

PROMT Professional and the Importance of Building and Updating Machine Translation Dictionaries

LORENA GUERRA MARTINEZ
Dublin City University

This article explains the main features of the machine translation system: @prompt Professional (English&Spanish), used in the comparative tests carried out by Guerra (2004). This software, as will be explained, provides a friendly and quite productive environment for users editing raw machine translation output.

In addition, the importance of building and updating machine translation dictionaries will also be addressed since this is one of the possible options to increase machine translation efficiency. It not only allows for improvement(s) in the accuracy and time competence of an automated system but it also facilitates the task of post-editing. However, this is not always a simple task. To build or update an entry can sometimes affect the automatic translation of a source text in undesirable ways. In these cases, the user is neither saving time nor improving translation quality.

REVIEW OF @PROMT PROFESSIONAL (English↔Spanish)

INTRODUCTION

Project MT Ltd. (PROMT), located in Saint Petersburg (Russia), was founded in 1991 as a provider of translation software for the Russian language. These days, PROMT offers translation solutions for the Internet, corporate Intranets, personal computers, handheld devices and Wireless Application Protocols. These translation products also integrate with translation memory systems (Trados Translator's Workbench) and proprietary terminology extraction systems. Under the trade mark @prompt™, the company provides different professional and translation solutions based on the new translation technology “PROMT XT” (excellent Translation, XT). *See Fig. 1, PROMT family products:*

<i>Professional solutions</i>	<i>Translation Solutions</i>	
PROMT for Trados	<i>Desktop</i>	<i>Server</i>
	@prompt Express Professional	@prompt NET
(@prompt Expert	@prompt Internet	@prompt Intranet Server
PROMT Terminology Manager	@prompt Standard	@prompt Internet Translation Server
	@prompt Professional	@prompt Development Kit

Fig. 1. *PROMT family products* (professional and translation solutions)

A review of @prompt Standard, @prompt Professional and @prompt Expert is in preparation for submission to Multilingual Computing and Technology Magazine (Allen, 2003a). Besides, a brief review of @prompt Professional, is already available, Guerra (2003). For further information on these machine translation (MT) systems, their comparative features and/or prices, see <<http://www.e-prompt.com>>.

@prompt Professional, the MT system used in this project, includes two language pairs and directions (English↔Spanish), although only the first one (English-Spanish) was tested. The current software, based on the core PROMT XT (excellent Translation) technology, as already stated, differs from the system architecture explained in Sokolova (2002) in the introduction of further and improved user features.

DICTIONARIES

PROMT translation systems offer three different types of dictionaries:

- *a single large general-purpose dictionary* per language direction is provided by the software company. It is integrated into the translation engine during the initial installation. This general dictionary contains, as a rule, about 100,000 entries and it is used as the main terminological reference during the translation process. Since this dictionary is locked, its entries cannot be modified or deleted. However, they can be overridden manually by creating entries in user dictionaries and attaching them.
- *domain-specific dictionaries* are not supplied by the company, but they can be purchased separately, although they are not addressed in this project.. They contain terminology related to specific domains as

well as general-purpose words that have some specific meaning in that particular area. These dictionaries cannot be edited. However, the user can open an entry, modify it and then save it in the user dictionary. The volume of a domain-specific dictionary varies from 5,000 to 50,000 stems.

- *user or customized dictionaries* basically represent one of the most interesting parts of this research. As the user creates these dictionaries, they mainly focus on the terminological weaknesses of the text(s) under correction. They are also crucial in speeding up the process of post-editing. In the case of *@prompt Professional*, the new multi-dimensional dictionary architecture allows for more precise and advanced options such as a dictionary editor, explained in next section, or the optional installation of electronic dictionaries developed by different companies and covering various topics. These electronic resources (e.g., Lingvo, Context or MultiLex) allow users to obtain several translation variants for any word. As in the case of domain-specific dictionaries, they must be purchased separately.

DICTIONARY EDITOR



- Dictionary Editor.lnk

Dictionary editor is an application intended for creating and editing user dictionaries. It is available from any of the applications of the MT system, and it can even be accessed by opening the “start” menu, and then “@prompt family”. With a customised dictionary, users can adapt the software to their specific needs, gaining control and speed over the translation of words and expressions in certain documents. *@prompt Professional* incorporates new functions that allow users:

- to achieve greater precision when coding linguistic information, as indicated in *Fig. 2, Coding an entry*. The system can adapt dictionaries to user's skills by choosing a “beginner” or “expert” mode depending on the user's linguistic knowledge,
- to select “inactive translations” implying that some alternatives can be used as references rather than choices in the process of translation. *See Fig. 2, Coding an entry*, number “2”. They can be activated at any moment simply by clicking the mouse. Then, they appear in the text as translation variants in between brackets,

