Recycling MT: A Course on Foreign Language Writing via MT post-editing

Ana Niño

Centre for Computational Linguistics, UMIST
PO Box 88, Manchester M60 1QD, England
A.Nino@postgrad.umist.ac.uk

Abstract

Far away as we are from FAMT (Fully Automated Machine Translation), we soon learnt to get the most from MT raw output. We learnt to see the bottle half full and found many sound applications in the translation industry. However, just a few have dared to explore the possibilities of using MT output for language learning.
This article seeks to explore the potentials of MT post-editing as a teaching method to promote foreign language writing skills.

Keywords: MT output, learners output, MT post-editing, writing skills, language production, error analysis, computer-mediated communication

1 Introduction

This paper considers the teaching of foreign language writing skills via MT post-editing. We will focus on the recycling of raw MT output for language learning.

The idea of reusing a system for different applications seems especially meaningful for researchers and commercial product developers exploring new areas in demand.

We acknowledge that MT can assist in breaking down the language barriers. Although current MT systems are far from perfect, they are actually providing an inexpensive solution for companies facing a growing international competence.

Writing nowadays constitutes one of our main means of communication online. Furthermore, writing is an important part of our personal experience and social identities and we are often evaluated by our control of it (Hyland 2002).

We claim that there is a correlation between writing skills and translation ability (Koenigs 1990). Moreover, Cook (2001) claims that there is a growing awareness that translation can develop accuracy since translation can focus attention upon subtle differences between L1 and L2 among other things.

Our discussion will focus on MT post-editing as a new skill that conflates translation and writing practice in the foreign language classroom.

2 Background

The use of MT in language learning has not yet received the attention it deserves. We believe this may be due to the belief that raw MT output is counterproductive for language learners, who should be exposed to good, representative examples of the target language (Krashen 1985). We do not believe, however, that this shortcoming should overshadow its compensations. For instance, Belam (2003:3) mentions source language competence and linguistic awareness as two transferable

skills to be acquired via MT. In her words, "[These skills] will, we hope, equip the students better for a career in translation or at least in languages".

In spite of being quite a recent and unexplored area, many authors (Anderson 1995, Corness 1985, French 1991, Shei 2002 [1]) support MT as a tool to improve second language skills, and as a useful vehicle for language learning.

Lewis (1997) proposes three main uses for MT:

- As a **rapid translation** of short, repetitive and standardised texts such as minutes of meetings and parliamentary questions.
- For **browsing** in texts whose language the user does not understand.
- For **drafting**, i.e. the user writes a text in his mother tongue and requests draft translations in other languages, presumably to cross-check the results.

As for the use of MT in the classroom, French (1991) quotes an article by Rodney Ball on the use of Micro CAT. In this article Ball distinguishes three modes of application of MT in the classroom:

- As a means of learning more about the foreign language (in this case, the **language** is the object of interest).
- As a way of introducing students to the general topic of MT (in this case **MT** is the object of study).
- As a way to show students how to use the various facilities available on a specific system and encouraging them to evaluate the systems'

knowledge (in this case the **software** itself is the object of interest).

Furthermore, French (1991:62) encouraged his students to make use of pre-editing and determine "which modifications would be needed to make the text more machine-friendly (pre-editing), what supplementary dictionary items should be added or what modifications are required to existing dictionary entries and what shortcomings in the raw translation should be left to the post-editing phase". These tasks could make sense as further training for trainee translators but seem pointless in a language course.

Another author that points us to the use of pre for language instruction is Richmond (1994) who made use of pre-editing to improve the language students' writing skills. He made use of the French Assistant system to translate the source language text in an interactive mode. Then he would ask the students to compare his proposed translation of the text to the target language text. Afterwards, the students would adjust the source language text to bring French Assistant's translation and the model translation closer together. Later they would print out the target language text and the modified source text using French Assistant' side-by-side printout function. This exercise, according to Richmond (1994:75), promotes effective language learning focused on the process and on the output of MT. He concludes that "learners who hear and see language that they must decode for meaning go further and faster in acquiring grammar than do those who only get a staple diet of exercises". This serves to support our belief in the use of authentic material in the language class. Ideally, the student

must be provided with a wide variety of text and audio material to work with rather than the traditional handouts with drill exercises to practise one specific grammar point without any context at all.

Shei (2002 [2]) also thought of using MT pre-editing as a tool for writing compositions for the English lesson. Shei proposes Chinese students of English write down their compositions in English and make use of an MT system to correct them by pre-editing the target language text (English) until the source language text (Chinese) sounds better to them.

