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: <u>http://translate.google.com/</u>

| Web Images Maps News Sho | opping <u>Gmail</u> more ▼ | | | | | |
|---------------------------|------------------------------------|--|----------------------|-------------------------------------|--|--|
| Google Mom | ne <u>Text and Web</u> <u>Tran</u> | nslated Search | Dictionary | <u>Tools</u> | | |
| Hanslate Sela | | Turu alata taut | | | | |
| | | Translate text | | | | |
| | | | opuge one. | | | * |
| | | | | | | - |
| | | Spanish 🔻 > E | nglish ▼ <u>swap</u> | | | Translate |
| | | | | | | |
| | | Search other | languages | | Languages avail | able for translation |
| | | Enter a search phi language to find in languages. <u>Translate and se</u> | formation in other | | Arabic Bulgarian Catalan Chinese Croatian Czech | Italian Japanese Korean Latvian Lithuanian Norwegian |
| | | Translation to | ols | | Danish Dutch | Polish Portuguese |
| | | Get translation wit make your webpag in other languages <u>Get translation to</u> | ge instantly availab | | English Filipino Finnish French German Greek Hebrew Hindi Indonesian | Romanian Russian Serbian Slovak Slovenian Spanish Swedish Ukrainian Vietnamese |
| | | | | <u>Home</u> - <u>A</u> ©2009 Goo | <u>bout Google Transla</u> ogle | <u>ite</u> |
| Web Images Maps N | ews Shopping Gmail | more V | | | | |
| Google | Home Text a | ind Web | Translated | Search | Diction | <u>ary Tools</u> |
| Translate 🔘 BETA | | | | | | |
| Translate text or w | vebpage | | | | | |
| Enter text or a webpage U | IRL. | | | | | |
| | | | | | | |
| Spanish • > English | ▼ <u>swap</u> | | | | | Translate |

| Number of language combinations | 34 * 34 |
|-------------------------------------|---|
| Automatic source language detection | yes |
| Language swap (reverse source and | yes |
| target language) | |
| 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: <u>http://www.systranet.com/</u>

| 🥞 SYS' | TRANet | Your free on | line trans | lator | | | | You are logged in as kdes |
|--------|------------------------|--------------|------------|----------|-------------|---------|----------|----------------------------------|
| Text W | /eb Page | Dictionary | RSS | File | My Dict | tionary | | |
| Dutch | • ÷ | English | • | Options: | | | 86 | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | Translate |
| | | | | | | | | |
| 🥱 sys | TRANet | Your free on | line trans | lator | | | | You are logged in as kde |
| Text W | /eb Page | Dictionary | RSS | File | My Dic | tionary | | |
| Dutch | • \$ | English | • | | | | | |
| Dutch | | DNT English | | Parl | t of Speecl | h | | |
| 2 | | | | Det | ect | • | . | |

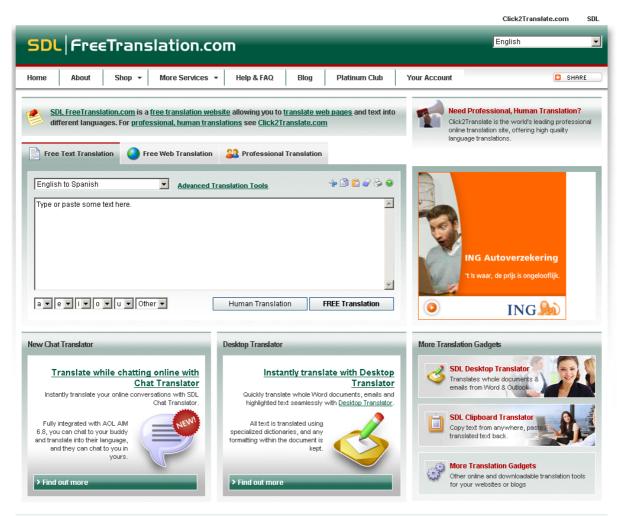
| Number of language combinations | 52 |
|-------------------------------------|--|
| Automatic source language detection | no |
| Language swap (reverse source and | no |
| target language) | |
| Text length limitation | 100k keystrokes |
| Dictionary search | yes |
| Text formats | doc, docx, rtf, html, htm, txt |
| URL translation | ves |
| 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: <u>http://www.windowslivetranslator.com/</u>

| Vindows Live~ | Start Hotmail Spaces OneCa | re MSN | Zoe | ken |
|--------------------------------------|---|--|---|---|
| Translator | | | | |
| Gratis online vertalingen waarm | ee het web echt <i>wereldwijd</i> wordt | | Toe | voegen aan Favorieten |
| Voer een tekstfragment in (to | ot 500 woorden) | Vertaalde tekst | | |
| | worden vertaald te typen of te plakken. | Automatisch vertalen helpt om de vreemde taal te begrijpen maar is vertaler te vervangen als een foutl | niet geschikt om een p | professionele |
| | <u>Alles wisser</u> | i . | | Vertaling door Microsoft [®] Translator |
| Of vertaal een webpagina 🛛 | nttp:// | Talen Engels - Nederlands | ¥ | Tekst vertalen |
| Nieuw: Assistent voor Engels als twe | ede taal | Introductie: | Vertaalbot voor Windows Voeg mtbot@hotmail.com | |
| Translator-knop voor Windows Live 1 | oolbar | - (a) | contactpersoon en nodig Messenger-gesprekken. | |
| Translator voor webpagina's toevoec | <u>en aan eigen website</u> | | məəsəriyər-yespi düldir. | |
| + Veelgestelde vragen over Translato | <u>r</u> | | | |