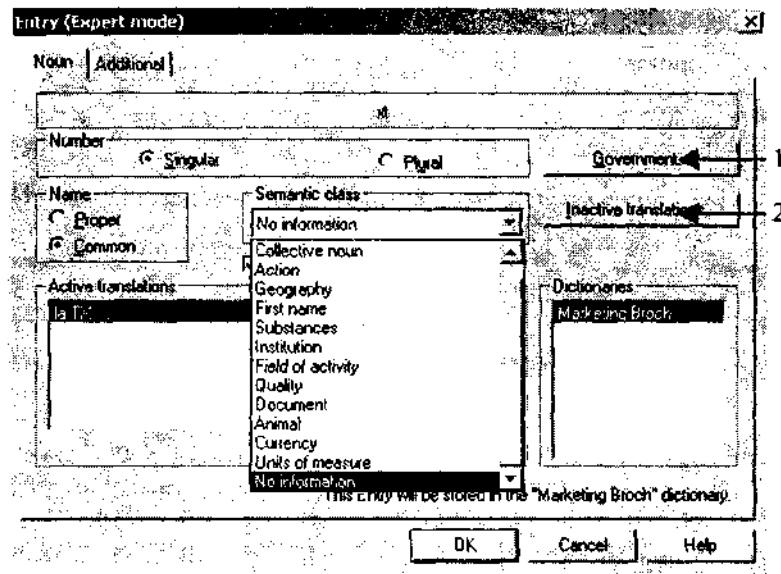


Fig. 2. Coding an entry (Expert mode)

- to improve the accuracy of the MT system by specifying some information on verb, noun and adjective syntax structure. This is possible by selecting the option "Government" only available in the expert mode. See Fig. 2 (ibid), number "1", and Fig. 3, Government,

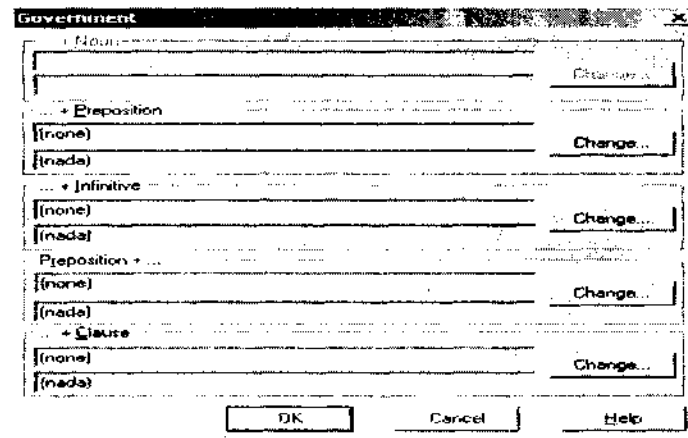


Fig. 3. Government

- to find any word in the dictionary, view the information about this entry in all the dictionaries of the system and, if necessary, to change the grammatical information. See Fig. 4, Entry references (a),

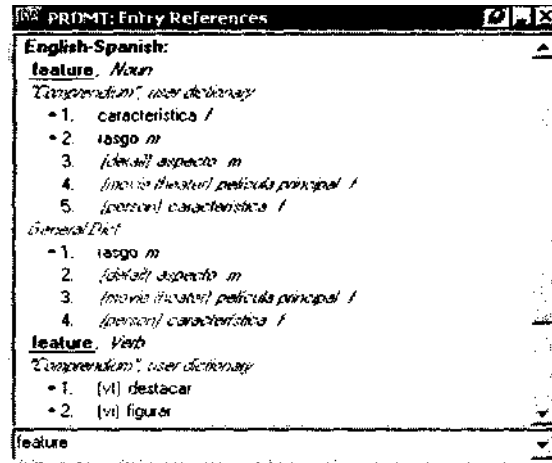


Fig. 4 Entry references (a)

- it is even possible to access information related to prepositions or conjunctions, which are not open for editing. See Fig. 5, Entry references.

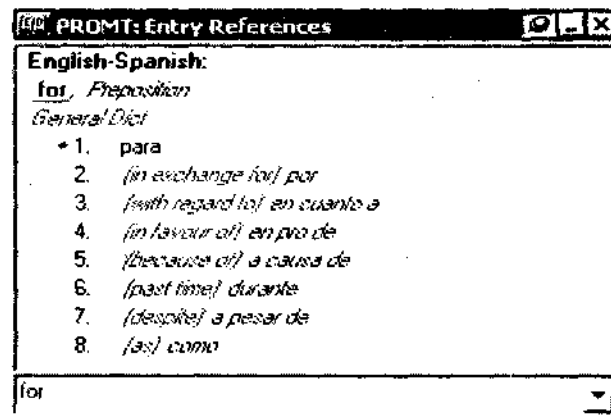


Fig. 5 Entry references (b)