Last, but not least, Somers (2001:3) also supported the use of MT in the class. He highlighted the idea that translation, and therefore MT, should go hand in hand with language learning. In his words, "…inasmuch as translation is often part of foreign-language learning, we can say that learning about MT and CAT tools should be part of the curriculum for language learners".

So far we have discussed the state of the art applications of MT in the class. Most of them referred to the idea of using MT pre-editing. The next section will focus on our area of research, i.e. MT post-editing in the foreign language class.

3 Post-editing as a teaching / learning tool

MT post-editing consists of revising the rough MT output. Traditionally, translators are the ones who have been trained in revision and MT.

It is important to distinguish between translation and post-editing MT. We should highlight here though that "the types of errors produced by MT systems do differ from those of human translators" (Hutchins and Somers, 1992:3).

As for translation, Krings (2001:7) points out that "When human beings translate, they construct meaning from the sentences they read, then they take this meaning and express it in another language, taking into account all of the nuances of the source and target cultures, the textual world of the text in both cultures, and their knowledge of the languages involved and the differences between them". However, With MT Post-editing, "the focus is on adjusting the machine output so that it reflects as accurately as possible the meaning of the original text"(Vasconcellos 1987:411).

We acknowledge that there is a correlation between revising human translation and revising MT output.

Mossop (2001) mentions the following skills as fundamental in revising human translation: copyediting, stylistic editing, structural editing, checking for consistency, use of computer aids, revision parameters, degrees of revision, self-revision and quality assessment.

In the same line, Wagner (1987:75) quotes the following skills as essential for MT post-editing: (1) excellent knowledge of the source language, (2) perfect command of the target language, (3) specialized subject knowledge, (4) word processing expertise and (5) tolerance. The latter requirement can be explained by the fact that machines are more consistent and therefore repetitive in their errors. With reference to this point, Löffler-Laurian (1986:227)

compares the situation of MT post-editors with the situation of teachers "who are frequently faced with repetitive errors when they correct written student assignments".

Both translation and MT post-editing involve lexical and grammatical skills, general reading and writing abilities, computer expertise and amending or editing skills. Editing is actually part of the teaching methodology for writing. In using MT post-editing as a teaching method, we should take into consideration the types of errors made by the language students and by the MT system. This could provide us with some insights as to the areas the students need to improve.

All in all, we acknowledge that the MT output, like any non-revised trainee translation or students' writing, is characterised by linguistic deficiencies. However, we believe that MT output provides basic linguistic structures which learners can build on and thus improve their language competence. We therefore aim to show that language students and/or translators can benefit from rewriting or correcting these faulty texts to improve their foreign language writing skills.

The novelty of this proposed research draws on the fact that MT post-editing into the target language is acknowledged as a language activity to be undertaken not only by translators, but also in the advanced foreign language lesson.

4 Objectives of the study

The main aim of this research is to investigate whether MT Post-editing activities into the target language enable advanced students of Spanish to develop their writing skills and learn new vocabulary and syntactic structures. Post-editing MT output is intended help students enhance their knowledge and competence in the production of various sorts of text. Thus, the teacher will be using MT as a teaching resource to design exercises to practise specific grammar points, as well as vocabulary (word choice, collocations, etc.) in different types of text. On the other hand, the student is presented with an exercise where s/he will have to concentrate on correcting errors to improve the translation. Given that the students already have a consistent knowledge of the language, they will be able to learn from the correctly translated part and learn from correcting the errors. The focus would be on the process of text production as a continuous problem-solving task.

In an MT post-editing environment, for students to correct MT output is different from and only a step towards correcting their own errors. That is, by making students correct the MT output, we are hoping that they may gradually acquire linguistic accuracy and become more confident in editing their own writing.
Our ultimate intention is to propose a pedagogical framework that considers MT Post-editing within a task-based methodology for the foreign language lesson.

5 Pilot studies

We started with a group of thirteen students of Advanced Spanish. They were asked to post-edit the faulty MT output from an advice column into more correct Spanish. They were provided with the source text in English and they

were allowed to use their dictionaries and reference books. Their errors were compared with the system errors (see table 1 below).
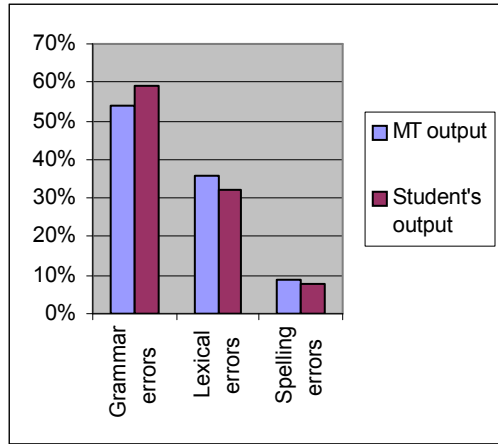

Table 1: MT vs. students' errors[1]

From the results we conclude that the students and the MT system had very similar error categorizations as far as grammar is concerned. Within grammar errors we found: confusion in the use of *ser / estar*, problems with the subjunctive mood, wrong use of possessive pronouns, confusion as to the use of the indirect object pronouns, agreement mistakes, problems with the relative pronouns and using infinitive for gerund or gerund for infinitive. Additionally, we found two word order errors only in the MT output, which were pretty obvious for the students to correct. We should point out that 8% of the students' grammar errors were induced by the MT output.

