

Teaching Post-editing: A Proposal for Course Content

Sharon O'Brien
SALIS
Dublin City University
Glasnevin, Dublin 9
Ireland
sharon.obrien@dcu.ie

Abstract

There is a growing demand for translation. To meet this demand, many translation companies are introducing a hybrid technology solution combining translation memory and machine translation. However, few trainee translators receive training in machine translation post-editing. This paper asks the question: Why should translator training programmes teach post-editing skills? Is post-editing the same as translation and traditional revision? The skill-sets required of a post-editor are listed and the usual list of skills is extended. An outline for a course in post-editing, divided into theoretical and practical components, is proposed. Finally, the question of when such a course should be given to trainee translators is addressed.

1. Why teach post-editing?

1.1 Growing Demand

The global market for translation was valued at around \$13 billion in 2000, and a growth to around \$22.7 billion by the end of 2005 has been predicted.¹ This increasing demand has led to an increase in the use of translation aids, including terminology management tools, translation memory (TM) and machine translation (MT) technology. At a recent conference on multilingual communication, leading translation companies reported that they are now testing and implementing a hybrid TM-MT technology solution to meet the growing demand for translation.²

¹ Allied Business Intelligence (1998).

² At the Society for Automotive Engineers Multilingual Communication TOPTEC Symposium, which took place from October 3-4 2002 in Nashville, three companies presented

Vasconcellos and Léon (1985:122) claim that a full-time, trained post-editor, working on-screen, can produce polished, standard quality output at a rate of between two and three times faster than traditional translation (i.e. 4,000 to 10,000 words per day). These figures suggest that MT and post-editing are viable solutions to meet the growing translation demand. Yet, how many translation professionals have received formal training in post-editing techniques? In the English translation of Krings's study of post-editing, the editor, Geoffrey S. Koby, suggests that "the translator must be trained in post-editing" (Krings and Koby, 2001:12). To the best knowledge of the current author, there are few translator training programmes offering full courses on post-editing at the time of writing.

1.2 Post-editing skills developed gradually

Vasconcellos (1986a:145) maintains that post-editing skills are developed gradually. The level of comfort with post-editing is greatly increased after 100,000 words (1 month of full-time post-editing). Somers (1997:201) also reports that it is recognised by many that post-editing is a skill that needs to be "honed". Companies wishing to implement machine translation technology would therefore benefit if translation graduates were already "com-

solutions combining machine translation and translation memory solutions (Bowne Global Solutions, SDL International and Telelingua Software).

comfortable” with post-editing. Additionally, post-editing skills would give translators an extra boost when it comes to finding employment opportunities.

1.3 Teaching post-editing means translators will embrace MT

Translators who do not have post-editing skills are frequently hostile to machine translation technology. Common arguments against MT include a dislike for correcting repetitive errors that a human translator would never make, a fear of losing language proficiency by working with poor MT output and a dislike of having one’s freedom of expression limited (Wagner, 1985:213). However, translators who embrace post-editing often report that their day-to-day work becomes much more interesting.³ Drawing on her experience of implementing Systran, Ryan (1988) maintained that the more and the earlier the translator was involved with the implementation of machine translation, the faster a usable system can be developed. Ryan correctly points out that “in an age when some systems can already translate over a million words in an hour, the time that it takes to run a translation becomes insignificant; the cost-effectiveness of the MT system must be measured largely by the effectiveness of the post-editing process” (Ryan, *ibid*:131). Involvement of a translator, therefore, improves the chances of success for machine translation. Senez (1998b:293) confirms this effect when she reports that a translator involved with an MT project eventually “no longer feels threatened by the machine, but has learned to reap as much benefit as possible from what the computer gives him”.

1.4 Conclusions

Post-editing skills should be taught because:

- it would help meet the increasing demand for translation and for faster production times;
- post-editing skills are different from translation skills and we cannot assume that a qualified translator will be a successful post-editor (c.f. 2.0 below);
- it would produce graduates who are already “comfortable” with post-editing and who are more ready to be productive in a machine translation environment upon graduation;
- and it could improve the uptake of machine translation technology by improving translators’ perceptions of MT and its capabilities.

2. Who is the Target Audience?

Having established that there are advantages to teaching post-editing skills, consideration should be given to the question “who are the target recipients of this teaching?”. It is logical to assume that trainee translators should be the primary target audience for post-editing training. However, this assumption encompasses an underlying assumption that translation and traditional revision are similar to post-editing and that translators are the best candidates for post-editing. It is interesting to consider whether, firstly, translating and post-editing are in fact similar activities and, secondly, whether translator training transfers the necessary skills to an individual for post-editing?

