Making MT Commonplace in Translation Training Curricula – Too Many Misconceptions, So Much Potential!

Elia Yuste Rodrigo

Computerlinguistik, IFI, Universität Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Schweiz – Switzerland
yuste@ifi.unizh.ch

Abstract

This paper tackles the issue of how to teach Machine Translation (MT) to future translators enrolled in a university translation-training course. Teaching MT to trainee translators usually entails two main difficulties: first, a misunderstanding of what MT is really useful for, which normally leads to the misconception that MT output's quality always equals zero; second, a widespread fear that machines are to replace human translators, consequently leaving them out of work. In order to fight these generalised prejudices on MT among (future) translators, translation instruction should be primarily practical and realistic, as well as learner-centred. It thus ought to highlight the fact that: 1) MT systems and applications are essential components of today's global multilingual documentation production; 2) the way in which MT is employed in large multilingual organisations and international companies opens up new work avenues for translators. This will be illustrated by two activities, one using commercial MT systems for quick translations, whose process outcome is improved through the trainees' interaction with the system; the other focusing on MT output comprehensibility by speakers of target language only. MT is thus a mainstream component of a translation-training framework delineated in Yuste (2000) that, by placing the trainee in workplace-like situations, also echoes Kiraly (1999).

Introduction

As a result of having successfully implemented a translation-training module that benefited from translation technology, this paper presents some teaching experiences with particular reference to MT instruction. Let us clarify that this was not a module exclusively devoted to MT (even if it would be worthwhile having one) but highly *enhanced* by it.

Translation Training Setting

The translation-training scenario contemplated in the paper takes place at university level. Nonetheless, the training principles here exposed could be applied to other translation training settings, which may often result from an urgent need for updating knowledge and fast developing new translation market skills, e.g. professional development seminars dealing with translation aids, held at university or elsewhere.

In this particular case, the translation training population consisted of final-year Modern Languages and Translation undergraduates from Spanish universities attending a translation course at a British university. They all had limited background in computing and translation technology applications. The lecturer, here the translation trainer, is specialised in Machine Translation.

Challenges behind MT Instruction for Translators

It is crucial to underline that, no matter how often specialised commentators and MT researchers argue that MT, and translation technology, is not aimed at replacing human translators, the latter still remain suspicious and somewhat reluctant to adapt to global translation market requirements.

Even if these fears and prejudices are usually not so strong in the younger generations of future translators, we still find quite a dramatic high proportion. An entry questionnaire administered at the beginning of term by the lecturer among a group of 17 students showed the following:

Over 80% of respondents (15 students) manifested to be concerned about the impact that MT inroads could make in the next few years, when they would still be finding their feet in the so-called **translation industry**. Most of them confessed not to know much about all the professions belonging to this industry.

Almost the same number of students, 13, strongly argued that a machine could never replace the quality offered by a human translator. Some of them (3) even pointed out that they did not understand why, despite their poor quality output, MT systems are being used so extensively by large governmental organisations and corporate entities. Obviously, this reflected a notorious knowledge gap on the process dynamics of multilingual documentation production, where there are many other imposing factors apart from quality of the end product.

In a question dealing with the adequacy of MT systems on the basis of text-type, only about 40% (that is, 7 students) believed that MT would be useful for translating technical instruction manuals or weather forecasts reports.

From these and other comments, it was inferred that their main concern was to find a job after graduation, *not to be seized by a machine*. They did not seem to fully know about the diversifying roles that translators could take throughout a translation project that makes use of MT or another translation technology application. Similarly, the fact that they had no clear idea about the usefulness of MT (according to which text types, in certain areas of bilingual or multilingual documentation production, etc.) impeded a better understanding on how or when they could use MT in their daily tasks as translators.

From an instructional point of view, it was therefore necessary at first to insist on proper explanations with an insight into the evolution of MT in its over four-decade history and some informed figures on the global impact that MT has been making recently, e.g. from ABI (1998).