- to set priorities, where the user establishes whether a word is more commonly translated as a verb or as a noun, as a noun or as an adjective. See fig. 6, Avoiding Ambiguities,

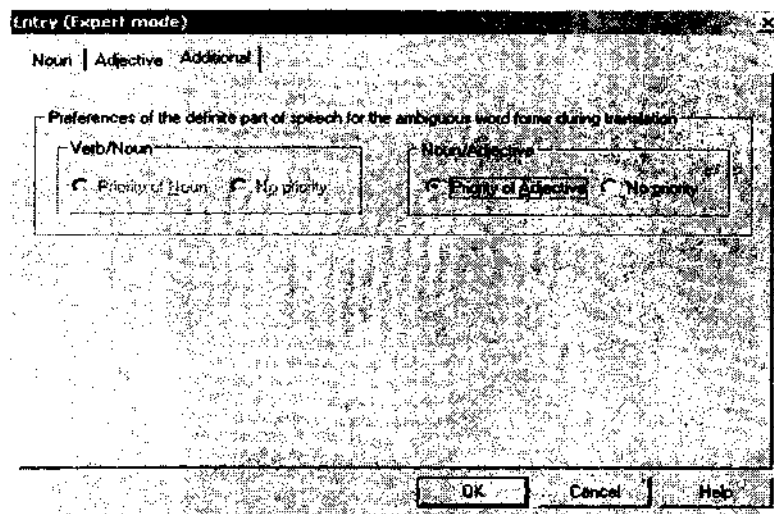


Fig. 6. Avoiding ambiguities

- to open and view the alphabetical list of headwords (stems, word forms and initial forms) in the dictionary,
- to display the alphabetic list of the dictionary as the ordered list of initial words or stems with the possibility of switching between these modes,
- to restore a deleted entry by selecting the “Restore” command from the “Entry” menu. *See Fig. 7, Restoring entries,*

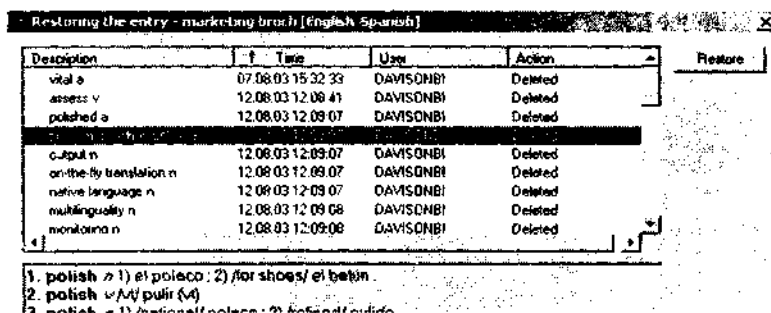


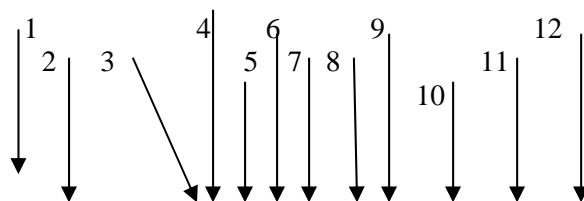
Fig. 7. Restoring entries

- to adjust the system to the topic of the text to be translated,
- to save the list of entries of an open dictionary in a text or RTF-file, or even in CSV (Comma Separated Values, a text-only formal used to store data separated by commas),
- to use the search capability in the dictionary and build an alphabetic list of entries that satisfy specific criteria,
- to save the source and final translated text as aligned bilingual texts, an advantage for combining with translation memory components.

These and other functions are outlined in Fig.8, *Commands in dictionary editor*. Their position in the toolbar is indicated in Fig. 9, *@prompt Professional: dictionary editor*.

-
- All dictionaries “1”
- To display information about the opened dictionary “2”
- To display the alphabetic list of the dictionary as the ordered list of stems “3”
- To display the alphabetic list of the dictionary as the ordered list of initial forms “4”
- To add new words or collocations into the dictionary which were prepared earlier in a text file “5”
- To enter a new word or collocation into a dictionary, or modify an existing entry “6”
- To open the selected dictionary entry for editing “7”
- To delete the selected entry from the dictionary “8”
- To form filter for searching in the dictionary “9”
- To modify filter for searching in the dictionary “10”
- To save the search results in a separate dictionary “11”
- To display information about dictionary editor “12”