2.1 The cognitive viewpoint

Krings and Koby’s (2001:360) unique study on post-editing demonstrates that cognitive processes relating to source-text comprehension during translation and post-editing differ. Also, Krings and Koby conclude that traditional translation is a significantly less linear process than post-editing (*ibid*:498). Therefore, there is reasonable evidence to suggest that post-editing differs from translation from a cognitive point of view.

³ Personal opinion expressed by members of the Luxembourg-based European Commission’s Spanish translation department in September 2002.

2.2 The practical viewpoint

Post-editing and translation differ on the practical level. Translation usually involves one source text and the creation of one target text to a level of publishable quality. Post-editing, on the other hand, involves two source texts, i.e. the text authored in the source language and the raw MT output, which a translator uses to help produce a final version. The task requirements also differ. The usual requirement from the translation process is to produce a target text that meets high quality criteria, whereas post-editing requirements can range from gisting to high-quality publication quality.

2.3 Post-editing and traditional editing or revision

McElhaney and Vasconcellos (1986) claim that post-editing is different from traditional revision of translation. For example, unlike traditional revision, the post-editor is more or less assured that no passages have been skipped and there are no spelling errors. They also point out that while misconstructions will be present in MT as well as Human Translation (HT), misconstructions in the former will likely be more local than in the latter (ibid:141). Löffler-Laurian (1985:71) also draws attention to the fact that the types of errors that occur in machine translation are different from those that occur in human translation.

2.4 Translation and post-editing objectives

Translator training focuses on accuracy and equivalence. Where specialised translation is the main focus, the trainee translator is taught to be as accurate as possible where terminology and meaning is concerned and to aim for cultural and textual equivalence. The trainee translator is taught to produce texts suitable for publication. It is for this reason that translator training, in the traditional sense, can act as a hindrance to post-editing where the aims are frequently different. On

the subject of differences between translating and post-editing, Senez (1998a) says: “A translator will always strive to disguise the fact that the text has been translated. In the case of post-editing, it is enough for the text to conform to the basic rules of the target language, even if it closely follows the source text” (pages not numbered).

2.5 What are the similarities?

Where translation and post-editing do not differ is in the requirement for ascertaining the target audience’s needs. Translation training programmes train translators to examine the expectations of the source language audience and to compare these to the expectations of the target language audience and to translate accordingly. Post-editors need to perform this task too.

2.6 Conclusions

We have seen evidence that post-editing is not the same as translation or traditional revision. In fact, some of the demands of post-editing are contrary to the skills and objectives of translators and probably represent one of the reasons why MT implementation has failed in the past. Nevertheless, McElhaney and Vasconcellos (ibid:142) believe that there are strong arguments in favour of training translators as post-editors. They argue that a translator is best able to identify linguistic errors, has a fund of knowledge about the cross-language transfer of concepts, and has the technical resources at their disposal to work efficiently.

The conclusion, then, is that translators should be trained as post-editors. However, this type of training should be optional rather than compulsory. Qualification for a module on post-editing should be made dependent on the strengths and personality of each student, if possible. The skill-sets required are outlined in section 3 below.

3. What skill-set does a post-editor need?

According to Johnson and Whitelock (1987), post-editing is a highly skilled task where the post-editor ought to be an expert in the subject area, the target language, the text-type and contrastive knowledge: “In effect, the post-editor should be at least as skilled in all of these domains as the original translator” (ibid:140).⁴

Wagner (1987:76) lists excellent knowledge of the source language, perfect command of the target language, specialised subject knowledge, word-processing experience and tolerance as the essential skills of a post-editor.

Vasconcellos (1986a:136-138) elaborates on the need for word-processing skills, listing full key proficiency, efficiency in cursor positioning, effective use of search and replace functions and ability to use macros as essential for the skill set of a post-editor. Knowledge of terminology coding for machine translation is mentioned later in the same article (ibid:142). In Vasconcellos (1986b), the entire paper is dedicated to the significance of text linguistic knowledge for effective post-editing.

In addition to these tangible skills, several authors and MT practitioners list a positive predisposition towards MT as an essential quality for a post-editor (Vasconcellos and Léon, 1985:135; Somers, 1997:201; Wagner 1987:73). Wagner (ibid) reports that translators who are forced to post-edit will not be as efficient as those who have volunteered. She also suggests that “a certain amount of confidence in one’s own translation ability and technical expertise is essential for this type of work” (ibid:204).

