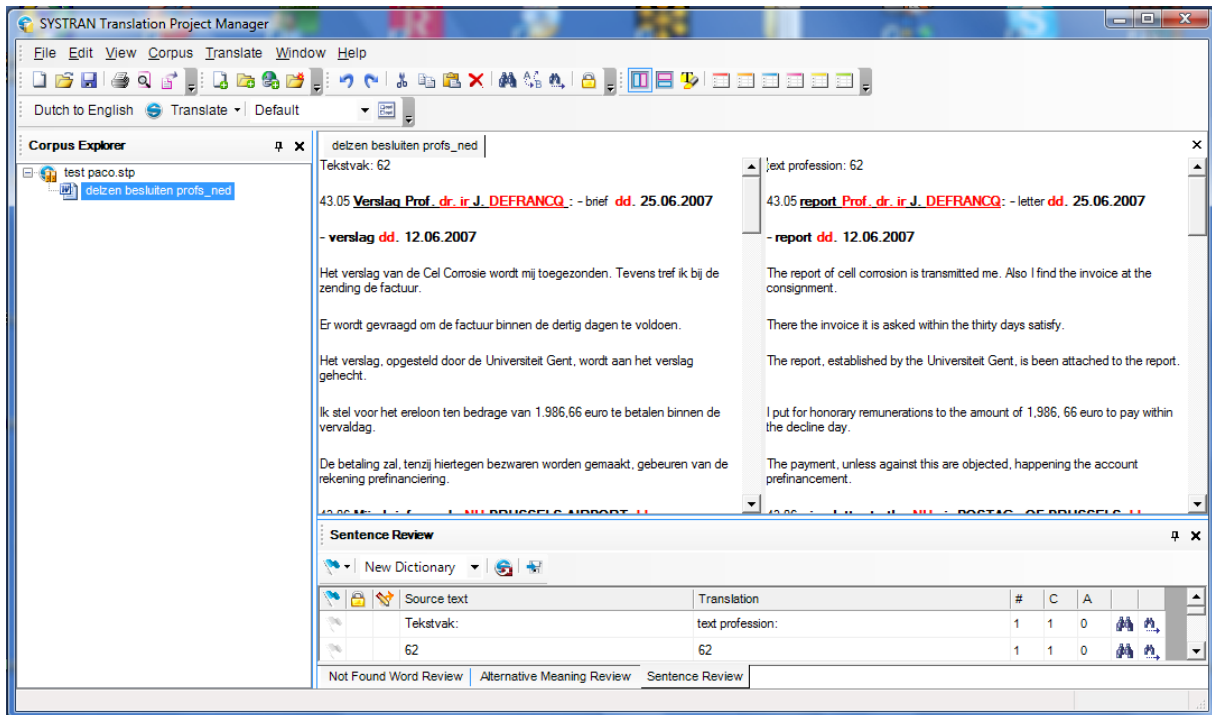
| | |
|---------------------------------|----|
| Number of language combinations | 68 |
|---------------------------------|----|

g. Conclusions post-editing in online MT engines

Post-editing in online MT applications is virtually nonexistent. Only Systran offers an interactive tool which the user can use to input his own word translations with part of speech and DNT (Do Not Translate) "negative" list. The user can also choose alternative meanings matching the context. Windows Live Translator allows for feedback and ratings of translation quality. Google Translation Center (not online yet) promises to allow human translators to review its translation results in the hope of improving its statistical engine by incorporating these human reviews for training purposes.

3. Desktop machine translation engines

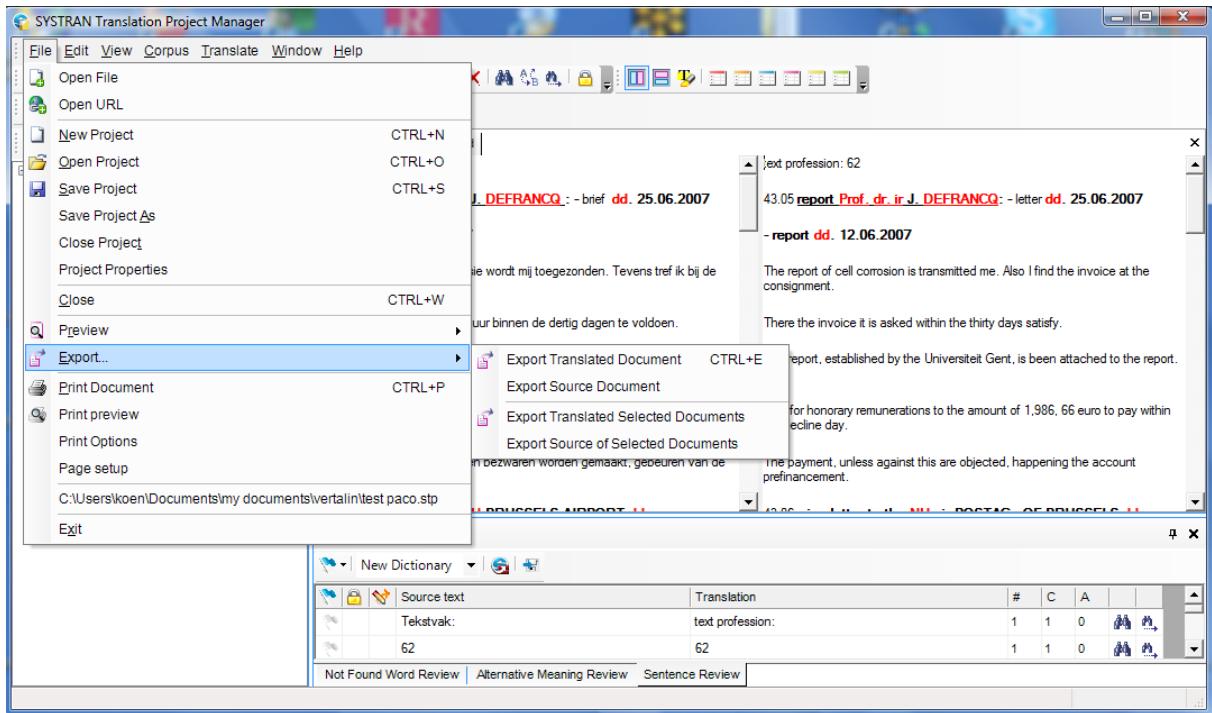
a. Systran Premium 6



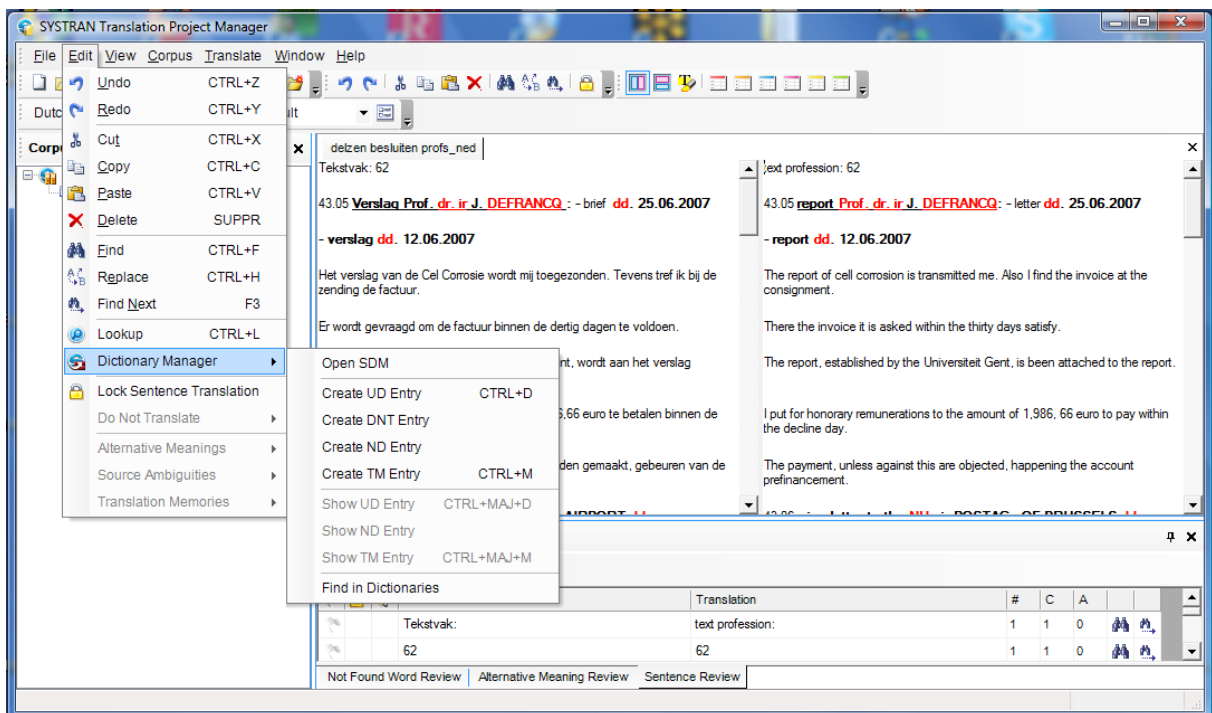
Side-by-side alignment of source text and translation. Marking in red of unknown segments.



Systran bar in ms word



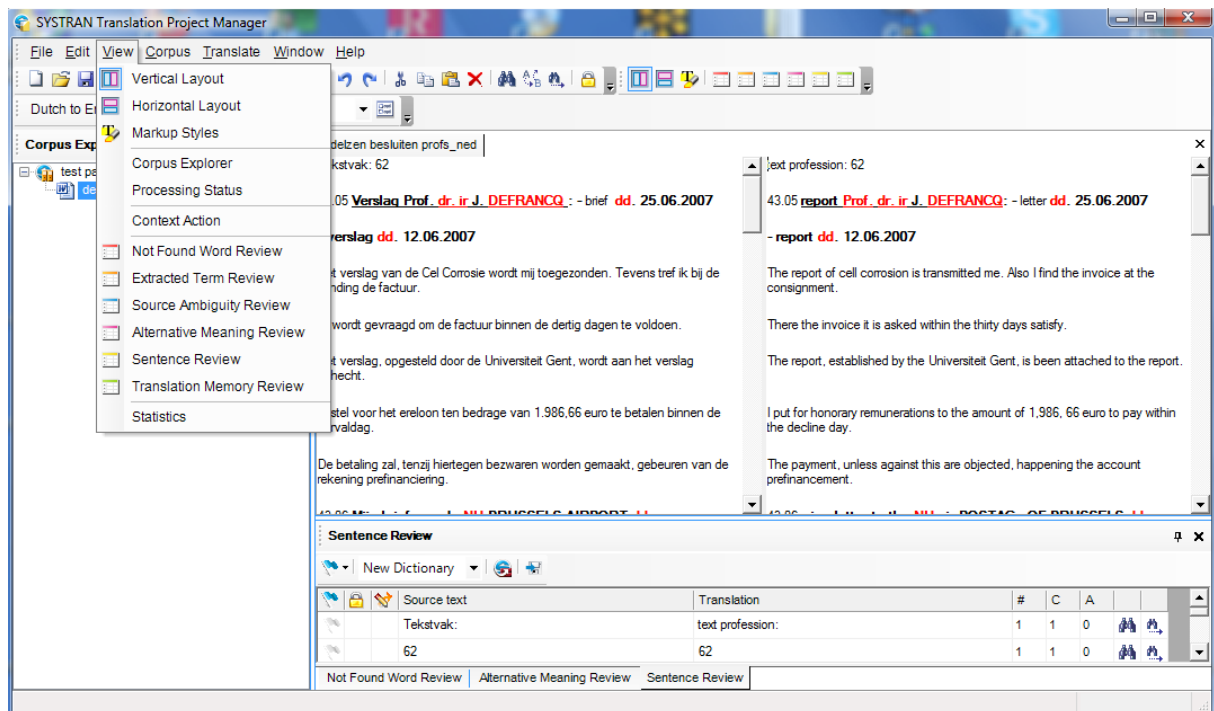
Export feature for source and target texts.