Fig. 8. *Commands in dictionary editor*



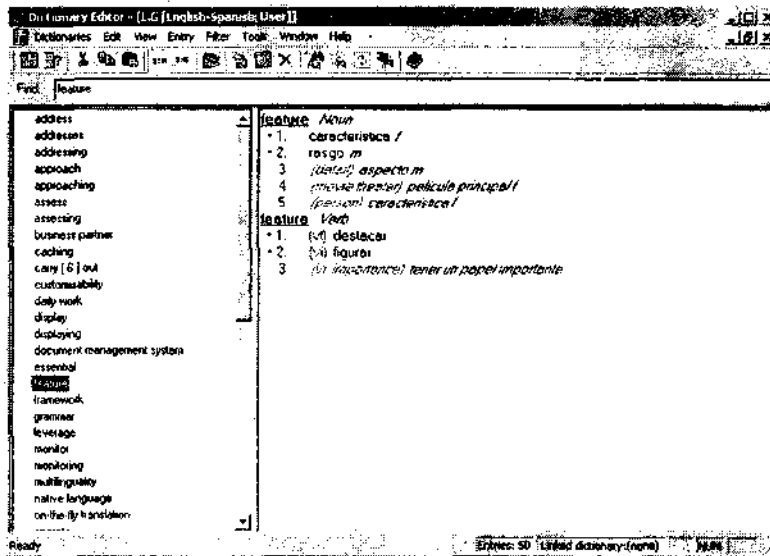


Fig. 9 @prompt Professional: dictionary editor

Specialised translation applications and general features

The following is a brief review of the specialised translation applications and general features available from the “start” menu, “@prompt family”



- installation of specialized dictionaries and topic templates.

TDSSetup is a program that allows the installation or removal of specialised dictionaries, even facilitating users in incorporating or removing different topic templates (lists of dictionaries, preserved words and pre-processors). These additional installations are not supplied by the company and must be purchased separately.



- Customize PromtE.Ink

Customise Plug-In for Microsoft Internet Explorer is a tab that allows users to modify the main options of the translation system. Thus, it is possible to change the direction of translation or the topic template, to edit

the list of topic templates for the selected direction of translation and even to select an automatic or semi-automatic topic template. See Fig. 10, *Customise PromtE*,

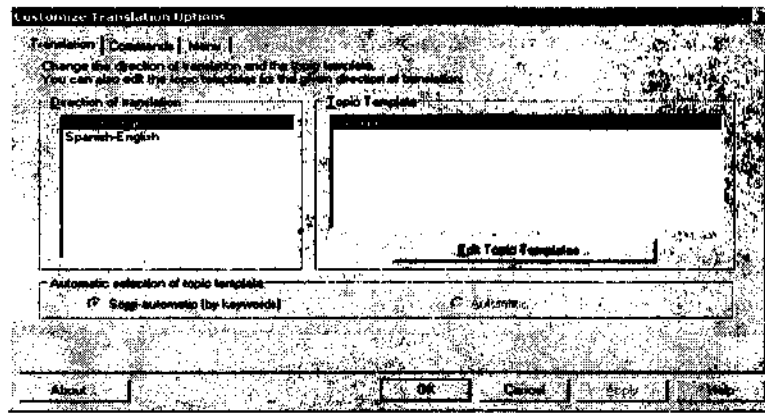


Fig. 10. *Customise PromtE*



- SmarTool Tuning.Ink

@*prompt Professional* can integrate translation capabilities into Microsoft Word, Excel, FrontPage, PowerPoint and Outlook 2000/XP simply by specifying the Microsoft Office 2000/XP applications in the SmarTool Tuning displayed after opening “start” menu and “@prompt family”. See Fig. 11, *SmarTool Tuning*,

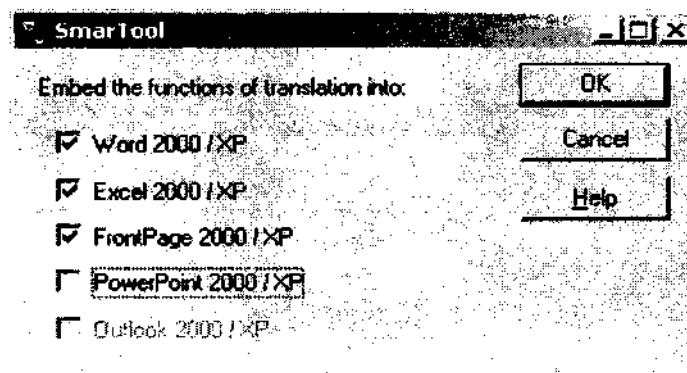


Fig. 11. *SmarTool Tuning*



- PROMT Backup.Ink

PROMT Backup is an auxiliary program for saving and restoring data and settings of applications included in the MT system. Thus, it is possible to save common objects of the translation system (topic templates, user dictionaries, associated memory databases or electronic dictionary bases) as well as user information.