There are few differences between the skills mentioned above and those

demanding of a professional translator. However, ability to use macros, to code dictionaries for MT, and a positive attitude towards MT are three attributes required of a post-editor that are not usually demanded of a translator.

This author would argue that several other skills are required for successful post-editing and these are addressed in the following sections.

3.1 Knowledge of MT

Knowledge of MT technology in general would go a long way towards helping the post-editor understand what is going on in the so-called “black-box” and why certain errors occur consistently. Understanding the history of MT development, its current status and future prospects would ensure that the post-editor had an appreciation for the technology, its limitations and how it might improve in the future.

3.2 Terminology Management Skills

While most trainee translators are taught the theory and practicalities of terminology management, the trainee post-editor would benefit from an extensive course in machine translation dictionary coding and term base management. In any one translation environment, multiple tools can be used to store and retrieve terminology both for source and target text production. This presents challenges when terms have to be used across multiple tools and processes. Trainee post-editors not only need to know how to code MT dictionaries, but they also need to know how to manage term bases. This requires knowledge of multiple term management tools and terminology exchange formats, which are only emerging at this time (see, for example, the OLIF, TBX, SALT and XLT initiatives).⁵

⁴ By “contrastive knowledge”, the authors refer to both SL and TL components that map between texts and deep representations of “interface structures” (Johnson and Whitelock, 1987:137).

⁵ For information on OLIF see <http://www.olif.net/>; on TBX and SALT see: <http://www.opentag.com/tbx.htm>; on XLT see <http://www.tt.org/oscar/xlt/dxlt.html>. (All websites last checked on October 12, 2002).

3.3 Pre-editing/Controlled Language skills

It has been documented on numerous occasions that authoring source text using controlled language rules improves MT output.⁶ A drawback to this approach is that authors are unwilling to be constrained by controlled language rules. An alternative solution is to use an intermediate editor who has the necessary skills to apply CL rules to a text before it is submitted to MT. Being an expert in both source language and target language makes the post-editor a good candidate for this job. There is also a significant incentive, i.e. it reduces the time spent on cleaning up tedious and non-sensical errors in *multiple* target language versions! Therefore, knowledge of controlled languages and controlled authoring tools would benefit post-editors.

3.4 Programming skills

Vasconcellos (1986a:136) mentions using macros as a necessary skill for post-editors. In the current author's opinion, a post-editor is an ideal candidate for writing macros to automatically clean-up texts since s/he has extensive experience of commonly occurring errors. These macros are the first step towards the concept of an automatic post-editing tool, as suggested by Ryan (1988), Knight and Chander (1994), Allen and Hogan (2000). If equipped with programming skills, the post-editor could develop his or her own programme for automatically correcting consistent errors for specific language pairs, text types and MT systems.

3.5 Text linguistic skills

As mentioned above, Vasconcellos (1986b) outlines the importance of knowledge of theme and rheme and other language-specific text type norms for post-editing. A good grounding in text linguistics would therefore seem to be of

benefit to post-editors. This knowledge could be applied not only for post-editing but also for programming macros and automatic post-editing modules.

4 Proposed outline for a course module in post-editing

We have so far established the need for teaching post-editing, the requirement that post-editing be taught to translation students and the core skills required. The additional skills of knowledge of MT systems, terminology management, controlled language, programming and text linguistic skills have been added to this list. In this section, a course outline that addresses these skill-sets will be proposed.

Since we have established that the best candidate for this type of training is a translation student, we will assume that the student has acquired certain skills before taking the module in post-editing. If this is not the case, then these skills would have to be taught prior to the post-editing module: specialised translation skills; basic linguistics; basic terminology management; IT skills; an introduction to language technology (focussing on translation memory tools). An assumption is also made that the student has excellent source and target language skills.

The module in post-editing is divided into two, with a focus on theoretical issues in the first half and a focus on practical issues in the second half.

4.1 Theoretical Component

The theoretical content will contain the following subjects:

- Introduction to Post-editing
- Introduction to Machine Translation Technology
- Introduction to Controlled Language Authoring
- Advanced Terminology Management
- Advanced Text Linguistics
- Basic Programming Skills

⁶ See Adriaens et al. 1996, Mitamura et al. 1998, Adriaens et al. 2000.

The *Introduction to Post-editing* would address the concept of post-editing. For example, why do we need post-editing, how does it differ from translation and revision, what levels of post-editing exist, how do we determine user requirements, what technology can we use for post-editing, can we classify typical post-editing errors, and so on?