After this indispensable explanatory stage, the training process would move into a more practical side that we describe below.

Machine Translation for Future Translators: From 'Food for Thought' to Empowered Translation Professionals

At this point, our future translators were theoretically acquainted with MT, its history, research trends, commercial systems, etc. It was high time that this new understanding was supplemented with more practice-oriented work.

Adopting Kilary's maxim that 'learning to be a professional translator means learning to act as one' (1999: 4), the next step would deal with MT as if these translators-to-be were already employing it in their workplace. The main goal here would be to collaborate (i.e. share and apply knowledge) in translation projects involving the use of commercial MT systems and Internet-based MT applications. The trainees would not only learn about the system's features but would also develop new professional and linguistic skills (e.g. preparation or text pre-editing, post-editing of MT output, MT lexicon updating, MT rules expansion, etc.).

This practical training stage will be exemplified here by two activities that, enhanced by MT, did nourish the human talent of our future translators.

***Activity One.*- Using MT to produce quick translations. Automation gets better through human intervention.**

After an introductory session on commercial MT systems (namely *Globalink Power Translator Pro 3.2* and *Systran Pro 2.0*, and its web version: http://www.systransoft.com/), the students were divided into two groups, one per MT system to work with and evaluate.

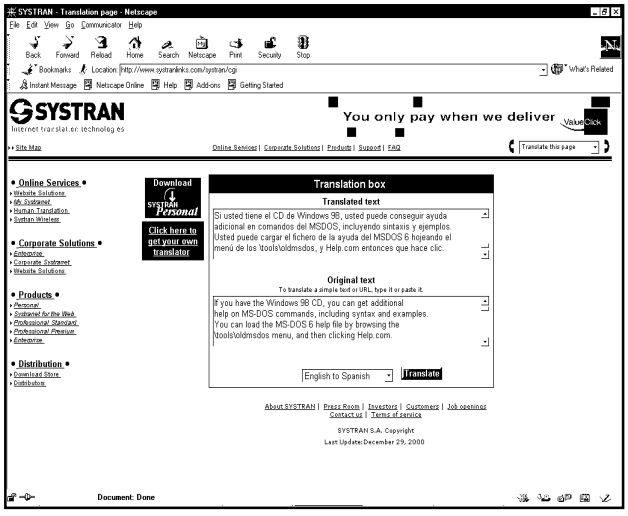

Figure 1: Using Systran on the Web: Translated text and original text.

The scope of this paper does not allow us to fully describe the evaluative work method that they followed. Yet the basic procedure consisted of providing the two groups with the same source text to translate (namely the *Microsoft Windows 98 README for MS-DOS Config.sys* file. Similar activities focused on other texts belonging to other specialised domains followed later on). With the help of another lecturer, it was possible to observe the two groups using a **think-aloud strategy** to discuss the limitations and assets of the systems used, before and while in front of the computer.

Both groups spotted the items that would be likely to cause problems to the system, due to their lexical or syntactic ambiguity (e.g. *command*) or for being domain-specific terms that would not yet belong to the system's lexicon (e.g. *config, readme*). When comparing translation outputs from the two systems, Systran programs (web and software product alike) tended to offer more fortunate translations of terms such as *command* (Sp. *comando*, in information technology (IT)/software context, as opposed to *el/la orden* appearing in the version produced by Power Translator Pro). It was then agreed that this was due to the fact that Systran's Translation Options allowed for the uploading of optional topical glossaries, e.g. computers, data processing, as shown in Figure 2.