- Promtx.Ink

PromtX is a simplified editor-translator that can automatically translate any text in Windows Clipboard or typed on the keyboard in any Windows application window. This module also permits users to work with .RTF documents, to open and translate documents in .doc and HTML formats, to read and translate texts from an Optical Character Recognition (OCR) program or to use the Microsoft Agent technology to speak the selected text. See Fig. 12, PromtX,

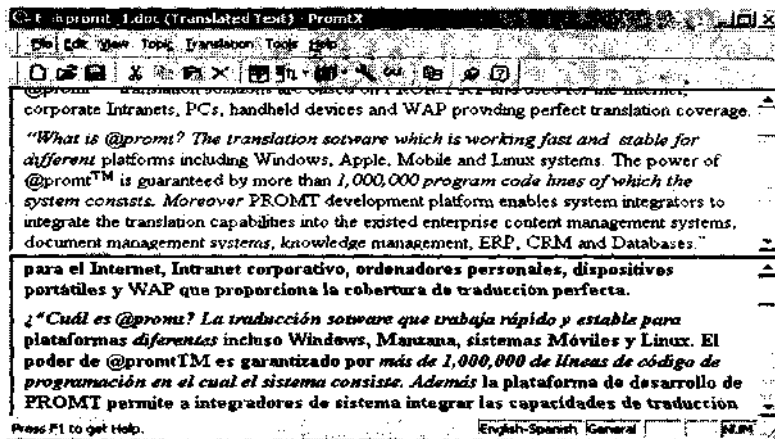


Fig. 12. PromtX



- Customize PROMT Plug-In for Acrobat.Ink

This module extends Adobe Acrobat 4.x-5.x and Adobe Acrobat Reader 4.x-5.x with the functions of text translation. It is designed for translation

of the preliminary selected text, omitting its formatting. The translated text is then displayed in a separate translation window as indicated in Fig. 13 (a), *Translation of .PDF files*. Additionally, Fig. 13 (b), *PROMT Plug-In for Acrobat*, indicates how Adobe Acrobat Reader 5.x can be embedded if required by the user.

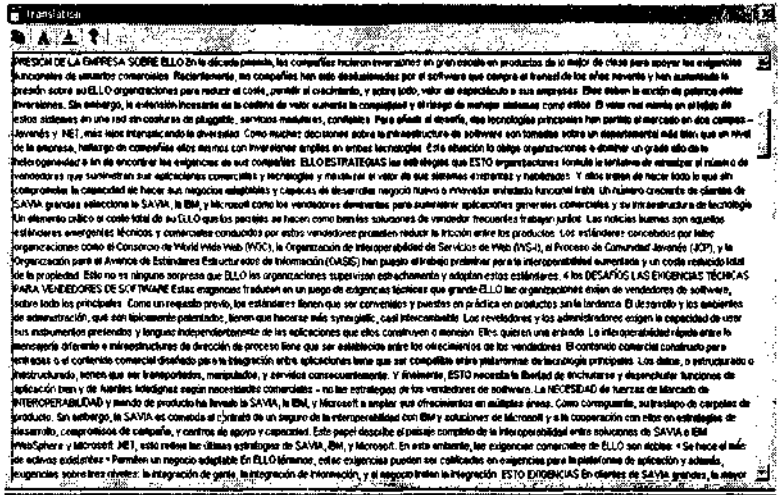


Fig. 13 (a). *Translation of .PDF files*

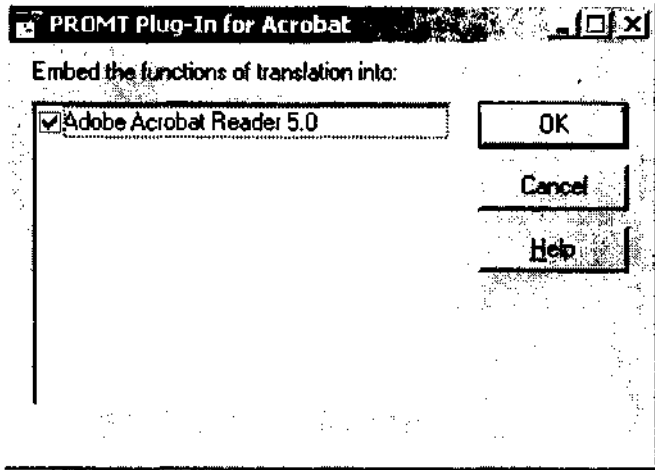


Fig. 13 (b). *PROMT Plug-In for Acrobat*



- Associated Memory Manager.Ink

This program is designed to work with Associated Memory (AM) databases. It permits the creation, deletion and manual editing of AM databases in any translation direction. Thus, in the process of translation, the system searches for identical fragments from previously machine-translated texts. If any matching is found, the system automatically takes the information from the database.