Dictionary manager:

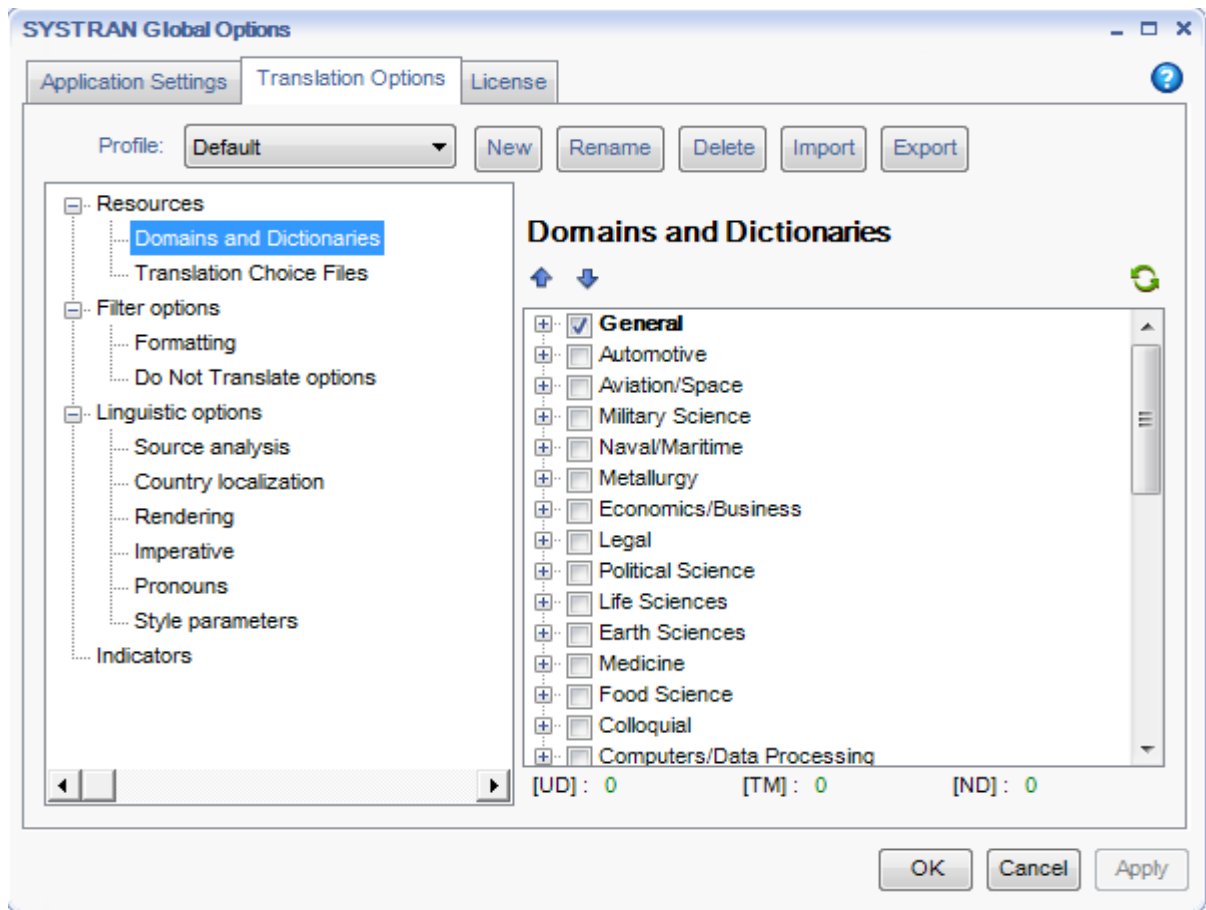
- SDM: Systran Dictionary Manager
- UD: User Dictionary
- DNT: Do Not Translate
- TM: Translation memory

Also DNT and TM are regarded as dictionaries.

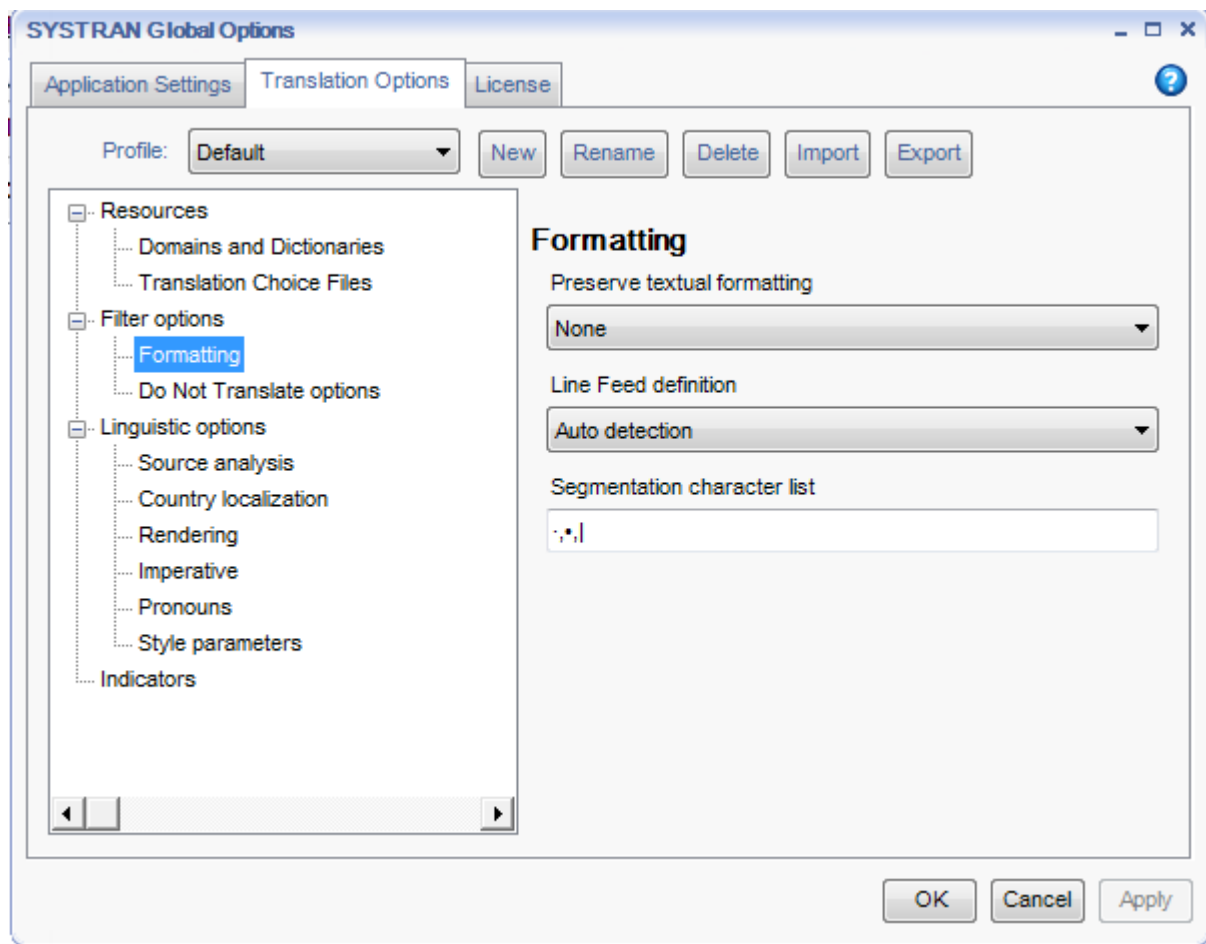


Marking of:

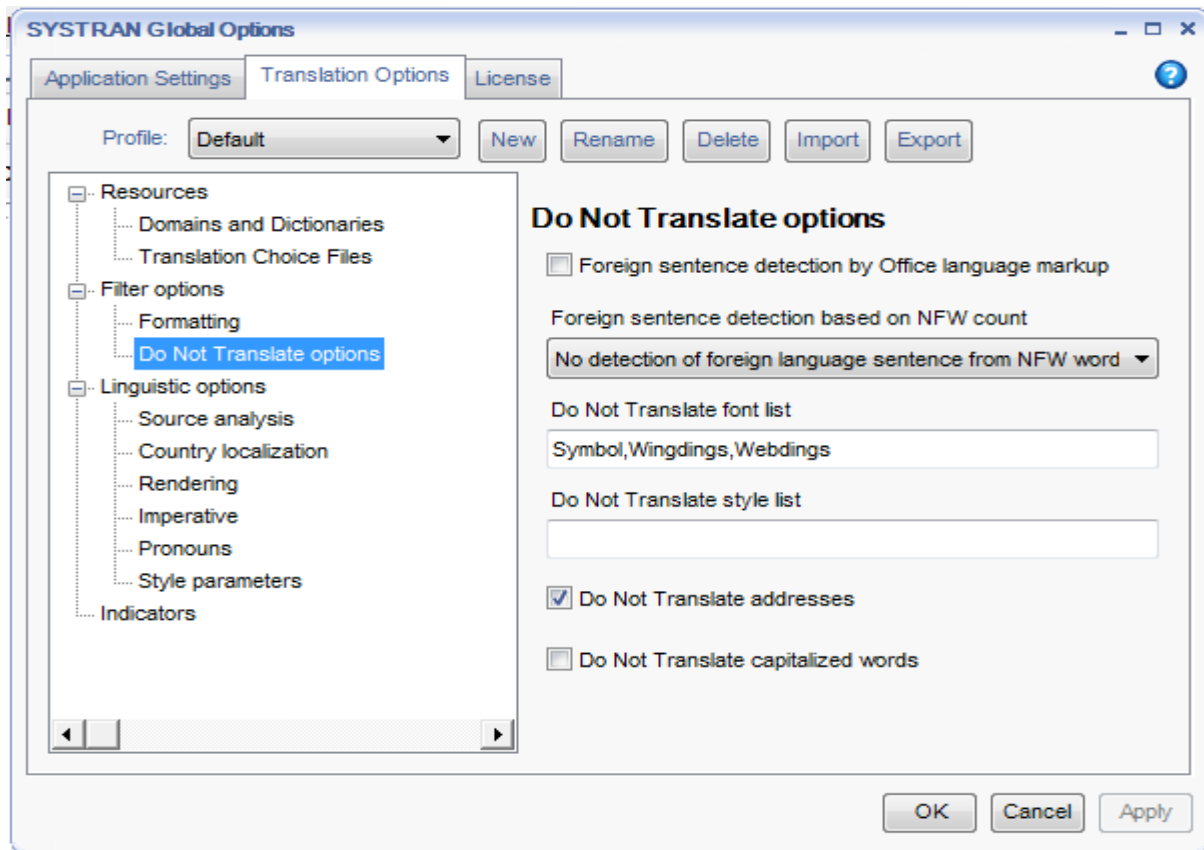
- Words not found
- Ambiguous source segments
- Alternatives meanings
- TM segments



Specialized dictionaries can be combined.