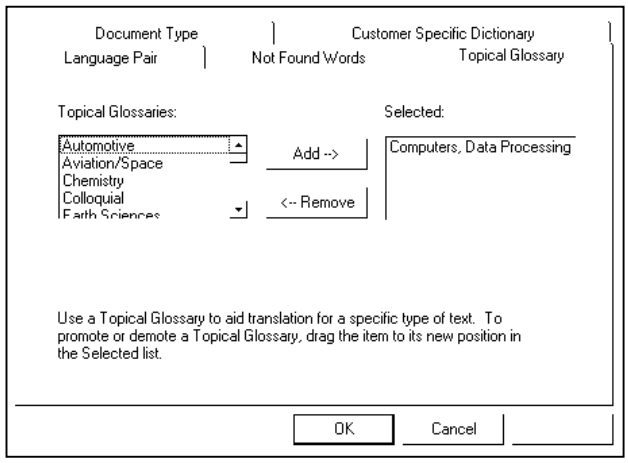

Figure 2: Systran Translation Options – adding

In this exploring of MT systems, it was extremely important to make future translators realise that they could **interact** with the software to get better translation results. In other words, they could contribute to and had an essential role to play in a **semi-automated translated process**.

This was represented by their intervention in the updating and improvement of the MT system's dictionary or lexicon. For instance, they learnt how to edit the dictionary entry for *command* in Power Translator so that it had the new Spanish translation of *comando*, more appropriate for this assignment and for any incoming IT-related projects. Figure 3 shows how to carry out this dictionary improvement in Power Translator Pro, by expanding the entry in question.

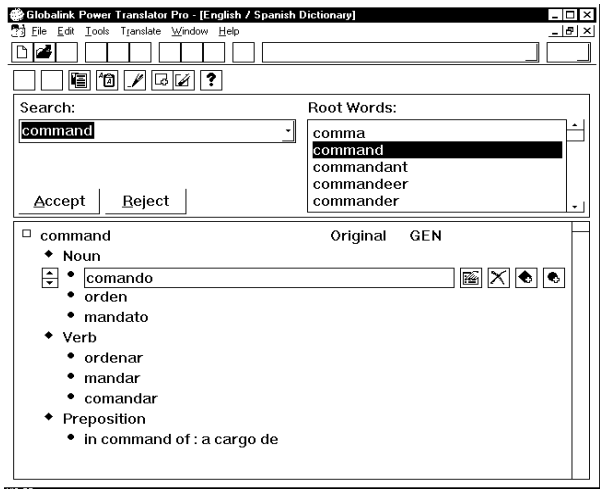

Figure 3: Power Translator Pro – English/Spanish Dictionary: Editing the entry *command*

This part of the activity laid the grounds for further MT system - user (i.e. translator) interaction that would come later on in the syllabus through the training in other translation technology applications, namely translation memory (TM) tools.

As a result of this activity, they reckoned that it would be much more challenging and rewarding to work with MT systems that allowed for translators' and/or terminologists' intervention (e.g. Power Translator's dictionary editing), unless systems were able to produce satisfactory results thanks to the integration of tailored or domain-specific components (e.g. Systran's topical glossaries) that, ideally, should be open to further expansion or improvement by the translator.

Furthermore, the more they were learning about the functioning of MT systems, the more they were keen to use MT in the future. The speedy process of translating with a machine became one of the most attractive aspects of MT. In fact, when throwing the above-mentioned source text to the systems, its 543 sentences were translated in two minutes. This helped them understand how MT systems are conveniently used in cases where on-the-fly translations are needed urgently. Although the quality issue still remained as an ultimate goal in their minds, there was now room for automating part, if not most, of the translation process.

Our future translators were getting ready for gradually developing new work skills conforming to the **multitasked translator**[1] profile, which is more and more needed in the translation market.

Next, they were asked to compile a list of potential users and situations in which MT could be particularly of use. It was interesting to note that some came up with the idea that freelance translators could benefit from MT to produce more translations in less time, and, consequently, make more money.

As this activity's epilogue, a piece of coursework was set to be handed in at the end of term. It consisted of an investigation into authentic users of MT technology in a particular industry sector, by a government/international organisation, etc. If possible, they had to provide information on how an authentic subject or party used MT in their translation workflow, either in the form of a report or an interview transcript. This was intended at raising their awareness about the full potential of machine translation from external real sources.