As to the lexical errors, these seemed to be easier for the students to spot and only a 4% of them were induced by the MT output.

We should highlight though that the students' output is prone to have a wider range of errors depending on the number of students. For a clearer overview of the total number of errors both overall and in each category see table 2 below:

	MT Output	Student-edited Output
Grammar errors	19	69
Lexical errors	12	38
Spelling errors	3	10

Table 2: Number of errors in MT output vs. student-edited output

In the second experiment we compared the MT output with the output provided by two different groups working with the same source text: the control group was asked to translate a source text English into Spanish, whereas the experimental group was asked to post-edit the MT output into more correct Spanish.

Table 3 below illustrates the comparison of error percentages found in the MT output, as well as in the translation and the MT post-editing groups.
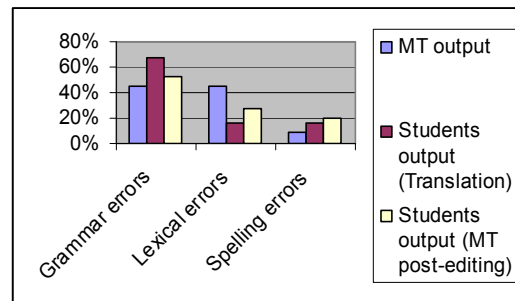

Table 3: MT errors vs. translation and MT post-editing errors

On average, the experimental group (MT post-editing) reported a lower total error percentage than the control group. The

---

[1] "Student output" means student-edited output

graphic suggests that spotting and correcting errors in the experimental group has focused students' attention on grammar and, consequently, this group presents a lower percentage of grammar errors than the translation group. On the other hand, the translation group made more use of the dictionary in their task and, consequently, reported a slightly lower percentage of spelling errors.

Among the most common grammar errors in the MT as well as in the students output we found the following subcategories: confusion of *ser/estar*, misuse of the definite article, misuse of the possessive pronouns, problems with the relative pronouns and agreement errors. In the students' output we found the use of the gerund instead of the present tense in Spanish, and in the translation group we found confusion *por / para* and many gender errors.

From the students' grammar errors only 7% were induced by the MT output and only 8% of the students' lexical errors were induced by the MT output.

Table 4 below summarizes the total number of errors in this second experiment:

	MT Output	Student-edited output (Translation)	Student-edited output (MT post-editing)
Grammar errors	11	38	21
Lexical errors	11	9	11
Spelling errors	2	9	8

Table 4: Number of errors in MT output vs. student-edited output (Translation and MT post-editing)

We acknowledge that command of grammar seems to be more difficult to achieve than command of vocabulary and we recommend MT post-editing as an activity to promote accuracy and therefore grammatical, lexical and stylistic correctness in the foreign language classroom.

As to the conclusions we can draw from this preliminary research, we can safely say: Firstly, that grammatical errors (followed by lexical and spelling errors) are the most predominant ones in the MT output as well as in the student-edited MT output, and, secondly, that students making use of MT post-editing reported to have a lower total error percentage than the ones practising translation on the same source text. These results, therefore, may serve to support our argument that MT post-editing into the target language can be suggested for the foreign language classroom as an activity to practice revision as part of the writing and the translation experience in the language being learnt.

6 Anticipated results

As a result of this investigation we intend to use a part-of-speech tagger to compile a learner corpus, i.e. the MT post-editing results from different text types from a group of advanced students of Spanish. This will help us identify the types of errors made by advanced language learners in MT post-editing and compare them with another tagged corpus of manually-produced student translations. The results, hopefully, will unveil a lot of information on the difference of structures produced by the students and in how far are these from being correct. This information will be

very useful to translation teachers and language teachers interested in teaching MT post-editing as a challenging skill to practice translation and writing skills.