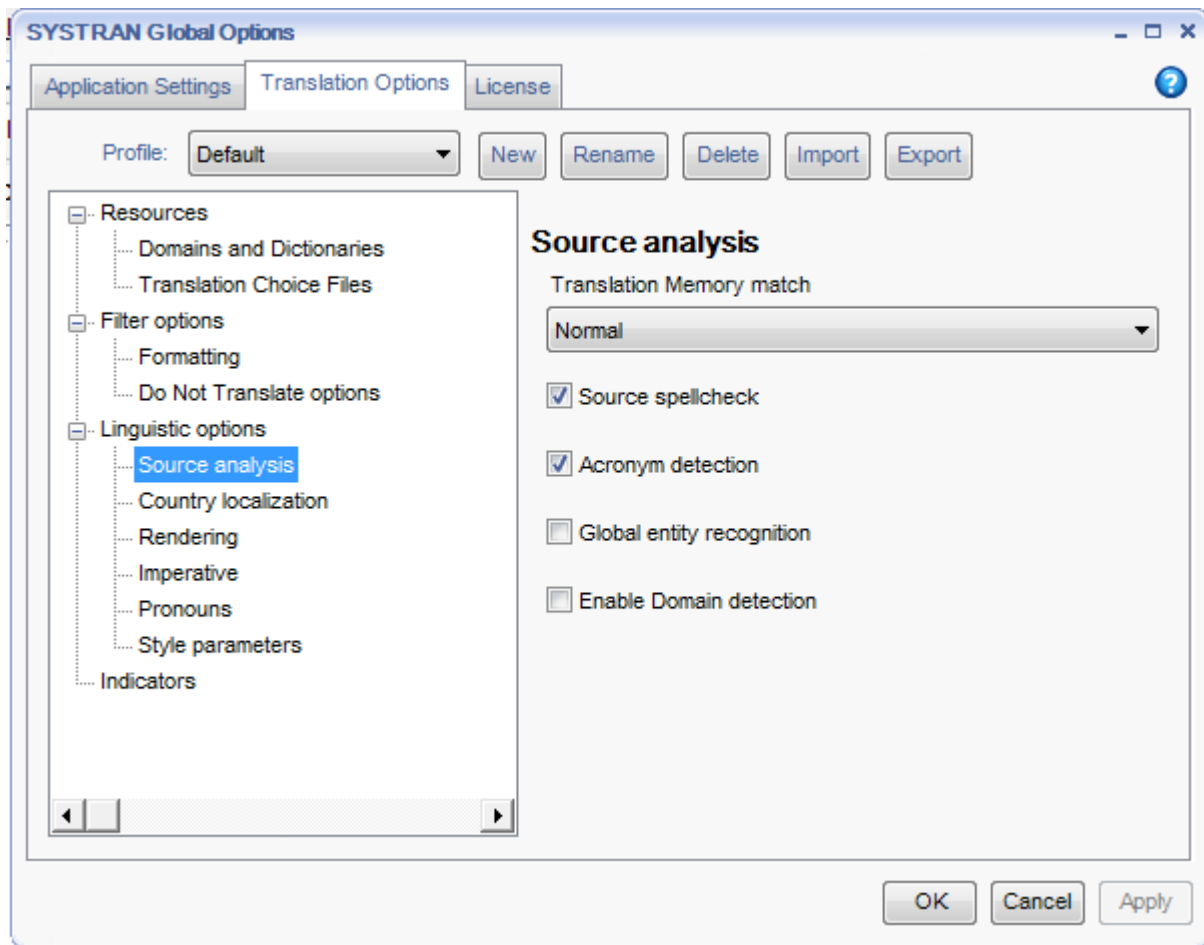


Textual formatting options and segmentation character list can be edited



Do Not Translate options:

- foreign words (automatic detection option)
- certain fonts
- certain styles
- addresses
- capitalized words



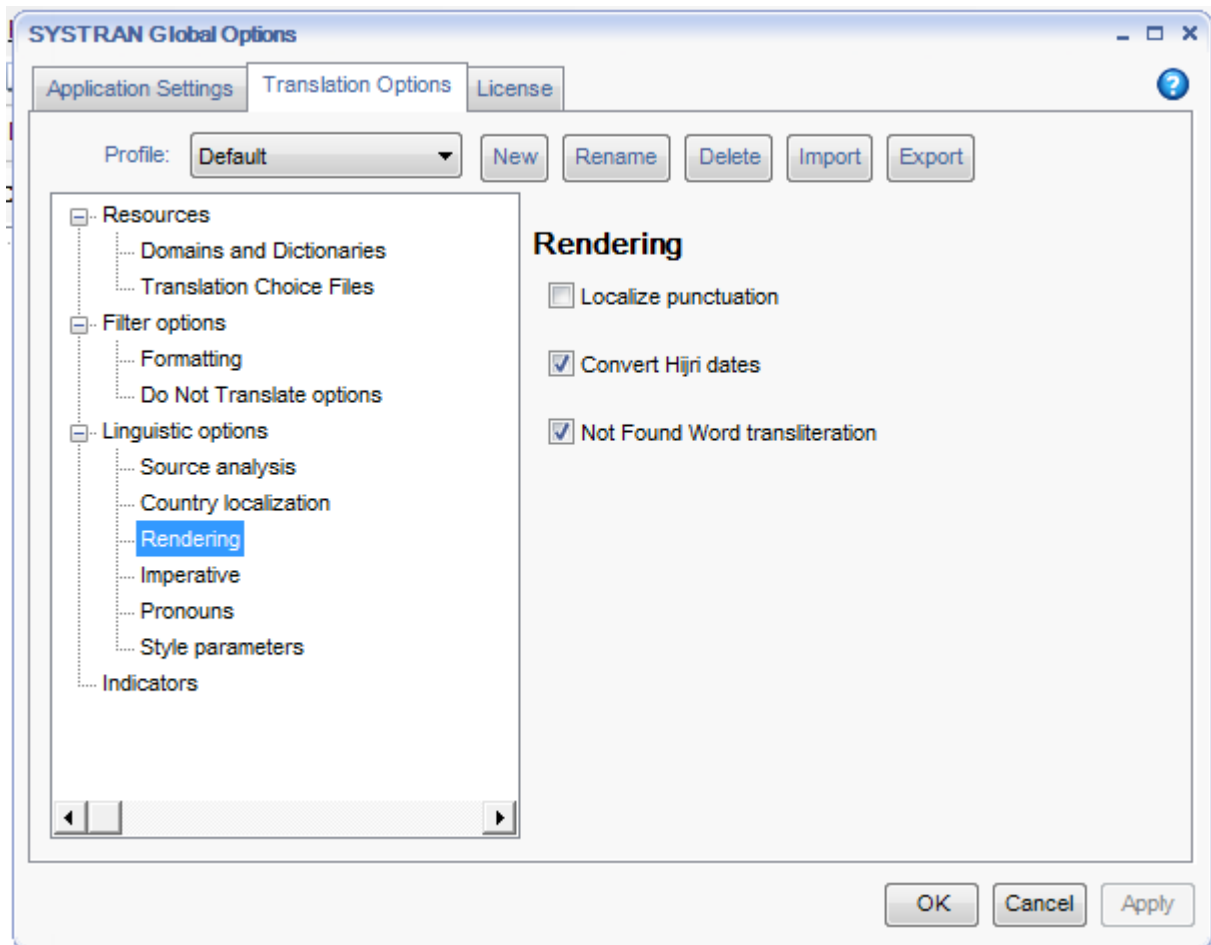
Linguistic options - Source analysis:

- translation memory matching
- spellcheck
- acronym detection
- domain detection

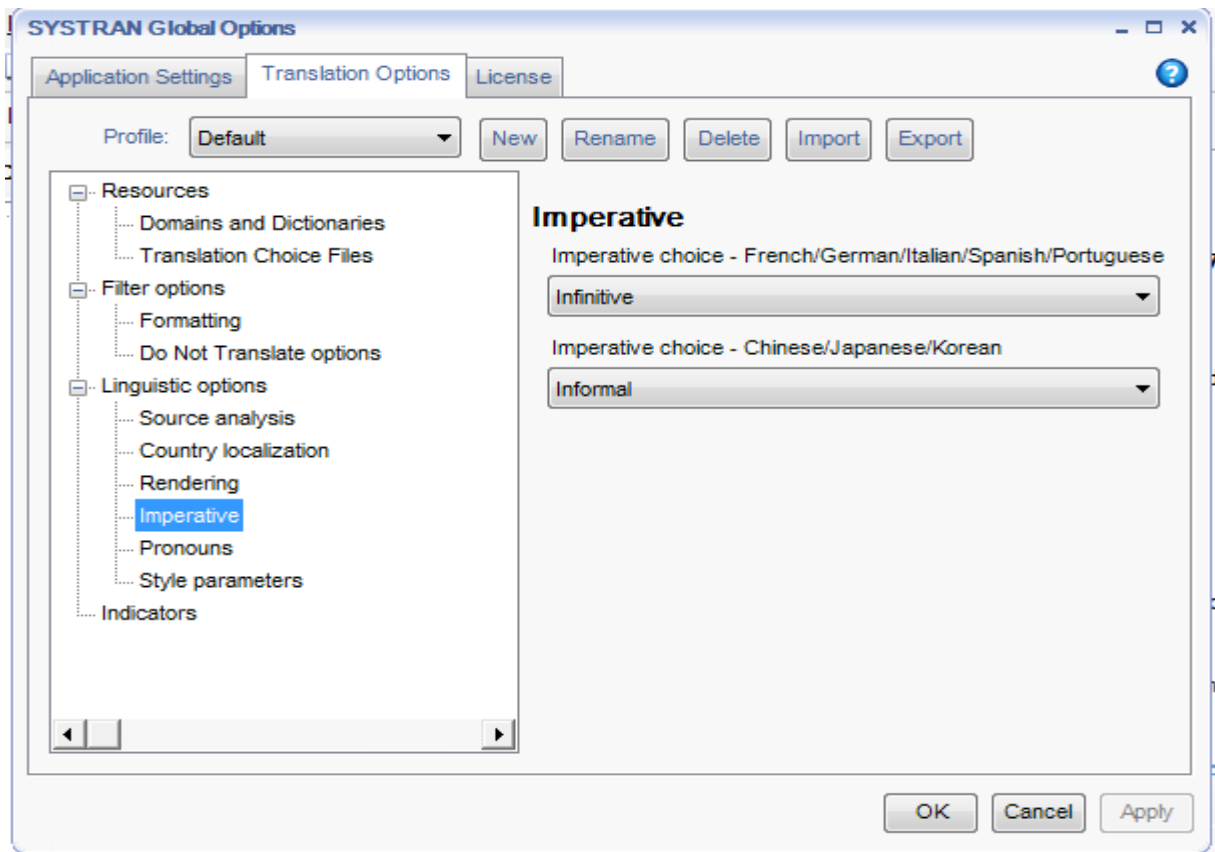


Linguistic options – Country localization:

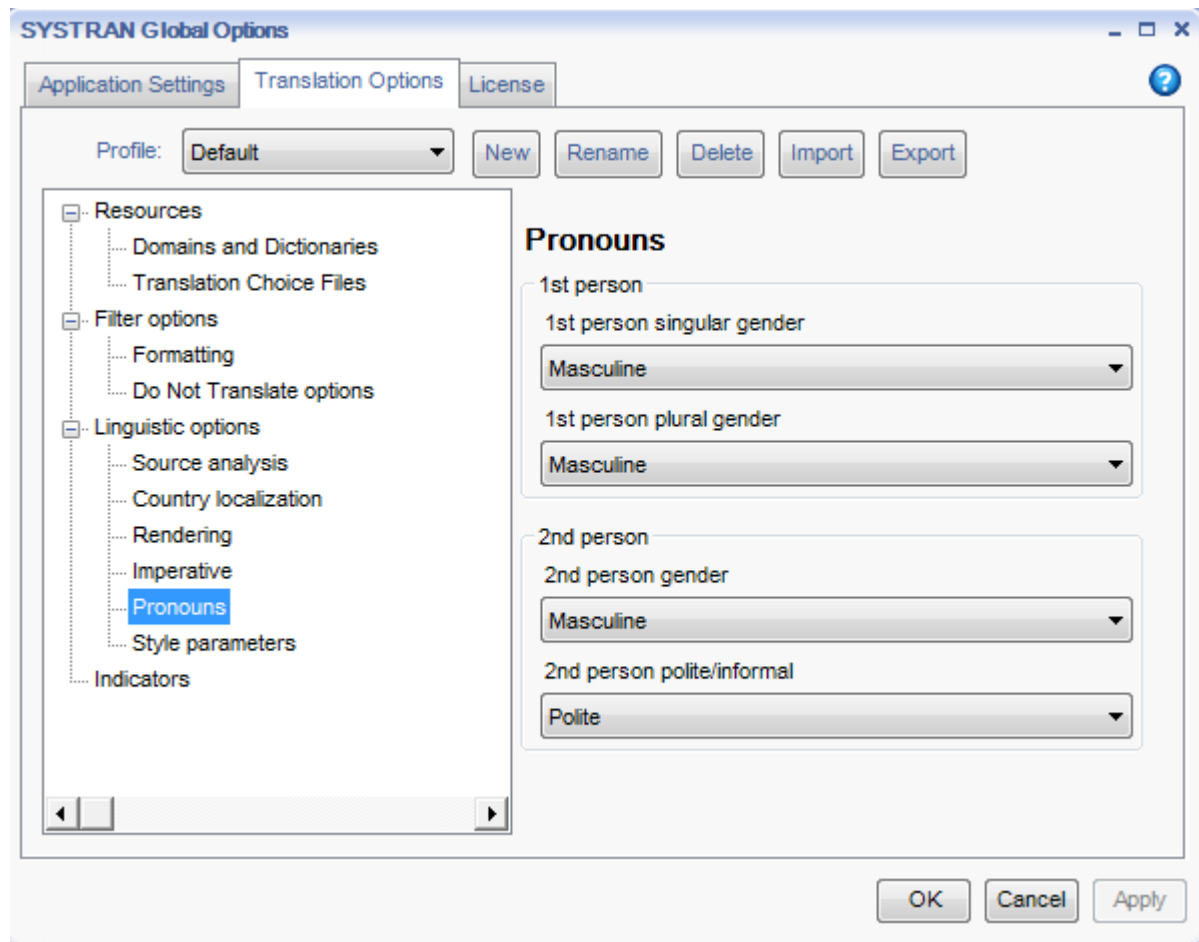
- regional variants
- sociolinguistic variants
- new or old spelling



Linguistic options – Rendering:
Local punctuation (e.g. numbers Anglo-Saxon vs Continental)

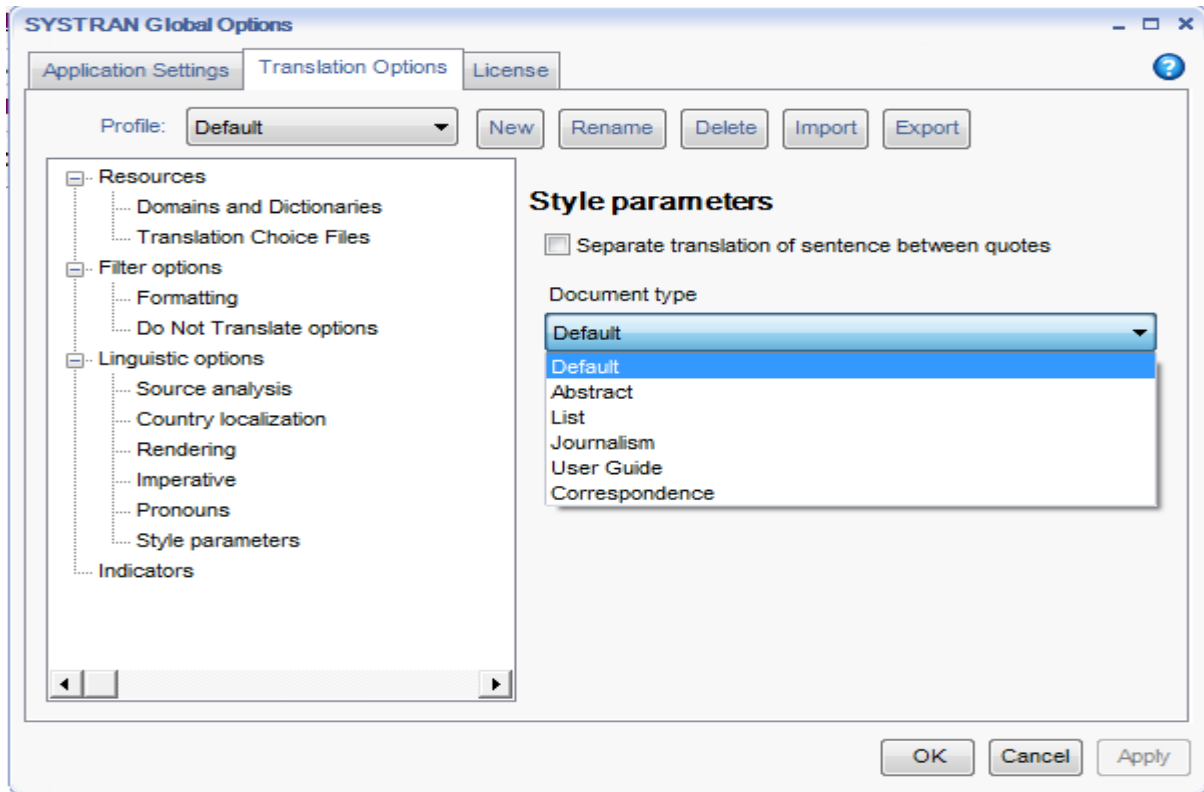


Linguistic options – Imperative or infinitive (for certain European languages)

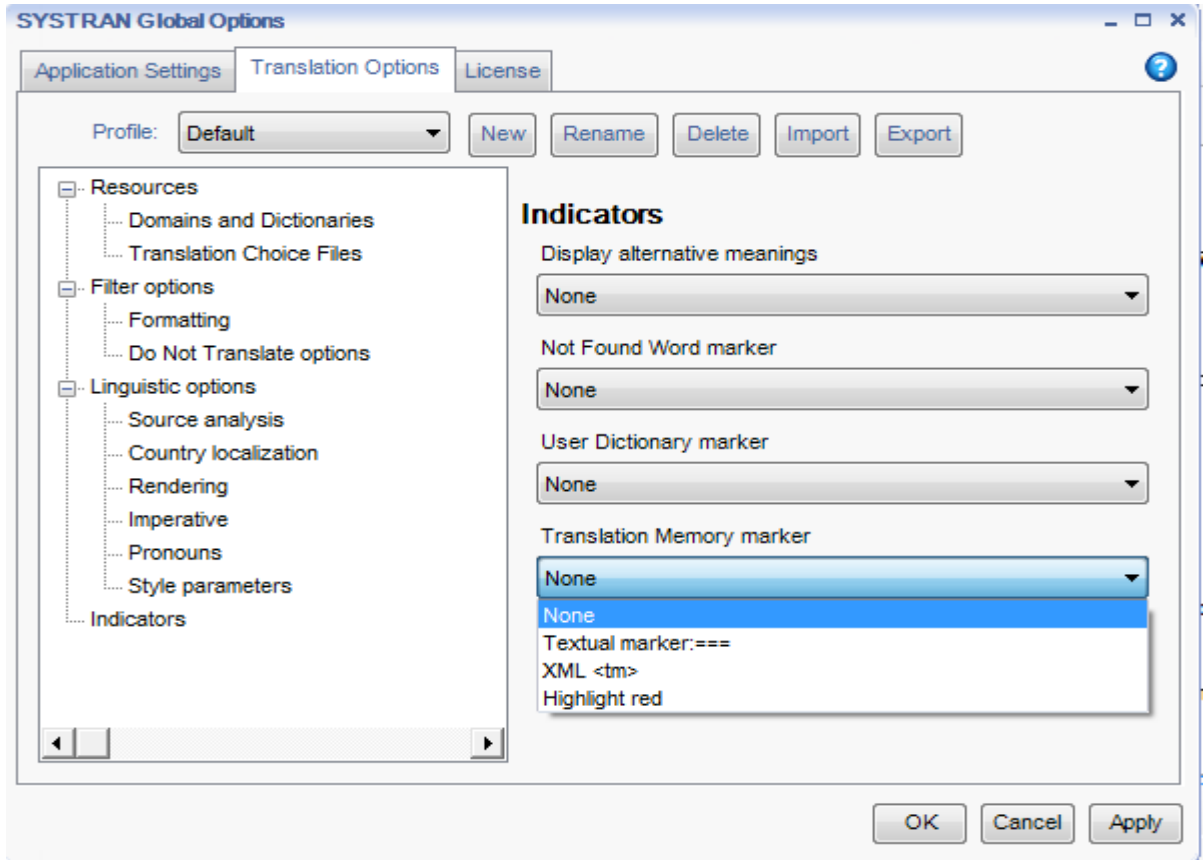


Linguistic options – Pronouns

- gender
- register (informal – polite...)