| Number of language combinations | 28 |
|-------------------------------------|--|
| Automatic source language detection | no |
| Language swap (reverse source and | no |
| target language) | |
| 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 <u>Translator home page</u> , or use the <u>Feedback</u> 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/



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 | no |
| target language) | |
| 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: <u>http://www.promt.com/</u>

| | | | | <u>Reqi</u> | <u>ster Sign In</u> | 🐱 English (<u>cha</u> |
|--|------------------|------------------------|-----------------------|----------------|------------------------|--|
| Ads by Google | Deutsch Englisch | Babel Fish Translation | <u>Translate Text</u> | <u>Transla</u> | ition Tools | |
| ext Translator Website | e Translator De | esktop Translators | | 📮 <u>m.</u> | online-transl | ator.com for mot |
| imple mode Advanced mode | | | | | | |
| Show variants of translation Transliterate unknown words Translation | ABC Spelling | Result | Keyboard Help | | | |
| ource text | 🛍 🗎 🗡 | Translated text | | a × | Transl | ation Software |
| ype or paste your text here | <u>×</u> | | | × | © Downla Or @pro | translators and translator uline Store mt Expert 8.0 RANSLATOR \$ 535 tuy now> |
| English to Russian translation | Translate | | | | | |

| | 1 |
|-------------------------------------|--|
| Number of language combinations | 24 |
| Automatic source language detection | no |
| Language swap (reverse source and | no |
| target language) | |
| Text length limitation | 3,000 keystrokes (10,000 for registered |
| | users) |
| Dictionary search | Yes |
| Text formats | txt (copy/paste only) |
| URL translation | yes |
| Interaction | The "Translation options" group includes |
| | the following settings: |
| | "Show variants of translation" (if a word |
| | has several meanings, alternative |
| | translations are displayed in brackets) |
| | "Transliterate unknown words" (words |
| | not found in the dictionaries are |
| | transliterated from Cyrillic into Latin script |
| | or vice versa). |
| | 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. |
| | |

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 nonexisting. 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

| SYSTRAN Translation Project Manager | | | | | | | |
|---|---|---|--|---------------|--|--|--|
| Eile Edit View Corpus Translate Wind | ow <u>H</u> elp | | | | | | |
| : D 📂 🖬 📣 Q 💕 , : Q 🛤 🧐 📂 , : 🤊 ∾ I 🛦 🖦 🎕 🗙 I 🖓 🐝 🍬 I 🚔 , : 🔟 🖂 🦻 🖃 🗇 🚍 . | | | | | | | |
| Dutch to English 🤤 Translate 🗸 Default 🔹 📰 | | | | | | | |
| Corpus Explorer 🛛 📮 🗙 | | | | × | | | |
| 🖃 🞧 test paco.stp | Tekstvak: 62 | <u> </u> |)ext profession: 62 | _ | | | |
| delzen besluiten profs_ned | 43.05 Verslag Prof. dr. ir J. DEFRANCO : - brief dd. 25.06.2 | 2007 | 43.05 report Prof. dr. ir J. DEFRANCQ: - letter dd | . 25.06.2007 | | | |
| | - verslag dd. 12.06.2007 | | - report dd. 12.06.2007 | | | | |
| | Het verslag van de Cel Corrosie wordt mij toegezonden. Tevens tref ik zending de factuur. | : bij de | The report of cell corrosion is transmitted me. Also I find the invoice at the consignment. | | | | |
| | Er wordt gevraagd om de factuur binnen de dertig dagen te voldoen. | | There the invoice it is asked within the thirty days satisfy. | | | | |
| | Het verslag, opgesteld door de Universiteit Gent, wordt aan het versla gehecht. | The report, established by the Universiteit Gent, is been attached to the report. | | | | | |
| | lk stel voor het ereloon ten bedrage van 1.986,66 euro te betalen binr vervaldag. | I put for honorary remunerations to the amount of 1,986, 66 euro to pay within the decline day. | | | | | |
| | De betaling zal. tenzij hiertegen bezwaren worden gemaakt, gebeuren van de rekening prefinanciering. | | The payment, unless against this are objected, happenin prefinancement. | g the account | | | |
| | | - | | | | | |
| | Sentence Review | | | ч х | | | |
| 🐎 🗸 New Dictionary 💌 🧙 🖶 | | | | | | | |
| | 💌 🚊 🥎 Source text | Translation | n # C | A _ | | | |
| | Tekstvak: | text profess | sion: 1 1 | 0 🦓 🛝 🗍 | | | |
| | · 62 | 62 | 1 1 | 0 🏘 🔥 👻 | | | |
| | Not Found Word Review Alternative Meaning Review Senten | ce Review | | | | | |
| | | | | | | | |