The *Introduction to Machine Translation Technology* should cover the history of MT, MT system types, description of commercial MT systems, evaluation methodologies, current state of the art, including integration with translation memory tools, and future prospects.

The *Introduction to Controlled Language Authoring* should include a history of Controlled Languages, a description of CL tools, evaluation methodologies for CL tools, current state of the art, integration with authoring and MT tools, and future prospects.

Advanced Terminology Management would aim to build on the basic terminology management skills the student brings to the course, by discussing the strengths and weaknesses of terminology management tools, dictionary coding for MT, and, most importantly, terminology exchange between tools using terminology exchange standards such as XLT, OLIF, and TBX.

Advanced Text Linguistics would build on the basic linguistic skills of the student by introducing them to the standards of textuality, text type classification, and the use of corpus linguistics and corpus analysis tools for analysing text types.

Basic Programming Skills would introduce the student to the basics of programming and would then instruct the student in macro programming and in a programming language suitable for Natural Language Processing, for example *Perl*.

4.2 Practical Component

Since post-editing is a practical skill and one of the objectives of teaching post-

editing is to allow a student to acquire the “comfort” factor Vasconcellos talks about *before* being recruited, practical experience of post-editing would form a major component in this module. According to Vasconcellos (1986a:145), a post-editor at PAHO (the Pan-American Health Organisation) post-edits 100, 000 words, or almost one full working month, before that level of comfort is reached. While it may not be possible for a student to attain this goal, especially considering the workload from the theoretical component of a programme, the student should be encouraged to practise post-editing both within and outside course hours. Post-editing of different text types from different MT systems should be carried out. If the student has more than one target language, post-editing into multiple target languages would be desirable. Also, since post-editing requirements sometimes vary between producing a text for information purposes and for publication purposes, students would practise these different “levels” of post-editing.

Practical experience with at least two commercially-available MT systems would also form part of the practical component. Students would be required to submit texts for translation to the MT system and to analyse and compare the results when system settings have been changed, user-specific terminology has been coded, and, where possible, linguistic rules have been altered. Students would also be required to investigate the pros and cons of a MT system’s integration with a translation memory tool.

To gain practical experience of terminology management tools, dictionary coding utilities, and terminology exchange formats, students could be instructed to create a term data base using a specific terminology management tool (let’s call this term management tool A), to code that terminology using a machine translation dictionary coding tool and then to export the terminology from the MT dictionary to a second term management tool (called

term management tool B) using different terminology exchange standards.

Practical experience of controlled authoring tools could be gained in the following manner: Students are asked to check and edit a text in the source language using a CL tool and to submit the controlled and uncontrolled texts to a number of MT systems. Post-editing of both versions would then reveal the pros and cons of controlled authoring for machine translation.

Corpus Analysis practical experience could be gained by compiling parallel corpora, tagging them, and analysing them for specific text linguistic features such as theme/rheme structure, voice, tense, cohesive ties, etc. using corpus analysis software such as Wordsmith tools.

Finally, students would acquire practical programming skills by writing macros to automatically apply common changes in target texts. They could also apply the programming language skills learned in the theoretical component of the course by designing a rudimentary automatic post-editing application.

When designing a course in an academic environment, it is usual to specify how many hours are dedicated to each component of the course, whether a component is considered core or optional, what weighting a component has in comparison to other components and how each component will be assessed. Since every academic institution differs, there is little point in specifying this information here. Instead, a general suggestion is made that equal status be given to the theoretical and practical components and that assessment be carried out on a continuous basis, using practical and written methods of assessment.

The reader will most likely acknowledge that the proposed course outline is quite extensive and that, unless this module was offered as a stand-alone programme, such as a graduate certificate or diploma, it would be difficult to cover all components adequately in the time-

frame of an academic semester. The proposed outline is an ideal, from the author's point of view. The module could, of course, be split over two semesters.

5 At what stage should post-editing be introduced?

Since translator training programmes are structured differently in every institution that offers this type of programme, it is impossible to say exactly where a course on post-editing fits in. However, as has already been mentioned, successful post-editing requires a high level of confidence in the post-editor's own work (Wagner, 1987:204). In the introduction to his book on traditional revising and editing, Mossop (2001) reports his finding that undergraduate students are rarely ready for self-revision or revision of others' work until after they have completed a practicum. This, too, suggests that experience and confidence are necessary ingredients for the task of post-editing.