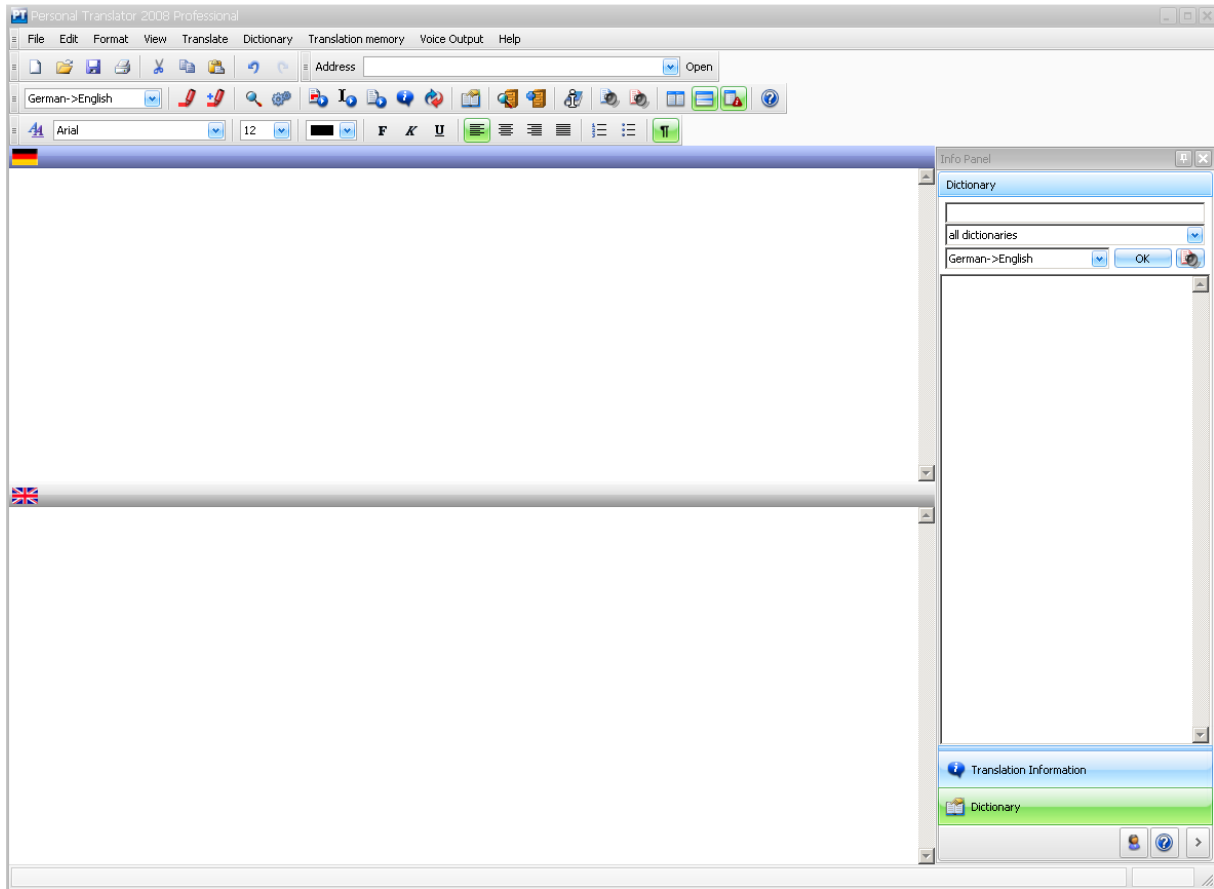


Style parameters and option to separate translation of sentence between quotes

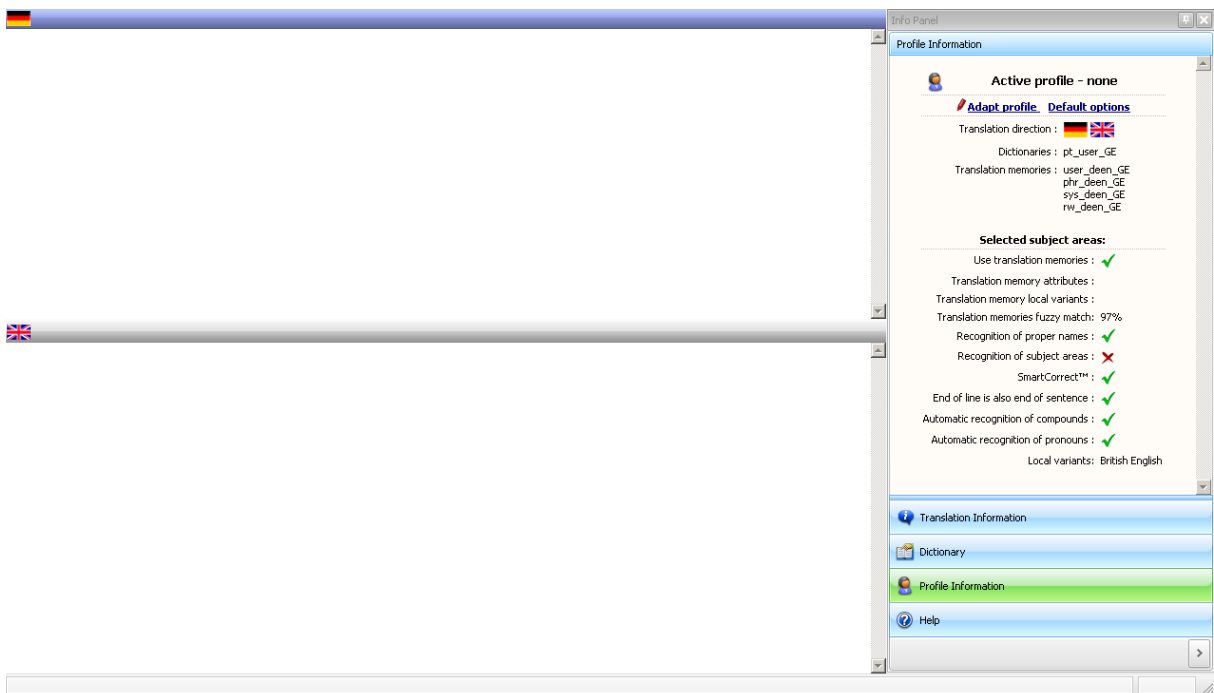


Markers for alternatives – dictionary entries - TM

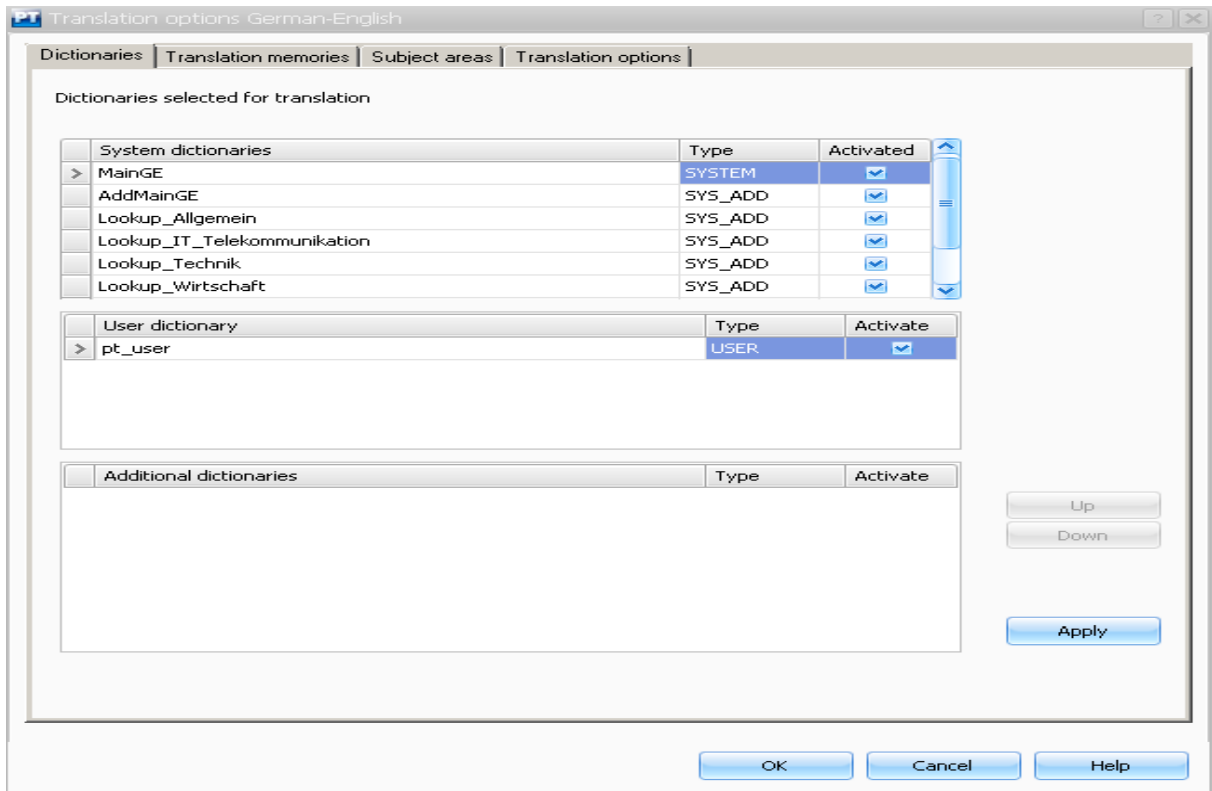
b. LINGUATEC Personal Translator 2008 Professional



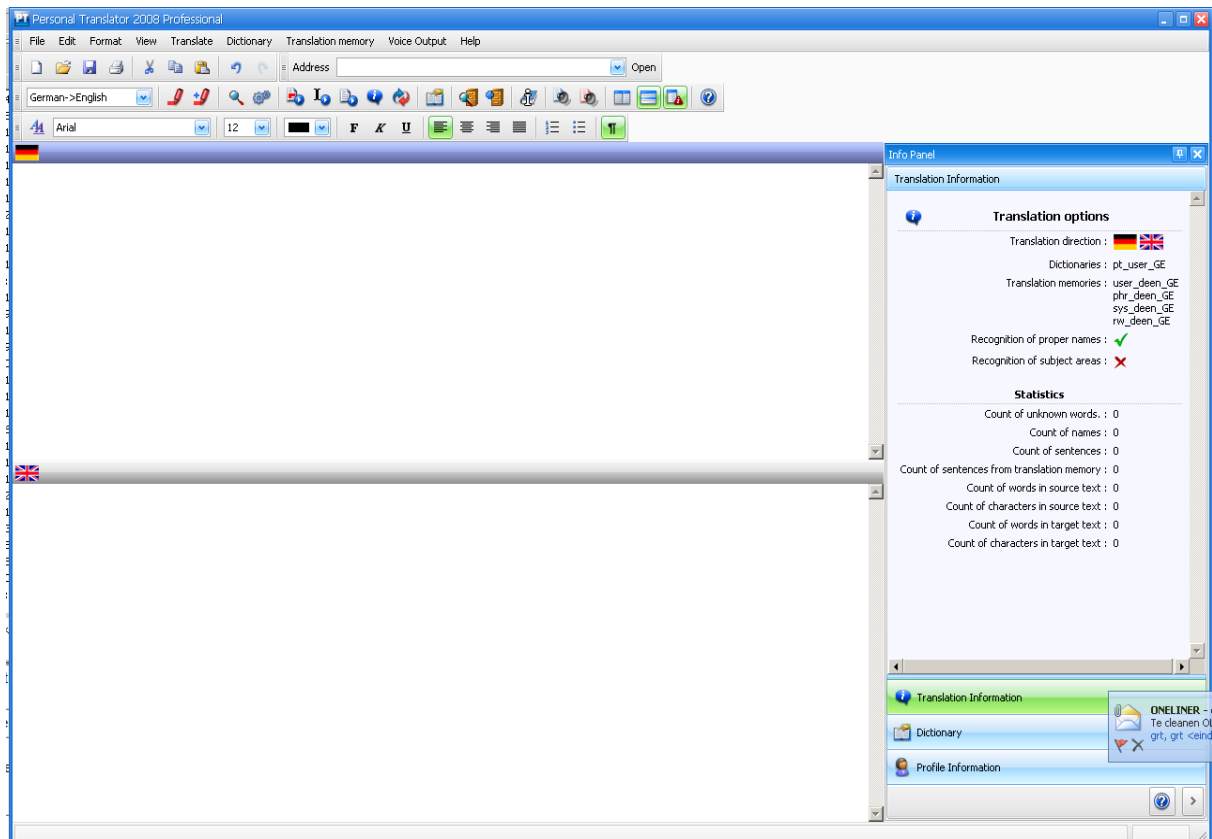
General survey of the user interface, with in the right pane all relevant meta-information: dictionary