7 Potential impact of the research

One of the potential impacts of this research is distance learning. The Internet has provided us with a new means for collaborative teaching and learning. More and more we are encouraged to an independent language learning experience where communicating ourselves in one or more target languages will be desirable. The following quotation illustrates my point: "It is more likely that there will be increased communication between native and non-native speakers, whether for rehearsal purposes, or as the only medium in which communication in the target language is likely to occur". Felix (2003:16)

Other advantage of this research is that it is portable in the sense that we can use an MT system as a support to learn different language pairs. It is especially useful with close-related languages such as Spanish and Portuguese or Spanish and Italian, where mastering writing (as compared to speaking) turns into a not-so-simple goal to attain.

7 Conclusion

In this paper we have discussed the use of MT post-editing in foreign language teaching as a means to enhance writing practice and translation into the target language. On the basis of our preliminary research we can conclude that MT post-editing seems a feasible activity to be practised in the foreign language class. Our hypothesis is to use the MT system as a "peer editing" that does not provide us with entirely reliable output, but rather presents us with a text that needs amendment. Through these amendments, we intend to promote learner awareness of the need for auto-correction, to promote the students' autonomy in the process of learning a language and to encourage the learner to produce a coherent target language text. This is just a first step in our purpose of investigating whether MT post-editing can enhance language production.

References

Anderson, D.D. (1995) "Machine Translation as a Tool in Second Language Learning" pp.68-97 in *CALICO Journal*, Volume 13, Number 1

Ball, R. (1989) "Computer assisted translation in the modern languages curriculum". *The CTISS File*. 1989 Apr. (8) pp.52-55

Belam, J. (2003) "Buying up to falling down": a deductive approach to teaching post-editing", *MT Summit IX Workshop on Teaching Translation Technologies and Tools (T⁴) (Third Workshop on Teaching Machine Translation)*, New Orleans. Available at www.dlsi.ua.es/t4/docum/belam.pdf

Cook, V. (2001) *Second Language and Language Teaching*. London: Arnold

Corness, P. (1985), "The ALPS computer-assisted translation system in an academic environment" pp.118-127 in Picken C. (Ed.) *Translation and the Computer 7*, London: Aslib

Felix, U. (2003), "An orchestrated vision of language learning online", in Felix U.

(Ed) *Language Learning Online: Towards Best Practice*. Swets and Zeitlinger: Lisse, The Netherlands.

French, J.R. (1991), "Machine translation*"*, in Brierley W. and Kemble I. R. (Eds) *Computers as a Tool in Language Teaching*, pp. 55-69. Chichester, West Sussex: Ellis Horwood Limited

Hutchins, J. and Somers H. (1992) *Introduction to Machine Translation* London: Academic Press

Hyland, K. (2002) *Teaching and Researching Writing* London: Pearson Education Limited

Koenigs (1990), "Übersetzen und Schreiben in der Fremdsprache: Psycholinguistische Beziehungen und Didaktische Kinsequenzen" pp.278-288 in Arnzt R. and Thome G. (Eds) *Übersetzzungswissenschaft: Ergebnisse und Perspektive*. Narr: Tübingen

Krashen, S. D. (1985) *The Input Hypothesis: Issues and Implications*. London: Longman

Krings H. P. (2001), *Repairing texts: Empirical Investigations of Ma chine Translation and Post-Editing Processes*. Kent, Ohio: The Kent State University Press

Lewis, D. (1997), "Machine Translation in a Modern Languages Curriculum" in *Computer Assisted Language Learning* Volume 10, Number 1, February 1997

Löffler-Laurian, A. M. (1986), "Post-édition rapide et post-édition conventionelle: deux modalités d'une activité spécifique" in *Multilingua* 5 (2): pp. 81-88

Mossop, B (2001), *Revising and Editing for Translators*. Manchester: St. Jerome Publishing

Richmond, I. M. (1994), "Doing it backwards: Using translation software to teach target-language grammaticality" pp.65-75 in *Computer Assisted Language Learning*, Vol. 7, N. 1

Shei, C. C. (2002), *Combining Translation into the Second Language and Second Language Learning: An Integrated Computational Approach*. PhD Thesis. Division of Informatics. University of Edimburgh [1]

Shei, C. C. (2002) "Teaching MT Through Pre-editing: Three Case Studies. 6th EAMT Workshop. Teaching Machine Translation. UMIST Manchester [2]

Somers, H. (2001), "Three perspectives on MT in the classroom" pp.25-29 in *MT Summit VIII Workshop on Teaching Machine Translation.* Santiago de Compostela, Spain.

Vasconcellos, M. (1987), "Post-editing on-screen: machine translation from Spanish into English" pp. 133-146 in Picken C. (Ed.) *Translating and the computer 8: A profession on the move* London

Wagner, E. (1987), "Post-editing: practical considerations" pp.71-78 in Picken C. (Ed.) *The business of translating and interpreting: ITI conference I*. London: Aslib.

Acknowledgements