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



Systran bar in ms word

| - | SYS | STRAN Translation Project Manager | | 10 | | × |
|---|--------------|--|------------------------|---|--|------------|
| : | <u>F</u> ile | <u>E</u> dit <u>V</u> iew <u>C</u> orpus <u>T</u> ranslate <u>W</u> indo | w <u>H</u> elp | | | |
| 1 | 3 | Open File | | KI 🙈 🐝 🐮 🖴 📑 🥅 🚍 🦻 🗔 🗉 | | |
| 1 | 8 | Open URL | | | | |
| | | New Project | CTRL+N | 1 | | × |
| i | 6 | Open Project | CTRL+O | - | ext profession: 62 | |
| | | Save Project | CTRL+S | J. DEFRANCQ : - brief dd. 25.06.2007 | 43.05 report_Prof. dr. ir J. DEFRANCQ: - letter dd. 25.06.2007 | |
| | | Save Project <u>A</u> s | | | | |
| | | Close Project | | | - report dd. 12.06.2007 | |
| | | Project Properties | | ie wordt mij toegezonden. Tevens tref ik bij de | The report of cell corrosion is transmitted me. Also I find the invoice at the consignment. | |
| | | Close | CTRL+W | | - | |
| | Q | Preview | • | uur binnen de dertig dagen te voldoen. | There the invoice it is asked within the thirty days satisfy. | |
| | 6 | Export | • | Export Translated Document CTRL+ | E report, established by the Universiteit Gent, is been attached to the repo | ıt. |
| | ۲ | Print Document | CTRL+P | Export Source Document | | |
| | o, | Print preview | | Export Translated Selected Documents | for honorary remunerations to the amount of 1,986, 66 euro to pay within ecline day. | |
| | | Print Options | | Export Source of Selected Documents | | |
| | | Page setup | | n bezwaren worden gemaakt, gebeuren van de | The payment, unless against this are objected, happening the account prefinancement. | |
| | | C:\Users\koen\Documents\my documents | vertalin\test paco.stp | | | - |
| | | Exit | | | | φ × |
| | _ | | 🔊 🖬 New Dictionary 🔹 | 6 | | |
| | | | Nource text | Translation | n # C A | • |
| | | | Tekstvak: | text profess | sion: 1 1 0 🆓 🛝 | |
| | | | % 62 | 62 | 1 1 0 🏘 🛝 | - |
| | | | Not Found Word Review | Alternative Meaning Review Sentence Review | | |
| | | | | | | at |

Export feature for source and target texts.

| SYSTRAN Translation Project Manager | | | | | | | 015 | 5 | | | | × | |
|-------------------------------------|--------------|-------------------------------|----------------------------|-------|---|---|--------------|--|-------------|----------|------------|------------|---|
| Eile | <u>E</u> dit | t <u>V</u> iew <u>C</u> orpus | <u>T</u> ranslate <u>V</u> | Vindo | ow <u>H</u> elp | | | | | | | | |
| 10 | 2 | <u>U</u> ndo | CTRL+Z | 2 | 📮 🤊 (° 🌡 🗈 🛍 🗙 🐴 👫 | 🤈 💌 👗 🖦 🛍 🗙 🔲 🖀 🥛 🛄 🚍 🦻 🖃 📰 📰 📰 📰 📰 📰 📰 | | | | | | | |
| Duto | ~ | <u>R</u> edo | CTRL+Y | ilt | - == _ | • E | | | | | | | |
| Corp | * | Cut | CTRL+X | × | delzen besluiten profs_ned | | | | | | | | × |
| ⊡• | 6 | <u>C</u> opy | CTRL+C | E | Tekstvak: 62 | | • | ext profession: 62 | | | | | - |
| 1 | | Paste | CTRL+V | | 43.05 Verslag Prof. dr. ir J. DEFRAN | CQ : - brief dd. 25.06.20 | 07 | 43.05 report Prof. dr. ir J. DEFRANCQ | - letter d | d. 25. | 06.200 | 7 | |
| | × | <u>D</u> elete | SUPPR | | - verslag dd. 12.06.2007 | | | - report dd. 12.06.2007 | | | | | |
| | М | <u>F</u> ind | CTRL+F | | _ | | | | | | | | |
| | ŜВ | Replace | CTRL+H | | Het verslag van de Cel Corrosie wordt mij te zending de factuur. | oegezonden. Tevens tref ik b | ij de | The report of cell corrosion is transmitted me. consignment. | Also I find | the inv | roice at t | he | |
| | <u>n,</u> | Find <u>N</u> ext | F3 | | Er wordt gevraagd om de factuur binnen de | | | The set when the set of the section of the state of the state of the section of t | | | | | |
| | ۲ | Lookup | CTRL+L | | Er wordt gevraagd om de factuur binnen de | e dertig dagen te voldoen. | | There the invoice it is asked within the thirty of | lays satis | у. | | | |
| | 5 | Dictionary Mana | ger 🕨 | | Open SDM | nt, wordt aan het verslag | | The report, established by the Universiteit Ge | nt, is bee | n attach | ed to the | e report | |
| | 8 | Lock Sentence T | ranslation | | Create UD Entry CTRL+D | | | | | | | | |
| | | Do Not Translate | ► F | | Create DNT Entry | 3,66 euro te betalen binne | n de | I put for honorary remunerations to the amount the decline day. | it of 1,980 | 5, 66 eu | ro to pay | within | |
| | | Alternative Mean | ings 🕨 | | Create ND Entry | den gemaakt, gebeuren v | ran da | The payment, unless against this are objected, happening the acco | | | | | |
| | | Source Ambiguit | ies 🕨 | | Create TM Entry CTRL+M | uen geniaakt, gebeuren v | anue | prefinancement. | a, napper | ing the | account | | |
| | | Translation Mem | ories 🕨 | | Show UD Entry CTRL+MAJ+D | | • | | 0 OF 1 | nuce | | | - |
| | | | | | Show ND Entry | | | | | | | ą | × |
| | | | | | Show TM Entry CTRL+MAJ+M | | | | | | | | |
| | | | | | Find in Dictionaries | | Translation | 1 | # (| : A | | | |
| | | | | | Tekstvak: | | text profess | ion: | 1 1 | 0 | 44 | n, | |
| | | | | | 62 | | 62 | | 1 1 | 0 | 4 | <u>n</u> , | - |
| | | | | | Not Found Word Review Alternative M | Meaning Review Sentence | Review | | | | | | |
| | | | | | | | | | | | | | |