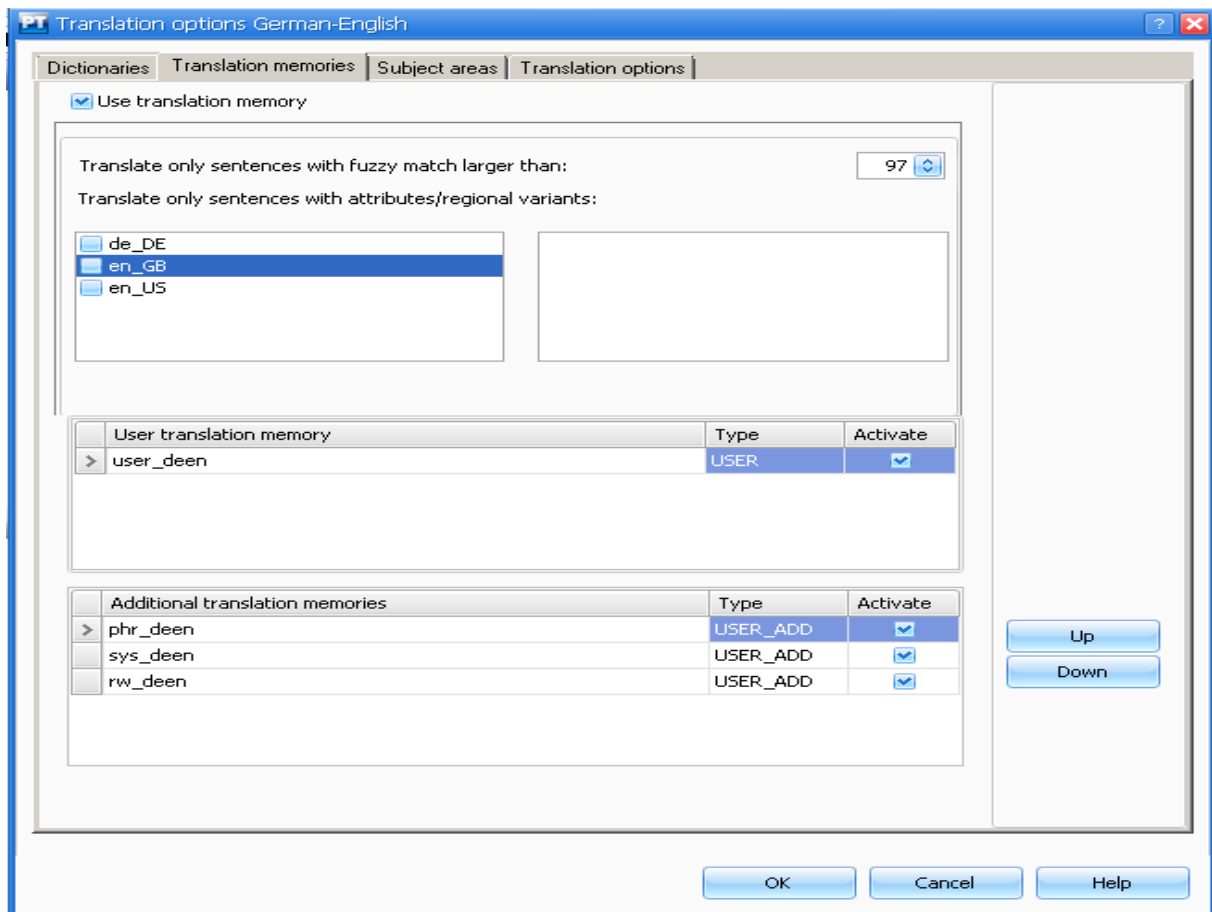
Right pane: profile options



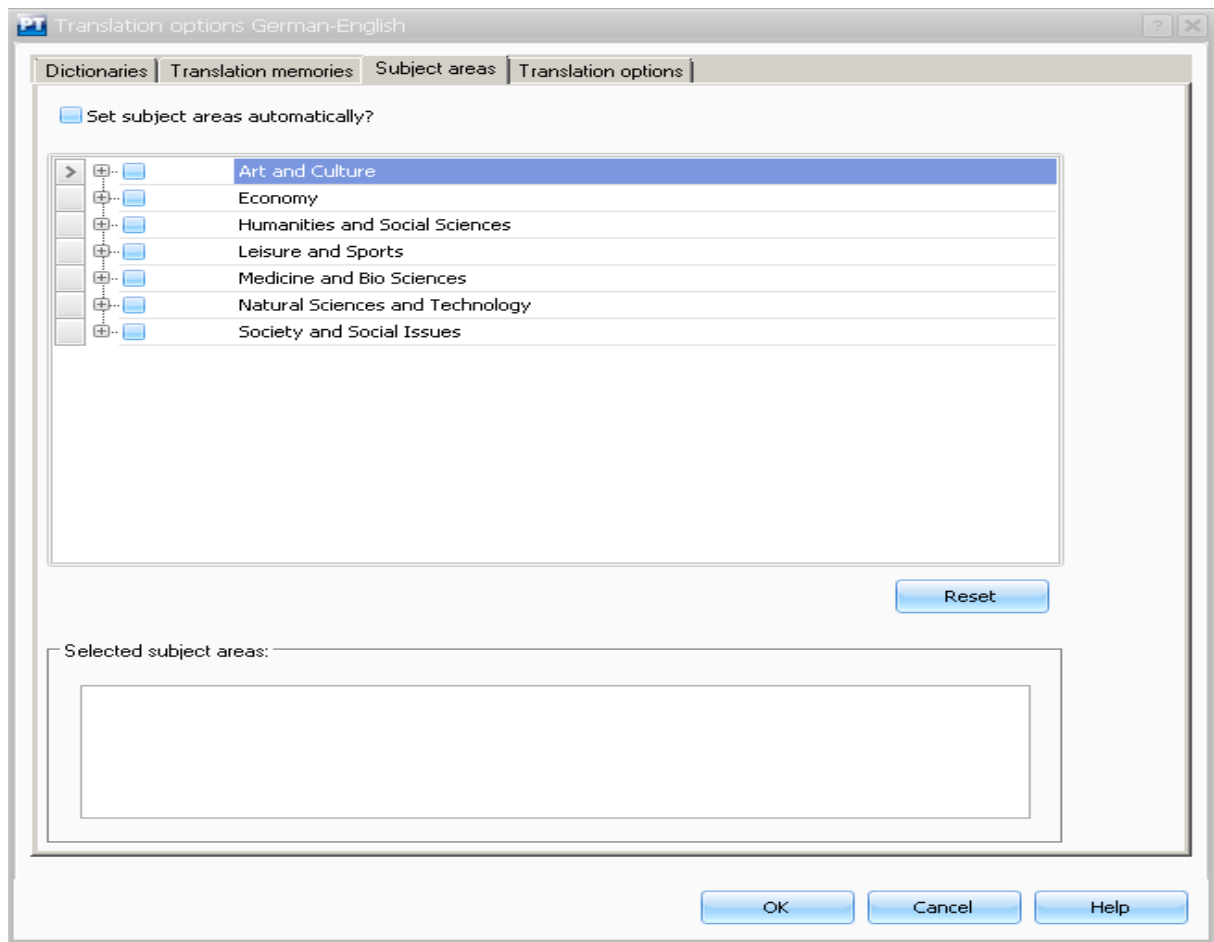
Dictionary options: combination of specialized dictionaries



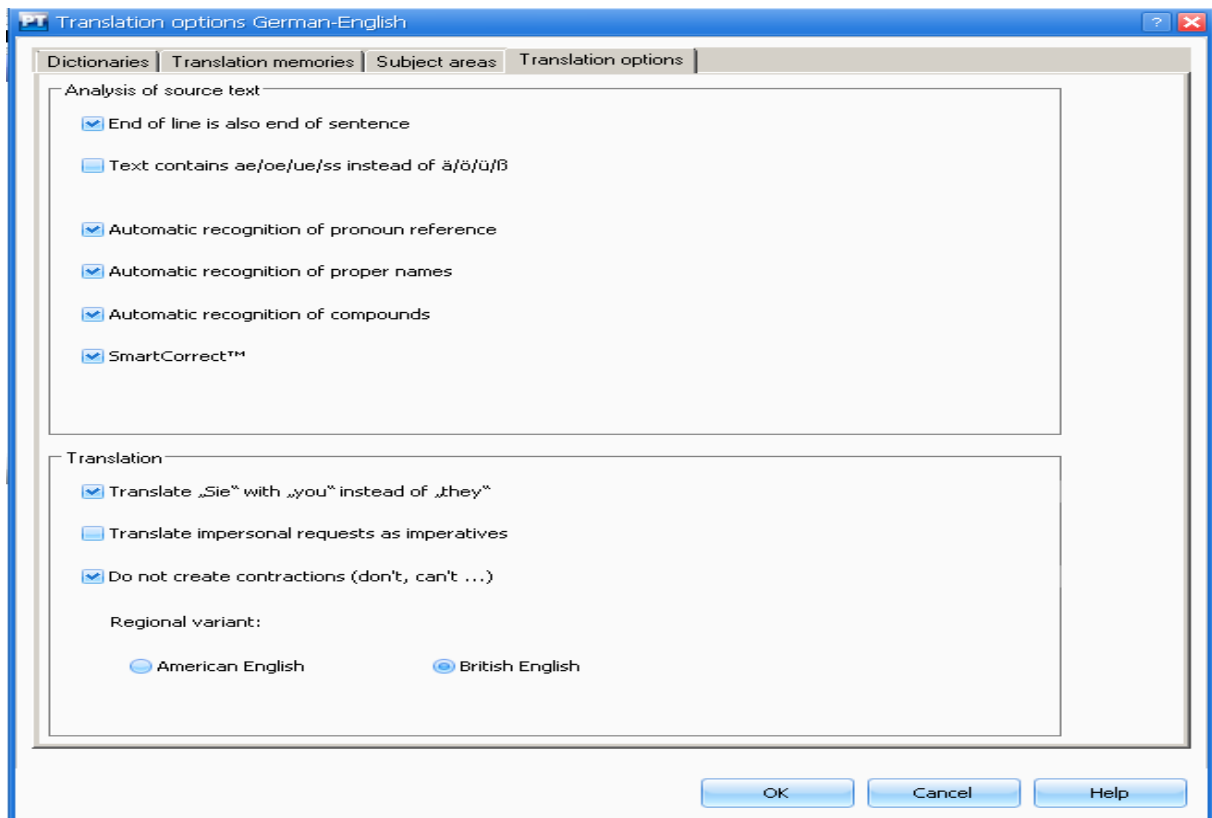
Right pane: translation options (e.g. use of TM (translation memory), dictionaries...) and statistics