- Finally, the software permits the translation of the file types outlined in Fig. 14, *Opening a file*,

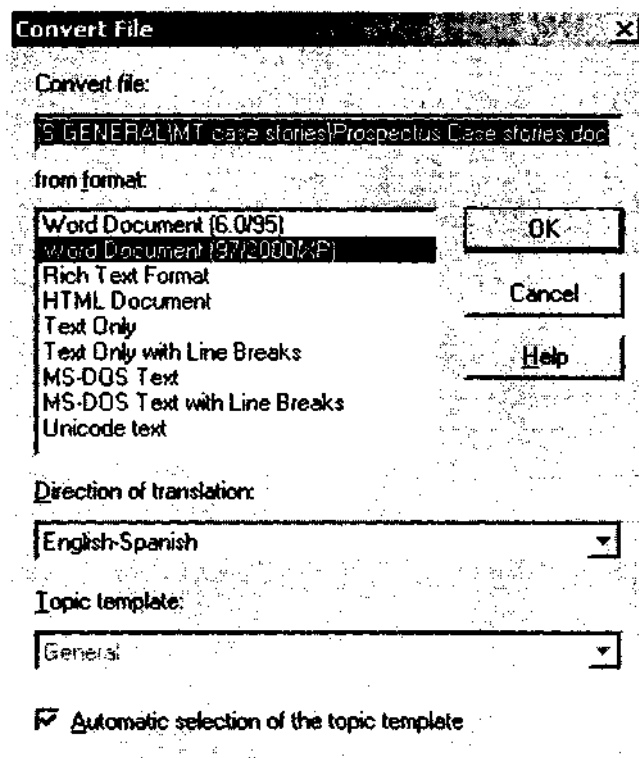


Fig. 14. *Opening a file*

EDITING FEATURES



- Promtlnk

@prompt Professional translation environment is based on a linguistic editor with a set of facilities which allow users not only to translate a text but also to edit the translation and update the dictionaries of all accessible directions of translation. This program allows the translation of simple (quick) and complex (professional) texts, together with a linguistic analysis of translation results, which may be used later to improve the translation quality of other texts of similar topics.

The following is a list of the most significant editing features offered by this software. Many of these functions have been introduced as improvements to the user interface over the last two to three years.

- Translation alternatives are given in curly brackets and underlined in blue (e.g., el trabajo de escribir{el papeleo}). Nevertheless, when new terms are coded into a user dictionary, these new variants of translation are underlined in purple (e.g., adaptabilidad{apoyo}) and presented in the order of priority specified by the user. If needed, users can choose to show or hide translation variants by pressing a button from the translation toolbar.
- Paragraph alignment is indicated by a mark whose (white, blue, black and green) colour depends on the status of the paragraph. Thus,
 1. White sign: the paragraph has not been translated yet;
 2. Blue sign: the paragraph has been translated;
 3. Black sign: the paragraph was modified since it was translated;
 4. Green sign: the paragraph has been marked as not requiring translation.
- If a word or phrase is selected in the target text, the source word or phrase is highlighted with the colour of the current paragraph in the source text, allowing users to locate easily the selected item. Besides, the corresponding information on the selected construction is displayed in the “Properties tab” of the “Info Pane”, as presented in *Fig. 15, Program interface*.
- Colour coding is provided for identification of unknown and preserved words, underlined in red and green respectively. Nevertheless, all

words in red change to pink when they are coded into an MT dictionary.

- It is possible to modify the colour of the selected text or even the background colour. Other options such as changing the font or the (lower/upper) case are also available from the "Format" menu.
- Hyperlinks are identified in the source text and it is even possible to access the Internet document or folder opened by the default browser.