Dictionary manager:

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

Also DNT and TM are regarded as dictionaries.

| 🐑 SYSTRAN Translation Project Manager | | | | | | | |
|--|---|----------------------------|---|------------|---|--|------|
| Ele Edit View Corpus Iranslate Window Help | | | | | | | |
| i 🗋 📂 🖬 | 🗋 📂 🔐 🚺 Vertical Layout 🧳 🐑 🐇 🐁 🛸 🛍 🗙 🖓 🍪 🖕 🗠 🚆 🛄 🚍 🦻 | | | | | | |
| Dutch to Er 🚍 Horizontal Layout | | | | | | | |
| Corpus Exp | B | Markup Styles | delzen besluiten profs ned | | | | × |
| : ⊡∵@nj test pa | | Corpus Explorer | kstvak: 62 | | ٠ | ext profession: 62 | |
| de 🔛 | | Processing Status | .05 Verslag Prof. dr. ir J. DEFRANCQ : - brief dd. 25.06.20 | 007 | | 43.05 report Prof. dr. ir J. DEFRANCQ: - letter dd. 25.06.2007 | |
| | | Context Action | | | | | |
| | | Not Found Word Review | -rerslag dd. 12.06.2007 | | | - report dd. 12.06.2007 | |
| | | Extracted Term Review | t verslag van de Cel Corrosie wordt mij toegezonden. Tevens tref ik nding de factuur. | bij de | | The report of cell corrosion is transmitted me. Also I find the invoice at the consignment. | |
| | | Source Ambiguity Review | | | | - | |
| Alternative Meaning Review Sentence Review Translation Memory Review | | Alternative Meaning Review | wordt gevraagd om de factuur binnen de dertig dagen te voldoen. | | There the invoice it is asked within the thirty days satisfy. | | |
| | | Sentence Review | t verslag, opgesteld door de Universiteit Gent, wordt aan het verslag | | | The report, established by the Universiteit Gent, is been attached to the rep | |
| | | Translation Memory Review | hecht. | | | | |
| | | Statistics | stel voor het ereloon ten bedrage van 1.986,66 euro te betalen binn valdag. | ien de | | I put for honorary remunerations to the amount of 1,986, 66 euro to pay with the decline day. | in |
| | | | De betaling zal, tenzij hiertegen bezwaren worden gemaakt, gebeuren ekening prefinanciering. | van de | | The payment, unless against this are objected, happening the account prefinancement. | |
| | | | | | • | | - |
| Sentence Review | | | | | | чх | |
| 🔭 🗸 New Dictionary 🔽 🌚 🐨 | | | | | | | |
| | | | 👏 🚊 🥎 Source text | Translat | ion | # C A | |
| | | | Tekstvak: | text profe | essio | n: 1 1 0 🆓 🖏 | |
| | | | 62 | 62 | _ | 1 1 0 🏄 🗛 | - |
| Not Found Word Review Alternative Meaning Review Sentence Review | | | | | | | |
| | | | | | _ | | . di |

Marking of:

- Words not found
 Ambiguous source segments
 Alternatives meanings
 TM segments

| Profile: Default Resources Domains and Dictionaries Translation Choice Files Filter options Formating Do Not Translate options Source analysis Country localization Rendering Imperative Pronouns Style parameters Indicators UDD: Vertical Science Earth Sciences Colloquial Computers/Data Processing | Application Settings Translation Options | License | (|
|---|---|--|---|
| □ Domains and Dictionaries □ Translation Choice Files □ Filter options □ Formatting □ Do Not Translate options □ Douge analysis □ Source analysis □ Country localization □ Rendering □ Imperative □ Pronouns □ Style parameters □ Indicators | Profile: Default - | New Rename Delete Import Export | |
| Filter options Formatting Do Not Translate options Linguistic options Source analysis Country localization Rendering Imperative Pronouns Style parameters Indicators General Automotive Aviation/Space Military Science Military Science Naval/Maritime Metallurgy Economics/Business Life Sciences Style parameters Medicine Food Science Colloquial Computers/Data Processing | Domains and Dictionaries | | |
| | Formatting Do Not Translate options Do Not Translate options Source analysis Country localization Rendering Imperative Pronouns Style parameters Indicators | Automotive Aviation/Space Aviation/Space Military Science Military Science Metallurgy Economics/Business Political Science Difference Earth Sciences For Earth Sciences For Medicine Food Science Colloquial Computers/Data Processing | |

Specialized dictionaries can be combined.

| SYSTRAN Global Options | _ 🗆 × |
|---|---|
| Application Settings Translation Options Lice | ense 📀 |
| Profile: Default | ew Rename Delete Import Export |
| Resources Domains and Dictionaries Translation Choice Files Filter options Formatting Do Not Translate options Source analysis Country localization Rendering Imperative Pronouns Style parameters Indicators | Formatting Preserve textual formatting None Line Feed definition Auto detection Segmentation character list |
| | OK Cancel Apply |

Textual formatting options and segmentation character list can be edited

| Application Settings Translation Options Li | cense | (|
|---|---|-----|
| Profile: Default Resources Domains and Dictionaries Translation Choice Files Filter options Formatting Do Not Translate options Linguistic options Country localization Rendering | New Rename Delete Import Export Do Not Translate options Foreign sentence detection by Office language markup Foreign sentence detection based on NFW count No detection of foreign language sentence from NFW word Do Not Translate font list Symbol,Wingdings,Webdings Do Not Translate style list | • |
| Imperative Pronouns Style parameters Indicators | Do Not Translate addresses Do Not Translate capitalized words | |
| | OK Cancel A | pph |

- Do Not Translate options: foreign words (automatic detection option) certain fonts certain styles

 - addresses
 - capitalized words

| STRAN Global Options | |
|--|---------------------------------|
| Application Settings Translation Options | License |
| Profile: Default | New Rename Delete Import Export |
| - Resources | |
| Domains and Dictionaries | Source analysis |
| Translation Choice Files | Translation Memory match |
| Filter options | Normal |
| Formatting | |
| Do Not Translate options | Source spelicheck |
| - Linguistic options | |
| Source analysis | Acronym detection |
| Country localization | Global entity recognition |
| Rendering Imperative | |
| Pronouns | Enable Domain detection |
| Style parameters | |
| Indicators | |
| | |
| | |
| | |
| | |
| | |
| | OK Cancel Apply |

Linguistic options - Source analysis: - translation memory matching - spellcheck - acronym detection - domain detection

| Application Settings Translation Options | License |
|--|---------------------------------|
| Profile: Default 💌 | New Rename Delete Import Export |
| | Country logglization |
| Domains and Dictionaries | Country localization |
| Translation Choice Files | Chinese |
| | Simplified Chinese - |
| Formatting | English |
| Do Not Translate options | |
| | American English |
| Source analysis | Portuguese |
| Country localization | |
| ···· Rendering | Brazilian Portuguese |
| Imperative | German new spelling |
| Pronouns | Comman new spenning |
| Style parameters | |
| Indicators | |
| | |
| | |
| • | |
| | |
| | OK Cancel App |

Linguistic options – Country localization: - regional variants - sociolinguistic variants - new or old spelling

| SYSTRAN Global Options | | - - × |
|---|--|--------------|
| Application Settings Translation Options | License | 0 |
| Profile: Default | New Rename Delete Import Export | |
| Resources Domains and Dictionaries Translation Choice Files Filter options Formatting Do Not Translate options Linguistic options Source analysis Country localization Rendering Imperative Pronouns Style parameters Indicators | Rendering Localize punctuation Convert Hijri dates Not Found Word transliteration | |
| • | | |
| | OK Cancel | Apply |

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

| SYSTRAN Global Options | _ □ × |
|--|--|
| Application Settings Translation Options | License |
| Profile: Default - | New Rename Delete Import Export |
| Resources | |
| Domains and Dictionaries | Imperative |
| Translation Choice Files | Imperative choice - French/German/Italian/Spanish/Portuguese |
| | Infinitive 🔻 |
| Formatting | |
| Do Not Translate options | Imperative choice - Chinese/Japanese/Korean |
| Enguistic options | Informal |
| Source analysis | |
| Country localization | |
| Rendering Imperative | |
| Pronouns | |
| | |
| Indicators | |
| maloutoro | |
| | |
| | |
| | |
| | |
| | OK Cancel Apply |
| | |