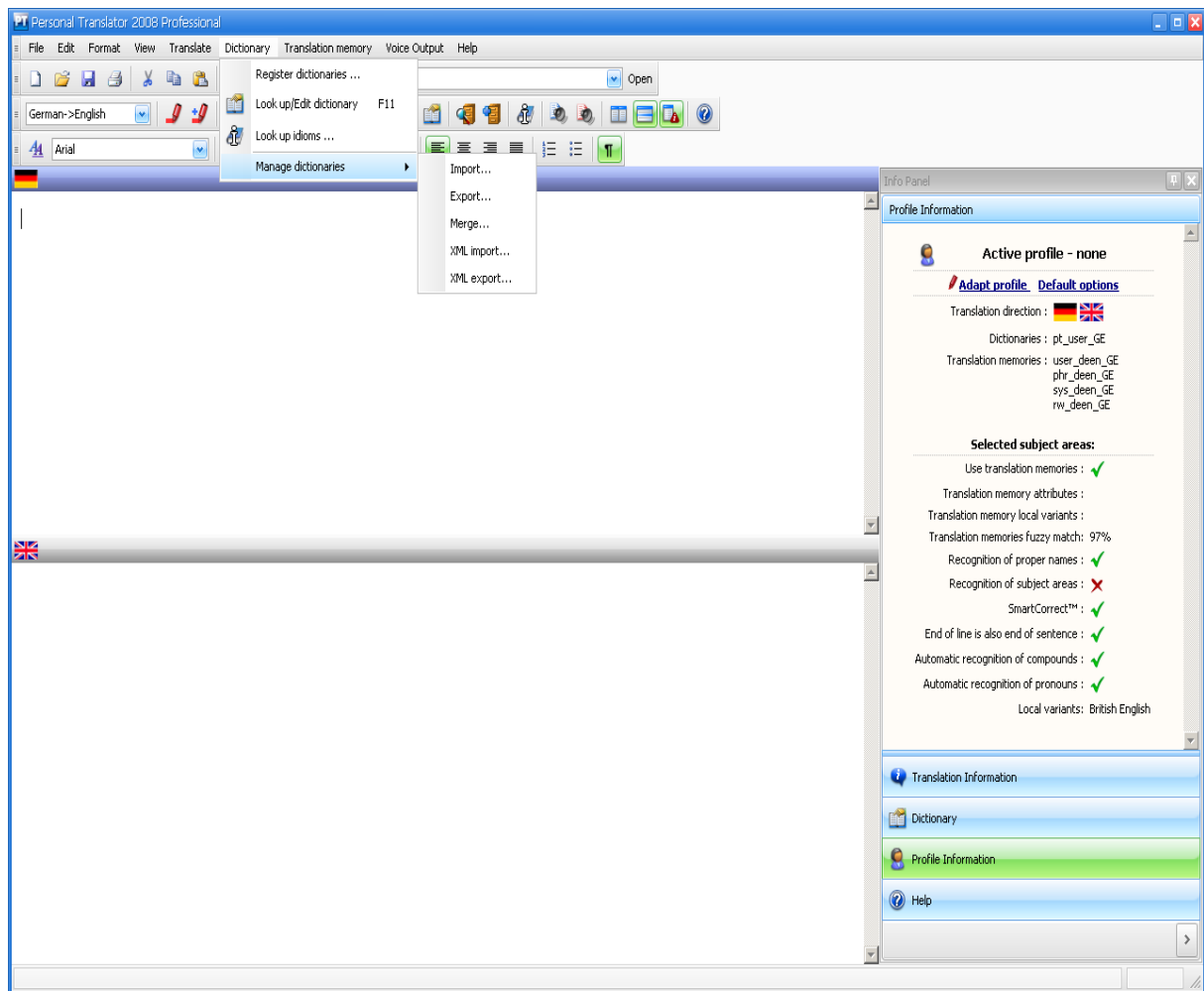
Translation memory options: fuzzy matching percentage and additional translation memories (exceptional feature, because most of the time, no more than one TM can be used)



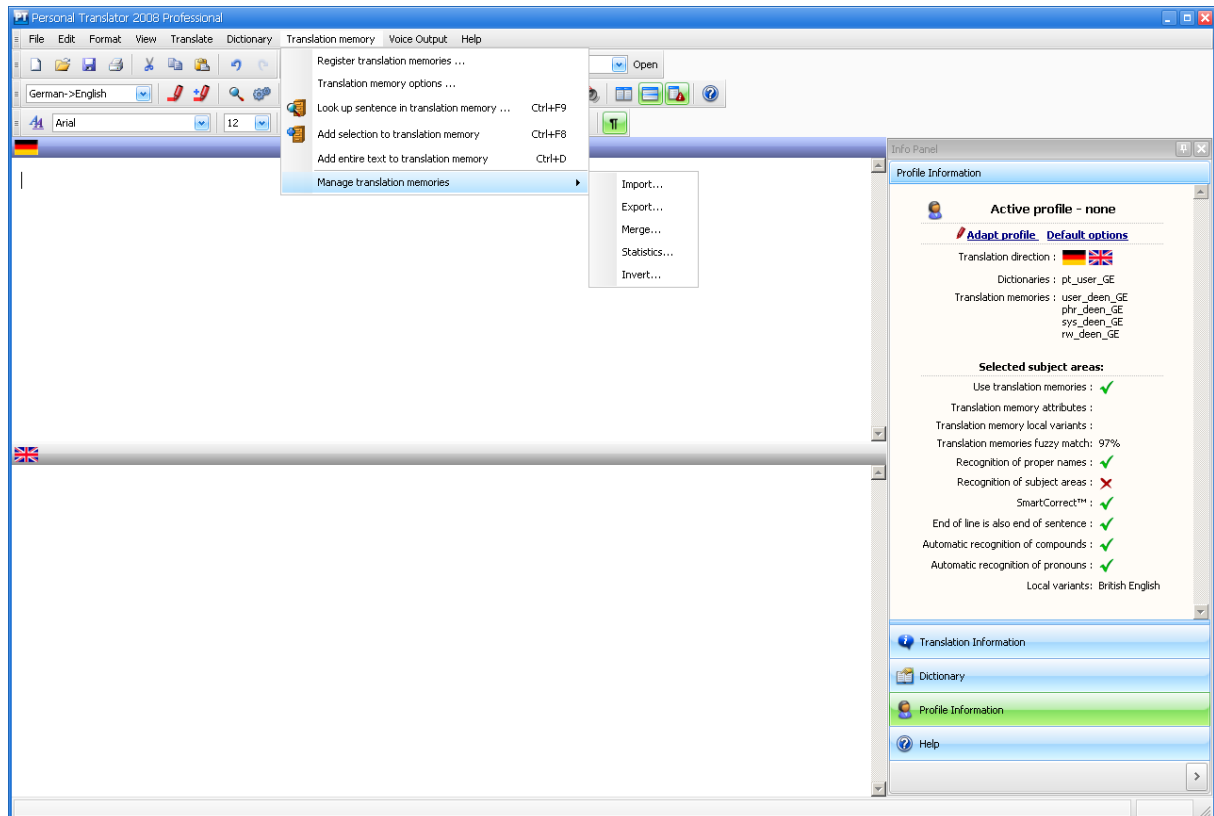
Domain or subject options: categories and subcategories



Translation options: some linguistic and stylistic options: pronouns, imperatives, formal/informal...



Dictionary management:



Translation memory management: import, export, merge, invert...in all common formats

c. Conclusion post-editing in desktop MT engines

We will now briefly discuss the most important post-editing features:

The two most popular desktop MT applications, Systran and Personal Translator (ex IBM), offer much more post-editing options than their online counterparts.

There are two reasons for this:

1. they are paid applications
2. an offline environment offers much greater versatility (web 2 being still in its infancy)

The summary below compares the most important features of Systran and PT.

d. Comparative survey

| | Systran Premium 6 | Linguec Personal Translator 2008 Professional |
|---|------------------------------------|---|
| Terminology extraction | yes | yes |
| Specialised dictionaries | yes | yes |
| User Dictionary | yes | yes |
| User Dictionary format | xmtf, csv, xml, dct, tmx, xls, txt | txt, xml |
| Dictionary manager (merge, export, import, reverse...) | yes | no reverse |
| Dictionary editor | yes | yes |
| Multiple Translation Memory support | yes (TM as dictionary) | yes |
| Supported TM format - input - output | tmx, txt tmx, txt | pt, tmx tmx, txt |
| Translation Memory match - 100% only (except for variables) - adaptation (conjugation, gender, inflection...) | yes no | yes no |
| Translation confidence indicator | yes | no |
| Integrated spellcheck | yes | yes |
| Automatic source language detection | yes | yes |
| Batch translation | yes | yes |
| Translation Project Manager | yes | yes |
| Side-by-side alignment | yes | yes |
| Supported source file formats | DOC, RTF, TXT, HTML, XHTML, PDF | DOC, RTF, TXT, HTML, PDF |
| Text to speech | no | yes |
| Speech to text | no | no |
| HTML editing | yes | no |

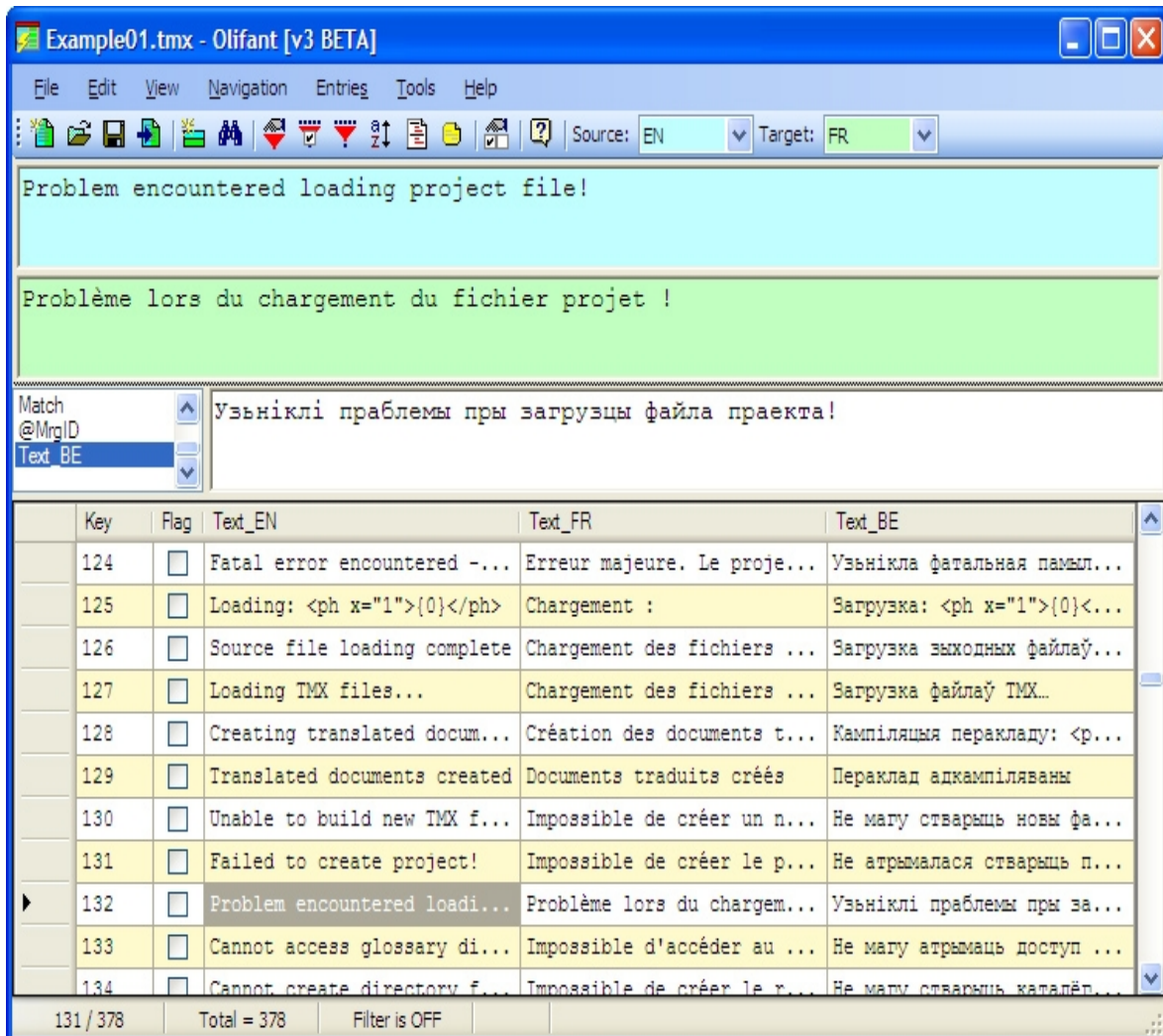
| | | |
|--|--|---|
| Translation review tool | yes | yes |
| Source text linguistic enrichment | yes | no |
| Alternative division of composite words | yes | no |
| Old or new spelling | yes | yes |
| Version control | yes | no |
| Do not translate list (words in capitals, addresses, proper names...) | yes | yes |
| Automatic localization (dates, numbers...) | yes | no |
| Language variants option (e.g. British, US...) | yes | yes |
| Register/style option (informal, speech...) | yes | yes (pronouns only) |
| Domain option (legal, medical...) | yes | yes |
| Imperative or infinitive option | yes | yes |
| Complex search/replace | yes | yes |
| Marking of <ul style="list-style-type: none"> - not translated - TM input - Alternatives - Spelling mistakes | <ul style="list-style-type: none"> yes yes yes yes | <ul style="list-style-type: none"> yes yes no yes |