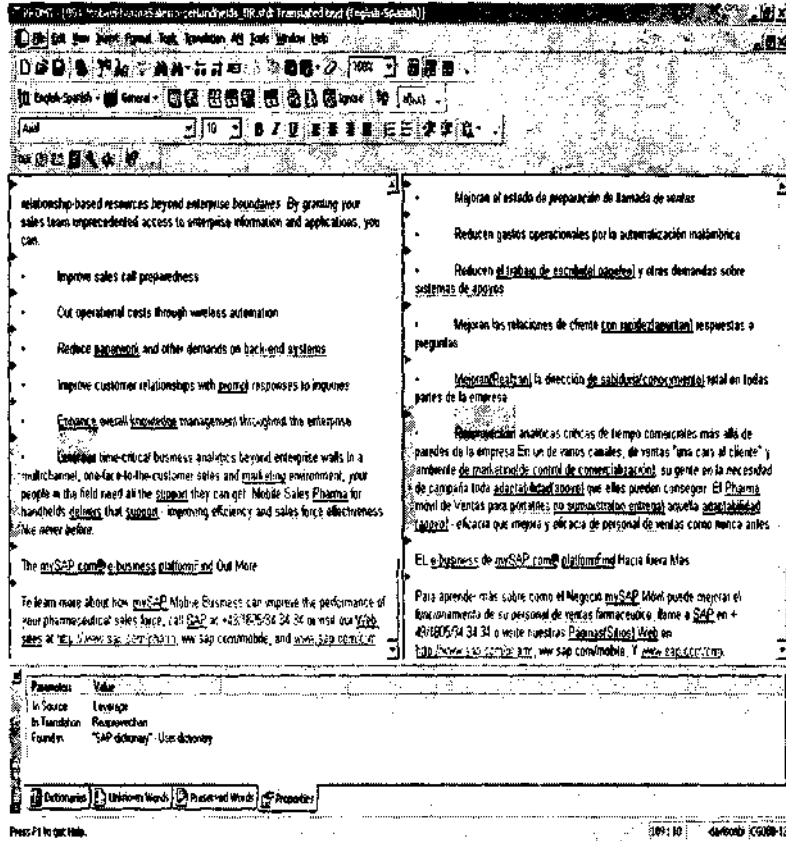


Fig. 15. Program interface

BUILDING AND UPDATING AN MT DICTIONARY

Building and updating a user dictionary is crucial for the effective implementation of any MT system. The time spent analysing the data, creating and updating dictionaries (at the beginning and during the entire translation process) will reap long-term rewards.

The following is a list of suggestions to consider while doing machine translation post-editing, especially to avoid miscoding and to use the tool faster. They are based on my experience while post-editing the marketing brochures presented in Guerra (2004), although Allen (2003b) also outlines these and some other points to bear in mind.

- When coding a new entry, it is very important to remember that, by default, the dictionary always selects the “Noun” tab. *See example in Fig. 16, New Entry, (“ofrece” is a verb),*

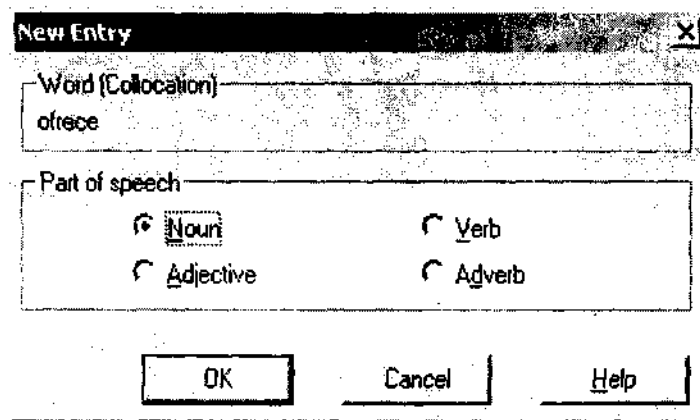


Fig. 16. *New Entry*

- If the user is coding an entry that already exists in the general dictionary but whose meaning is inaccurate for the current translation, it is possible to add and prioritise the "user" entry by moving it up. By creating and prioritising an entry, the user deletes the general dictionary mirror image entry from the custom dictionary.
- When coding an expression, the user must specify only the grammatical category of the headword. Then, the conjugation (or inflected forms) for each of the variable components must be added.

- When coding an entry that has been wrongly translated by the MT as a consequence of an inaccurate analysis (e.g., a noun is translated as a verb), the user must code the entry (the noun), in the verb form, in order for it to be identified and replaced by the system.
- Coding an entry in lowercase letters means that this word will be recognised in all typographical forms (e.g., lowercase and uppercase forms). However if an entry is coded with an initial uppercase letter, the system will recognise only the word form as it is displayed. This reduces MT efficiency and therefore, capital letters are used only when specific words or expressions need to be distinguished. For example, “WHO” (World Health Organisation) versus “who”
- If the source text has a word written in uppercase letters, e.g., "ROI" (stands for “return on investment”), and the target language does not have such an abbreviation, coding the corresponding meaning “rendimiento del capital invertido” implies that all these words in the target text appear in capital letters (inappropriate in this case). Allen (2003b) claims that the only solution for this problem is to code the entry with definite and indefinite articles as adverbial fixed expressions.
- To enter words with punctuation marks such as hyphens, commas or apostrophes, it is necessary to leave a space before the variable component and the punctuation mark. This is the only way in which the MT system will recognise these entries as being related and variable words.
- Sometimes words are not recognised by the software because in the source text they are spelt differently from what has been coded in the dictionary. Thus, an efficient way of improving MT efficiency is by coding all possible spelling variants when entering a word. A faster way to do this is by pressing “F8” and then, moving the up and down arrows in both the source word and the target language entry dialog boxes. Thus, the original word can be selected and edited or even deleted, if necessary.
- Accented characters are sometimes problematic. However, as this is a problem encountered when translating from Spanish into English and this is not the language pair under investigation, this point will not be elaborated upon here.