***Activity Two.*- Using a commercial MT system to understand foreign language text. Setting an evaluative experiment.**

From the previous activity, our future translators had learnt by themselves that 'computer-based translation systems are not rivals to human translators, but… aids to enable them to increase productivity in technical translation.' (Hutchins, 1997:1). Among the possible users that they had listed earlier on, subjects belonging to the communications sector had been mentioned, such as world-wide news analysts and reporters. This showed that they had also grasped the link between MT systems and their functionality as 'aids for communication' (Hutchins, 1997: 3).

In order to test this facet of MT, they embarked on a simple evaluation experiment about intelligibility of MT output by individuals with zero knowledge of source language, i.e. speakers of target language only.

The goal was to check whether English-speaking subjects with neither Spanish knowledge nor linguistic background could really get the gist of a machine-translated newspaper article originally written in Spanish. The article, which had to do with the *Golden Globe Award* to the Spanish filmmaker Pedro Almodóvar for his film *All About My Mother*, was taken from *El País Digital* (http://www.elpais.es).

Each student passed on the rough translated article, without the original article, together with an end-user questionnaire to a friend or acquaintance of the above-mentioned characteristics. The questionnaire contained a scale of comprehensibility, clarity and overall acceptability (0-5, 5 representing highest level) that respondents had to use in their answers. These were also asked to determine, if any, segments of the text that were hardly understandable or particularly odd, striking or ambiguous in any way. Most respondents coincided in that they did manage to understand 70-80% of the article, despite the funny *All envelope my madre* instead of *All about my mother*. But this relatively high level of intelligibility was also due to the fact that Almodóvar is well known by the British public, who also knew about the new film and the award.

Thanks to the respondents' comments, the students compiled a list of *laughable horrors* followed by full explanations of why the system did not translate

[1] In order to ensure optimum work prospects and client satisfaction, translators now play a number of varied tasks, which would not correspond to the traditional translator's role. Far from producing translation work *only*, they are expected to keep pace of information technology advancements applied to their profession, such as the latest translation technology, Internet resources for translators, etc. This enables them not only to optimise their daily linguistic activities, e.g. computer-aided terminology research and management, but also to market their translation skills, be a key team player or manage a translation project in a global setting.

adequately plus one or more human versions. Nonetheless, they were aware of the fact that the raw output firstly produced by the system was open to an improvement made on their part, by expanding or editing dictionary entries, making the program translate in sentence mode instead of in document mode, etc.