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

| SYSTRAN Global Options Application Settings Translation Options Lice | ense 🥥 |
|--|--|
| Profile: Default N •• Resources •• Domains and Dictionaries •• Translation Choice Files •• Filter options •• Formatting •• Do Not Translate options •• Source analysis •• Country localization •• Rendering •• Imperative •• Pronouns •• Style parameters •• Indicators •• Indicators •• Indicators | ew Rename Delete Import Export Pronouns 1st person 1st person singular gender Masculine Ist person plural gender Masculine 2nd person 2nd person gender Masculine V 2nd person polite/informal Polite V |
| | OK Cancel Apply |

- Linguistic options Pronouns gender register (informal polite...)

| SYSTRAN Global Options Application Settings Translation Options Licen | _ 	 × |
|--|---|
| Profile: Default New Resources Domains and Dictionaries Translation Choice Files Filter options Formatting Do Not Translate options Source analysis Country localization Rendering Imperative Pronouns Style parameters Indicators | Rename Delete Import Export Style parameters Separate translation of sentence between quotes Document type Default Import Default Abstract Import List Journalism User Guide Correspondence Import Export |
| | OK Cancel Apply |

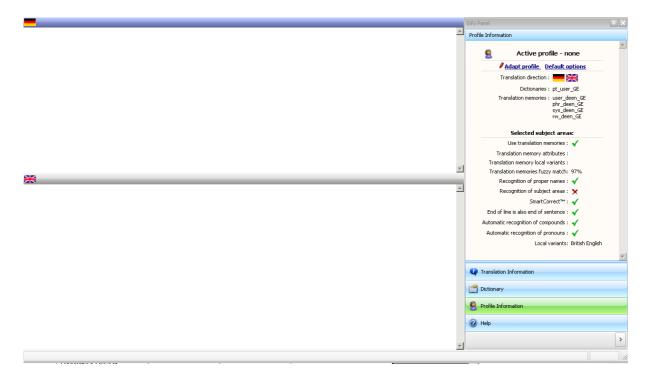
Style parameters and option to separate translation of sentence between quotes

| SYSTRAN Global Options | _ 🗆 × |
|---|---|
| Application Settings Translation Options License | • |
| Profile: Default New | Rename Delete Import Export |
| Resources Translation Choice Files Filter options Formatting Do Not Translate options Linguistic options Source analysis Country localization Rendering Imperative Pronouns Style parameters Indicators | Indicators Display alternative meanings None Not Found Word marker None User Dictionary marker None Translation Memory marker None Textual marker:=== XML <tm> Highlight red</tm> |
| | OK Cancel Apply |
| | |

Markers for alternatives – dictionary entries - TM b. Linguatec Personal Translator 2008 Professional

| Personal Translator 2008 Professional | _ 🗆 🗙 |
|---|---------------------------|
| File Edit Format View Translate Dictionary Translation memory Voice Output Help | |
| = 🗋 🧀 😹 🔏 🐘 🧌 🔿 🔿 👘 Address | |
| i German->English 💽 🍠 🗐 🧇 🧐 🞝 🔽 🔁 🗣 🏟 🗂 🥥 🧐 🦓 🖄 🕲 🖿 🥅 🙆 🎯 | |
| = <u>44</u> Arial | |
| _ | Info Panel 👎 🗙 |
| | Dictionary |
| | |
| | all dictionaries |
| | German->English 💽 OK 🔯 |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| 2 | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | * |
| | |
| | 😳 Translation Information |
| | Dictionary |
| | S O > |
| Y | |

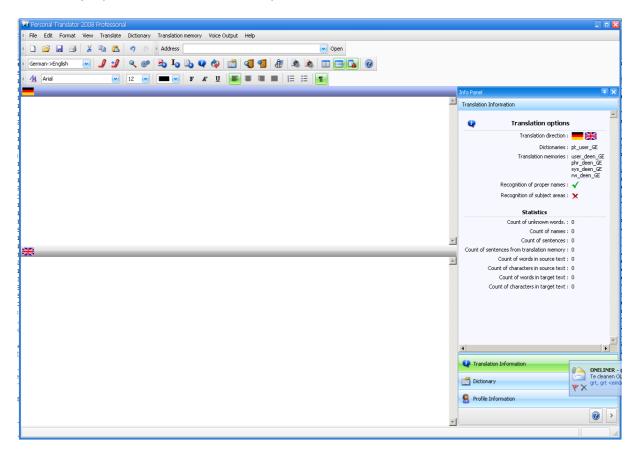
General survey of the user interface, with in the right pane all relevant metainformation: dictionary