4. Translation memory editor: Olifant – open source Okapi tools: <http://okapi.sourceforge.net/Release/Olifant/ReadMe.htm>

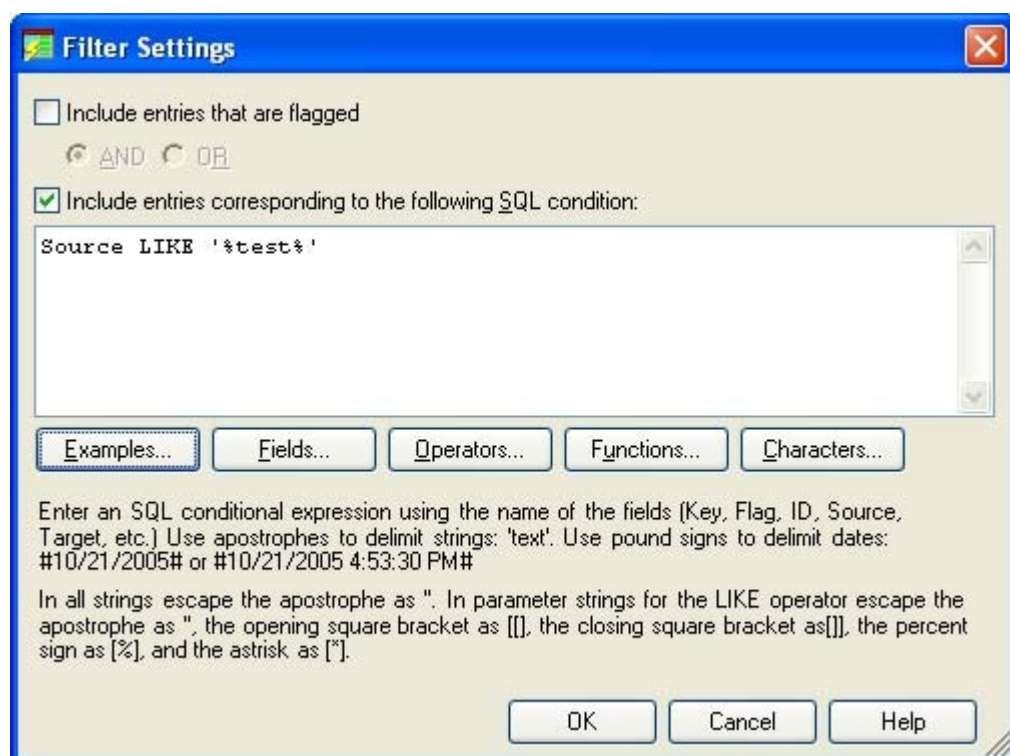
Olifant is a .NET application that allows you to load or import translation memories in different formats (such as TMX or tab-delimited). You can edit the translation units, their attributes and any other associated data. Olifant allows you to save or export your data in TMX or TXT formats.

The tools provide a table-driven interface to easily navigate through the translation memory. It offers functions such as flagging duplicates, search and replacing text using regular expressions, filtering the entries based on various criteria, merging other TMs into an existing one, exporting part of the TM, and many other utilities.

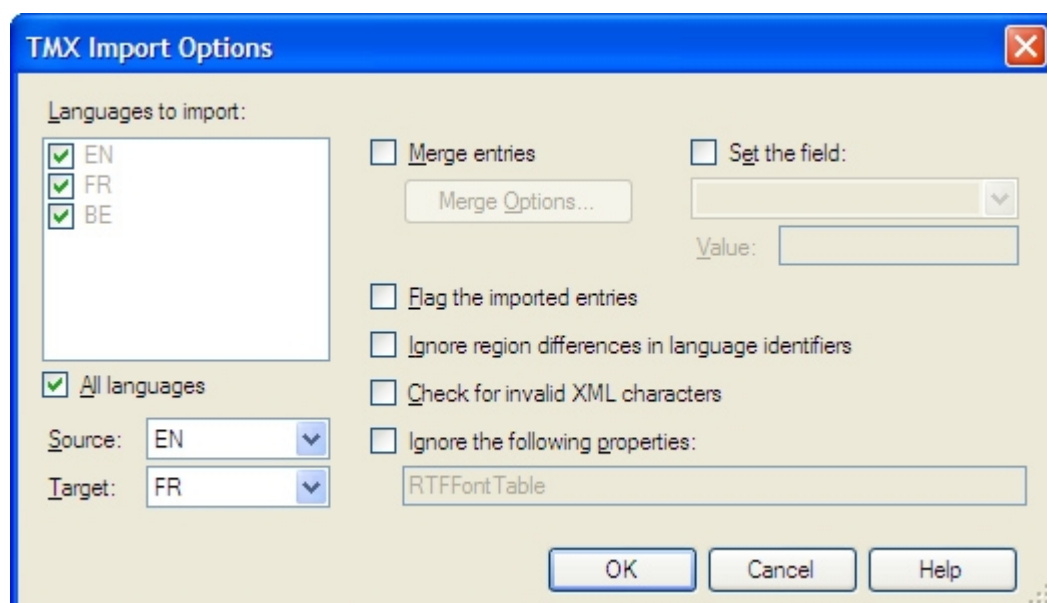
Screen 1: The main window with its edit pane and its list of entries.



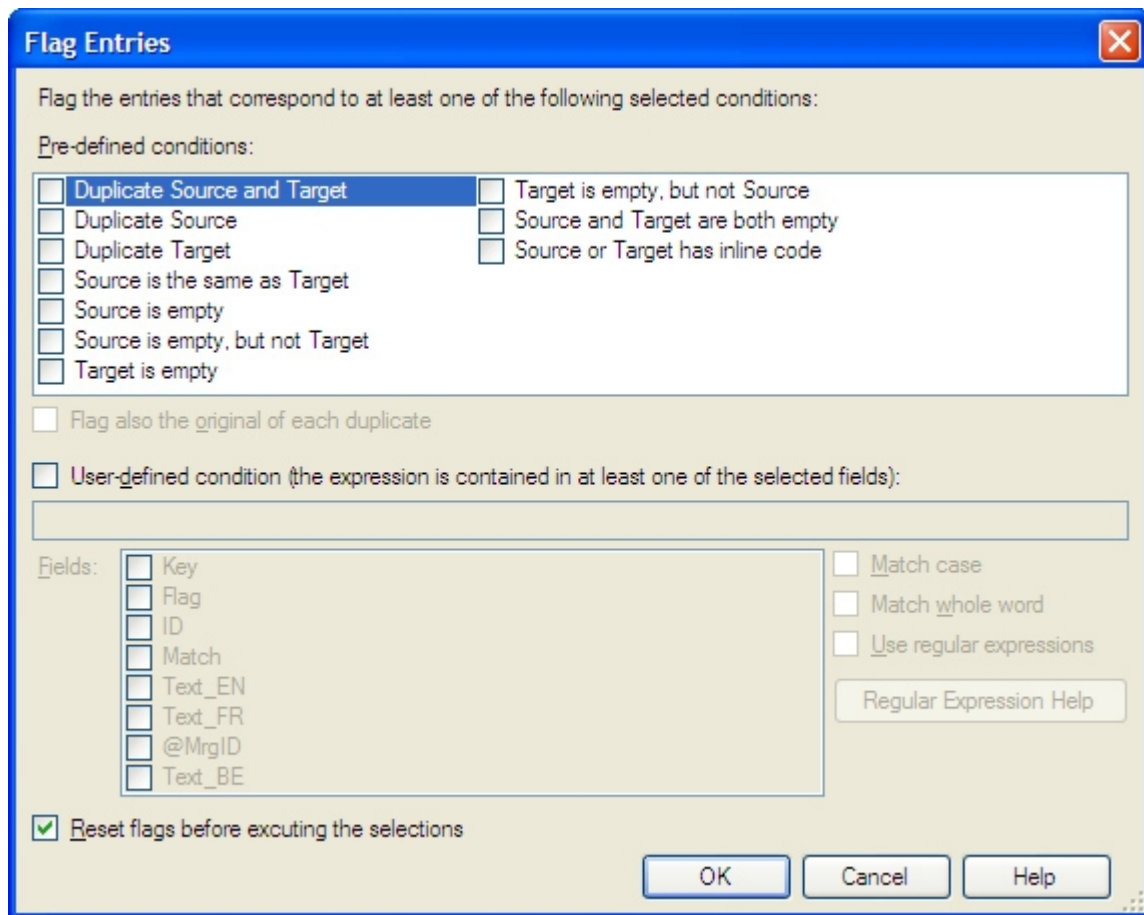
Screen 2: The [Filter Settings](#) dialog box where you can set criteria for the entries to display.



Screen 3: When importing a TMX document you can select which language within the document to import.



Screen 4: The **Flag Entries** dialog box where you can set the criteria of the entries to flag. For example, you can flag all the entries containing a given regular expression.



5. Conclusions for an outline of the PaCo interface

The above survey of both the online and desktop TM engines and the desktop TMX editor has led us to identify the following features as being important to the outline of the PaCo interface.

- module for input (copy/paste) of the source text: text length limitation?
Character set?
- module for synchronous viewing of the source and target texts (side-by-side alignment)
- choice of source (or automatic source language detection) and target language: 4 combinations: Dutch<>English and Dutch<>French
- TRANSLATE button
- language swap (reverse source and target language) button?
- module for input of TMX (translation memory) – TM takes precedence over MT
- module for input of dictionary (TBX, CSV...)
- module for input of DNT (Do Not Translate) list
- module for some options: register, specialty, marking of DNT/unknown words/TM input/variants and parser-assisted editing
- translation editing can be done within a text processing environment (search/replace, spellcheck, layout...)
- TMX editing can be done within Olifant