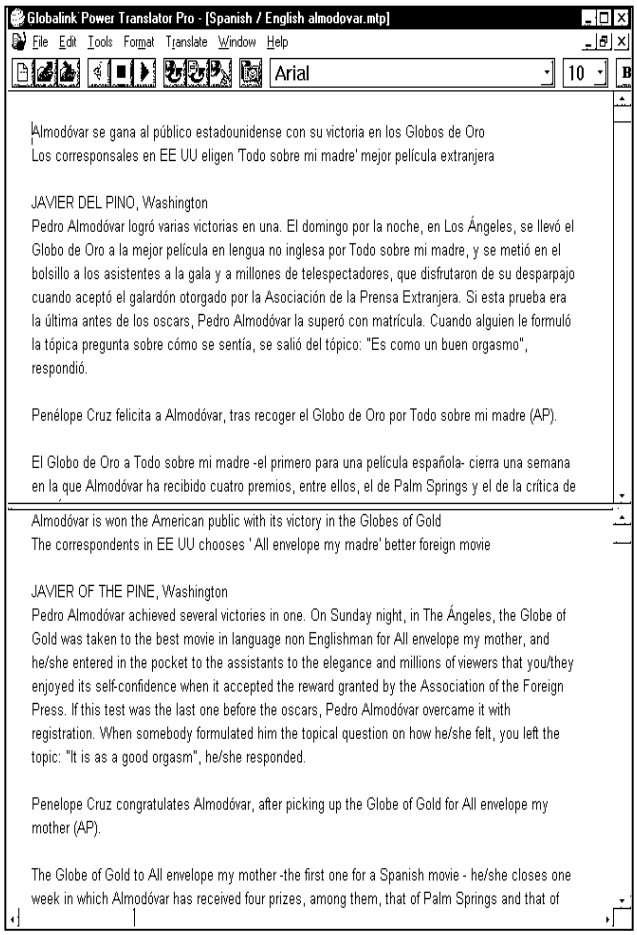

Figure 4: Globalink Power Translator Pro: Spanish > English

This helped them enormously not only to get even more familiarised with MT systems' limitations and output improvements (i.e. post-editing), but also to raise their cross-linguistic and cross-cultural awareness. For instance, they noted cases where the system resembles a misinformed rather than a bad translator, especially when producing multiword items that are grammatically correct but culturally invalid, as they are simply not expressed in that way in the target language.

More MT Teaching Ideas

We have presented two activities where future translators learn about the possibilities of MT by exploring with the proper guidance. More preparation and intervention on the lecturer's part may be needed to devise other MT-teaching activities that make use of corpora or other language technology techniques, such as alignment[2]. Alignment tools are commercially available and have been formally defined as 'software for the creation of bilingual text databases where sentences (or phrases) of source texts are linked to corresponding text segments of a target language' (Hutchins, 2000). In this line of work, our students were introduced to the alignment utility of the Viewer Tool in the text-processing suite **WordSmith Tools**[3].

The MT/translation lecturer can use such utility to present a new text, out of two files, with interspersed sentences to compare a translation with its original. In our training scenario, translation refers either to human or machine output. Apart from comparing originals with translations, it is worth confronting potentially different translations of an original to perform cross-translational analysis (translation vs. translation, instead of original vs. translation). To serve our purposes, we aligned MT output with another translation taken as reference[4]. Traditionally speaking, this reference version would have been fruit of a human mind only. However, if handling a multilingual official document now, it may well be that in the production of that English, Danish or Italian versions, there has been a great deal of automation in the translation process. After pointing that out, the aim is that students compare and analyse this polished and publicly accepted version with the raw version produced by one of the MT systems we have access to.

The teaching possibilities are many, but what is clear is that students have to keep exploring data, get involved, and draw their own conclusions as language professionals.

Conclusion

Our goal has been to present ideas that may prove useful to other translation trainers that are willing to teach MT to future translators for the first time.

Thanks to this practical framework, our trainees changed from a naively reluctant to an informed and welcoming standpoint as regards to MT and other translation technology components that followed in our syllabus. This also helped them to act more confidently and resourcefully in their coursework and future assignments as translators.

Acknowledgements

We wish to thank the participation of the students who attended the English<>Spanish Translation module at *Staffordshire University*, UK, academic year 1999-2000.

[2] Alignment is a text processing technique that takes one string from language A and subsequently places the same string from language B, repeating the operation throughout the text, until a new aligned bilingual (or multilingual) version of the text is presented. This is useful to visualise linguistic contrasts or translation problems as well as additions and omissions, style changes, word order preferences, etc, and thus constitute a very helpful exploitation technique in the translation-training classroom. Moreover, alignment lies behind the principle of translation memories, which would deserve a more careful attention out of this paper's scope.

[3] More information on this tool suite can be obtained from his author, Mike Scott, by e-mail or at his website at http://www.liv.ac.uk/~ms2928/ and the tools installation website at http://www.lexically.net/wordsmith/install.htm.

[4] This is a language version (say the English one) of a text, originally produced in more than two languages, that the lecturer can take as a reference, i.e. some kind of linguistic model for classroom analysis. Ready available multilingual corpora or *lecturer-compiled* corpora (as explained in Yuste (2000)) represent an excellent source.

References

Abaitua, J. (1999). Is it worth learning translation technology? In 3rd Forum on Training Translators and Interpreters: New Directions for the Millennium. 12--15 May 1999. (pp 209--220)Vic, Spain: Universitat de Vic. http://sirio.deusto.es/abaitua/konzeptu/ta/vic.htm (as of April 2001).