Right pane: profile options

| :ti | ionaries selected for translation | | | | |
|-----|------------------------------------|---------|-----------|------|------------|
| | System dictionaries | Туре | Activated | | |
| > | MainGE | SYSTEM | ~ | | |
| | AddMainGE | SYS_ADD | | | |
| | Lookup_Allgemein | SYS_ADD | × | | |
| | Lookup_IT_Telekommunikation | SYS_ADD | ~ | | |
| | Lookup_Technik | SYS_ADD | ~ | | |
| | Lookup_Wirtschaft | SYS_ADD | ~ | ~ | |
| _ | User dictionary | Туре | Activat | | |
| | | | | e II | |
| > | pt_user | USER | | | |
| > | - | USER | | | |
| > | pt_user Additional dictionaries | | | | |
| > | - | USER | | | Up |
| > | - | USER | | | Up Down |
| > | - | USER | | | |
| > | - | USER | | | |
| > | - | USER | | | Down |
| > | - | USER | | | |
| > | - | USER | | | Down |

Dictionary options: combination of specialized dictionaries



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

| | naries Translation memories Subject areas Trans | lation options | | | | |
|---|---|------------------------------|----------|------------|--|--|
| ~ | Use translation memory | | | | | |
| | | | | | | |
| | anslate only sentences with fuzzy match larger than: | 97 📀 | | | | |
| Translate only sentences with attributes/regional variants: | | | | | | |
| | de_DE | | | | | |
| | en_GB | | | | | |
| r | en_US | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | T | | | | |
| | User translation memory user_deen | Туре | Activate | | | |
| / | | | | | | |
| | | USER | | | | |
| | | USER | | | | |
| | | USER | ~ | | | |
| | | USER | | | | |
| | Additional translation memories | Type | Activate | | | |
| > | Additional translation memories | | | Up | | |
| > | Additional translation memories | Туре | Activate | Up | | |
| > | Additional translation memories | Type USER_ADD | Activate | Up Down | | |
| > | Additional translation memories phr_deen sys_deen | Type USER_ADD USER_ADD | Activate | | | |
| > | Additional translation memories phr_deen sys_deen | Type USER_ADD USER_ADD | Activate | | | |
| > | Additional translation memories phr_deen sys_deen | Type USER_ADD USER_ADD | Activate | | | |
| > | Additional translation memories phr_deen sys_deen | Type USER_ADD USER_ADD | Activate | | | |
| > | Additional translation memories phr_deen sys_deen | Type USER_ADD USER_ADD | Activate | | | |

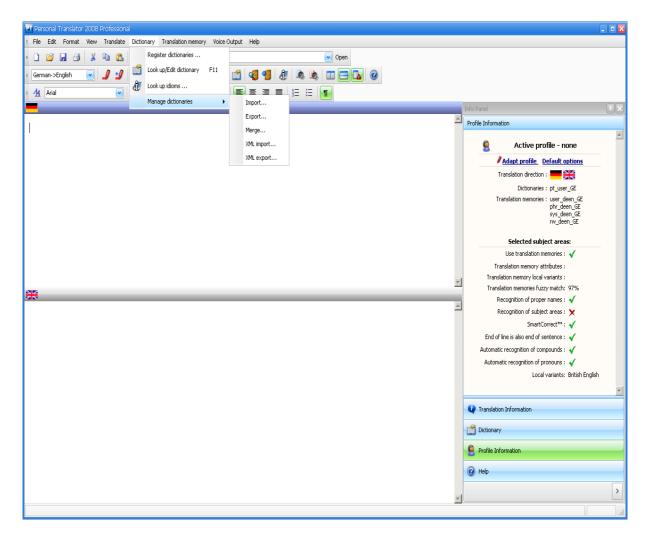
Translation memory options: fuzzy matching percentage and additional translation memories (exceptional feature, because most of the time, no more than one TM can be used)

| Tra | | | | | | | |
|--------|------------|--------------------------|-------------------|---------------------|----|--------|-------|
| Dictio | onaries | Translation memories | Subject areas | Translation options | 1 | | |
| | Set subi | ect areas automatically | 7 | | | | |
| | 500 300) | cee al cas accomatically | | | | | |
| > | ₽ 📒 | Art and Cultur | e . | | | | |
| | ⊕ | Economy | | | | | |
| | 🗄 · · 📒 | Humanities an | d Social Science: | s | | | |
| | ⊕ <u></u> | Leisure and Sp | | | | | |
| | ÷. 📒 | Medicine and B | | | | | |
| | <u>ا</u> | | es and Technolo | ogy | | | |
| | ÷ 🦲 | Society and So | ocial Issues | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | Reset | |
| | | | | | | | |
| ⊏Se | ected s | ubject areas: | | | | | |
| | | | | | | | |
| Γ | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| L | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | ок | Cancel | Help |
| | | | | | | | Ticip |