In the proposal outline above, it was assumed that students would take a post-editing course only if they had some pre-requisite skills, i.e. excellent language skills, specialised translation skills, basic linguistics, basic terminology management, IT skills and an introduction to language technology. Since experience, confidence in one's own work, and a number of pre-requisite skills are required before taking a course in post-editing, this suggests that such a course should only be offered in the last part of an undergraduate translator training programme, or, even more ideally, in a post-graduate programme because the students enrolled in the latter type of programme are more likely to be experienced and to have more confidence in their own work.

6 Summary

Current industry trends seem to suggest that machine translation will form an increasing part of the technology solutions put in place to meet the growing demand for translation. If this turns out to be true,

an growing number of translators will have to deal with machine translation output. Let the educators of translators prepare future generations for this by teaching students about machine translation and post-editing. This paper outlines the skill-sets required and proposes course content and structure.

References

- Adriaens, Geert, Jeffrey Allen, Arendse Bernth, Kurt Godden, Teruko Mitamura, Eric Nyberg, Rick Wocjik, Rémi Zajac (2000), *Third International Workshop on Controlled Language Applications, CLAW 2000*, Seattle, Washington.
- Adriaens, Geert, Roger Havenith, Rick Wocjik and Bruno Tersago (1996), *First International Workshop on Controlled Language Applications, CLAW 96*, Centre for Computational Linguistics, Leuven, Belgium.
- Allen, Jeffrey and Christopher Hogan (2000), "Toward the Development of a Post editing Module for Raw Machine Translation Output: A Controlled Language Perspective", in Adriaens et al. (2000), 62-71.
- Allied Business Intelligence, Inc. (1998), *Language Translation: World Market Overview, Current Developments and Competitive Assessment*, Oyster Bay, New York.
- Johnson, Roderick L. and Peter Whitelock (1987), "Machine Translation as an Expert Task", in S. Nirenburg (ed) *Machine Translation: Theoretical and Methodological Issues*, Cambridge University press, Cambridge, 136-144.
- Knight, Kevin and Ishwar Chander (1994), *Automated Post-editing of Documents, 12th National Conference on Artificial Intelligence*, Seattle, Washington, 779-784.
- Krings, Hans P. and Geoffrey S. Koby (eds) (2001), *Repairing Texts: Empirical Investigations of Machine-Translation Post-Editing Processes*, Kent State University Press, Kent, Ohio.
- Löffler-Laurian, Anne-Marie (1985), *Traduction automatique et style, Babel* 31, 70-76.
- McElhaney, Terrance and Muriel Vasconcellos (1988), "The Translator and the Postediting Experience", in Vasconcellos (1988), 140-148.
- Mitamura, Teruko, Eric Nyberg, Geert Adriaens, Linda Schmandt, Rick Wocjik, Rémi Zajac (1998), *Second International Workshop on Controlled Language Applications, CLAW 98*, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania.
- Mossop, Brian (2001), *Revising and Editing for Translators*, St. Jerome Publishing, Manchester, UK.
- Ryan, Joann P. (1988), *The Role of the Translator in Making an MT System Work: Perspective of a Developer*, in Vasconcellos (1988), 127-132.
- Senez, Dorothy (1998a), "Post-Editing Service for Machine Translation Users at the European Commission", *Translating and the Computer* 20, London.
- Senez, Dorothy (1998b), "The Machine Translation Help Desk and the Post-Editing Service", *Terminologie et Traduction* 1, 289-195.
- Somers, Harold (1997), "A Practical Approach to Using Machine Translation Software: 'Post-editing the Source Text'", *The Translator* 3, 193-212.
- Vasconcellos, Muriel (1986a), "Post-editing On-screen: Machine Translation from Spanish into English", *Proceedings of Translating and the Computer* 8, London.
- Vasconcellos, Muriel, (1986b), "Functional Considerations in the Post-editing of Machine Translation Output: Dealing with V(S)O versus SVO", *Computers and Translation* 1, 21-38.
- Vasconcellos, Muriel (1988) (ed), *Technology as Translation Strategy*, American Translators Association Scholarly Monograph Series, Vol. II, State University of New York at Binghamton (SUNY),
- Vasconcellos, Muriel and Marjorie Léon (1985), "SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization", *Computational Linguistics* 11, 122-136.
- Wagner, Emma (1985), "Post-Editing Systran – A Challenge for Commission Translator", *Terminologie et Traduction* 3.
- Wagner, Emma (1987), "Post-editing – Practical Considerations", in *ITI Conference I: The Business of Translating and Interpreting*, London, 71-78.