ABI – Allied Business Intelligence Inc. (1998). Language Translation: World Market Overview, Current Developments and Competitive Assessment. Oyster Bay, NY, USA. http://www.infoshop-japan.com/study/ab3365_languagetranslation_toc.html and http://www.alliedworld.com/product/TRAN00.html (as of April 2001).

Balkan, L., D. Arnold and L. Sadler (1997). Tools and Techniques for Machine Translation Teaching: a Survey. Colchester, Essex, UK: University of Essex. http://clwww.essex.ac.uk/group/projects/MTforTeaching/index.html (as of April 2001).

De la Fuente, J. M. (1999). Glorias y miserias de la traducción automática. In Jornadas de Lengua e Informática. Asociación de Técnicos de Informática – SIMO. Madrid, Spain. 4 November 1999. http://www.ati.es/gt/lengua-informatica/simo1/SIM1_jmf.html (as of April 2001).

Hutchins, J. (1997). Translation Technology and the Translator. In Proceedings of the 11th Conference of the Institute of Translation and Interpreting (ITI) (pp. 113--120). The Crown Hotel, Crown Place, Harrogate, 8--10 May 1997. London, UK: ITI. [Also in Machine Translation Review, issue no. 7 (April 1998), pp. 7--14] http://ourworld.compuserve.com/homepages/WJHutchins/ITI-97.htm (as of April 2001).

Hutchins, J. (1998). Computer-based translation tools, terminology and documentation in the organizational workflow. In Proceedings of the International Conference on Professional Communication and Knowledge Transfer, Vienna, 24--26 August 1998. 4th Infoterm Symposium: Terminology work and knowledge transfer — Best practice in terminology management and terminography. (pp. 255--268).Vienna, Austria: TermNet. http://ourworld.compuserve.com/homepages/WJHutchins/Vienna.htm (as of April 2001).

Hutchins, J. (2000). The IAMT Certification initiative and defining translation system categories. In online Proceedings of EAMT Workshop 'Harvesting Existing Resources', Ljubljana, May 2000. http://www.eamt.org/archive/ljubljana/Hutchins.htm (as of April 2001).

ISLE – International Standards for Language Engineering (2000). The ISLE Classification of Machine Translation Evaluations. Draft One, October 2000. http://www.isi.edu/natural-language/mteval/ (as of April 2001).

Kiraly, D. (1999). From teacher-centered to learning-centered classrooms in translator education: control, chaos or collaboration? In Innovation in Translator and Interpreting Training (ITIT) – An on-line symposium (17--25 January 2000). http://www.fut.es/~apym/symp/kiraly.html (as of April 2001).

Maegaard, B. (ed.) (1999). Chapter 4: Machine Translation. In Hovy, E., N. Ide, R. Frederking, J. Mariani and A. Zampolli (eds.). Multilingual Information Management: Current Levels and Future Abilities. Report commissioned by the US National Science Foundation and also delivered to the European Commission's Language Engineering Office and the US Defense Advanced Research Projects Agency. http://www.cs.cmu.edu/~ref/mlim/chapter4.html (as of April 2001).

Rico Pérez, C. (2001). From Novelty to Ubiquity: Computers and Translation at the Close of the Industrial Age. In Translation Journal, 5 (1). January 2001. http://www.accurapid.com/journal/15mt2.htm (as of March 2001).

Sager, J. C. (1993). Language Engineering and Translation. Consequences of Automation. Amsterdam, Netherlands / Philadelphia, USA: John Benjamins.

Yuste, E. (2000). Translation Instruction in the Y2K – Electronic Corpora, Internet and Translation Instruction. In CD-ROM Proceedings of the 7th Conference of the International Society for the Study of European Ideas (ISSEI), Section V, Workshop No. 501 'Teaching Translation at the Information Age'. Bergen, Norway: HIT Centre at the University of Bergen. 14--18 August 2000.