POST-EDITING GUIDELINES

The following is a list of suggestions to follow or, at least, to consider when doing full post-editing of marketing brochures. These guidelines are based on the various tests and difficulties encountered in the project carried out by the author (Guerra, 2004).

First of all, it is necessary to identify and to code into a user dictionary those terms unknown by the MT system, together with all mistranslations. This must be followed by an identification of every word and short expression that do not require translation and consequently, that need to be preserved (acronyms, abbreviations, company or product names, individual vocabulary words, etc.). Punctuation or misspellings should also be rectified to avoid further MT inaccuracies.

By coding new entries into a user dictionary, the quality of the preliminary raw MT output has improved only lexical issues. Therefore, more changes need to be made in order to produce a final target text indistinguishable from human translation. All features that, at a syntactic or grammatical level had caused difficulties for the MT system either in source text analysis, in transfer to the target language or in both, must be removed. The following are some suggestions:

> DELETE:

- words or expressions that are too literal to keep in a marketing brochure,
- unnecessary definite or indefinite articles, especially in lists,
- subjects that are superfluous, (in Spanish, verb forms often make subject pronouns unnecessary),
- repetitive words or phrases. Ideally, code them into a user dictionary as translation alternatives.

> CORRECT:

- number and gender disagreements,
- syntactical errors mainly due to word order,
- wrong or inaccurate meaning of prepositions,
- nouns followed and/or preceded by one or two adjectives, where the system cannot determine which is the head of the noun phrase,
- anaphoric forms if the antecedent is wrongly assigned (personal and demonstrative references; nominal, verbal or clausal substitutions; nominal, verbal or lexical ellipsis).

> WHEREVER POSSIBLE, TRY:

- to look for synonyms, and to code them as variants of translation. This will avoid the repetitive style caused by MT consistency,
- to be explicit, to simulate the performance of a human translator. This will avoid the literal output provided by the software.

> DON'T:

- follow general rules of word order and, in most cases, place adjectives before nouns or noun phrases to state specifically the attribute quality of the noun group,
- rewrite the entire sentence, paragraph or text. Use raw MT output as a base for further (stylistic) modifications.

> REMEMBER:

- these guidelines can be further developed with practice and a better knowledge of the software and the area or areas tested.

CONCLUSION

It is very important to bear in mind that the best MT software is quite far away from replacing professional human translators. This language technology is an aid not a replacement. The ways in which MT engines are employed in large multilingual organizations and international companies open up new work avenues for translators. This technology is an added and challenging component of today's global multilingual documentation production. As pointed out by Lou Cremers, Translation Services Manager at Océ Technologies:

MT should be regarded as enabling technology that speeds up routine work, leaving the translator to apply special domain and language skills (Euromap 2002, pages not numbered).

The successful implementation of this technology relies on a software application with built-in editing features, as this will facilitate machine-human interaction at the back-end of any translation process. Without doubt, *@prompt Professional* has demonstrated its friendly and quite productive environment for users editing raw MT output. Now, it is up to the user and the translation community, in general, to accept and implement human-

assisted machine translation (HAMT). So far, I have proved that the right software, if properly used, can demonstrate its efficiency even when translating materials other than highly repetitive technical and/or scientific manuals.

REFERENCES

- Allen, J. 2003a (forthcoming). Review of @prompt Standard, @prompt Professional and @prompt Expert. To be submitted to *Multilingual Computing and Technology Magazine*.
- Allen, J. 2003B (forthcoming). *Guidelines for MT Dictionary Building and MT Post-editing*. Current document version: v0.03 dated 08 July 2003.
- Euromap Language Technologies. 2002. How HLT applications can lead to higher quality translations at lower costs—. The experience of Océ Technologies [Online]. Available from: <<http://www.hltcentral.org/page-1026.shtml>>.
- Guerra, L. 2003. Review: Machine Translation with @prompt Professional. *IN: Irish Translators' and Interpreters' Association (ITIA) Bulletin*. Ireland, pp 5-6. Also available at PROject MT web site: <<http://www.prompt.ru/news-e/index.phtml>>. October 2003.
- Guerra, L. 2004. *Summary of: Human Translation versus Machine Translation and Full Post-Editing of Raw Machine Translation Output*. Master's Thesis. Dublin City University. Ireland. *IN: International Journal of Translation*, March-April 2004.
- Sokolova, S. 2002. An Overview of PROMT Translation Software. *IN: Multilingual Computing and Technology Magazine*. 13. (3). April/May 2002. 43-46.

LORENA GUERRA MARTINEZ
DUBLIN CITY UNIVERSITY
IRELAND
E-MAIL: lorena.guerra2@mail.dcu.ie