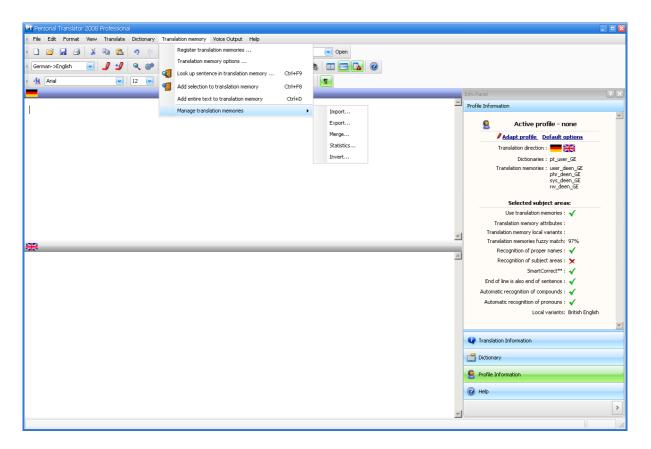
Domain or subject options: categories and subcategories

| Translation options German-English | 2 🔀 |
|---|-----|
| Dictionaries Translation memories Subject areas Translation options | |
| Analysis of source text | |
| End of line is also end of sentence | |
| 🦲 Text contains ae/oe/ue/ss instead of ä/ö/ü/β | |
| Automatic recognition of pronoun reference | |
| ✓ Automatic recognition of proper names | |
| Automatic recognition of compounds | |
| SmartCorrect™ | |
| | |
| □ Translation | |
| ✓ Translate "Sie" with "you" instead of "they" | |
| Translate impersonal requests as imperatives | |
| ☑ Do not create contractions (don't, can't) | |
| Regional variant: | |
| American English | |
| | |
| | |
| OK Cancel Help | |

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 | Linguatec 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 - not translated - TM input - Alternatives - Spelling mistakes | yes yes yes yes | 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 of 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.

🗾 Example01.tmx - Olifant [v3 BETA] File Edit View Navigation Entries Tools Help 🎁 🚅 🔚 🎦 🎽 🦓 🅰 📅 🍸 🏆 😫 🕒 🔗 😰 Source: EN ✓ Target: FR ¥ Problem encountered loading project file! Problème lors du chargement du fichier projet ! Match 🔨 Узьніклі праблемы пры загрузцы файла праекта! @MrgID Text BE Flag | Text_EN Text FR Text BE Key 🔲 Fatal error encountered -... Erreur majeure. Le proje... Узьнікла фатальная памыл... 124 Loading: <ph x="1">{0}</ph> Chargement : 125 Загрузка: <ph x="1">{0}<... 126 🔲 Source file loading complete Chargement des fichiers ... Загрузка зыходных файлаў... Loading TMX files... Chargement des fichiers ... Загрузка файлаў ТМХ... 127 128 Creating translated docum... Création des documents t... Кампіляцыя перакладу: <p... 129 Translated documents created Documents traduits créés Пераклад адкампіляваны Unable to build new TMX f... Impossible de créer un n... He магу стварыць новы фа... 130 Failed to create project! Impossible de créer le p... Не атрымалася стварыць п... 131 Problem encountered loadi... Problème lors du chargem... Узьніклі праблемы пры за... 132 Cannot access glossary di... Impossible d'accéder au ... He магу атрымаць доступ ... 133 Cannot create directory f... Impossible de créer le r... Не мару стварыць каталёт 134 Total = 378 Filter is OFF 131/378

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

Screen 2: The <u>Filter Settings</u> dialog box where you can set criteria for the entries to display.

| 🥦 Filter Settings | × |
|---|-----|
| Include entries that are flagged AND OB Include entries corresponding to the following <u>SQL</u> condition: | |
| Source LIKE '%test%' | < > |
| <u>Examples</u> <u>Fields</u> <u>D</u> perators <u>Functions</u> <u>C</u> haracters Enter an SQL conditional expression using the name of the fields (Key, Flag, ID, Source, Target, etc.) Use apostrophes to delimit strings: 'text'. Use pound signs to delimit dates: #10/21/2005# or #10/21/2005 4:53:30 PM# | |
| In all strings escape the apostrophe as ". In parameter strings for the LIKE operator escape the apostrophe as ", the opening square bracket as [[], the closing square bracket as[]], the percer sign as [%], and the astrisk as [*]. | |
| OK Cancel Help | |

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

| TMX Import Options | | × |
|--|--|---|
| Languages to import: EN FR BE | Merge entries Set the field: Merge Options Value: | < |
| All languages | Flag the imported entries Ignore region differences in language identifiers Check for invalid XML characters | |
| <u>S</u> ource: EN ✓ Target: FR ✓ | Ignore the following properties: RTFFont Table | |
| | OK Cancel Help | |

Screen 4: The <u>Flag Entries</u> 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.

| Flag Entries | |
|--|--|
| Flag the entries that correspond to at least one of the following selected conditions: <u>P</u> re-defined conditions: | |
| Duplicate Source and Target Target is empty, but not Source Duplicate Source Source and Target are both empty Duplicate Target Source or Target has inline code Source is the same as Target Source is empty Source is empty Source is empty, but not Target Target is empty Target is empty | |
| Rag also the <u>o</u> riginal of each duplicate | |
| User-defined condition (the expression is contained in at least one of the selected field | ls): |
| | |
| Flag | <u>M</u> atch case Match <u>w</u> hole word Use regular expressions Regular Expression Help |
| <u>R</u> eset flags before excuting the selections OK Can | ncel Help |
| | |

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