Post-editing or no post-editing?
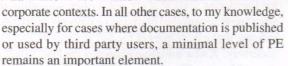
ost-editing (PE) is a task that is primarily associated with Machine Translation (MT) whereby a post-editor edits, modifies and/or corrects pre-translated text that has been processed by an MT system from a source language into one or more target languages. The implementation of MT into new translation and localisation workflow processes is thus constantly raising the question about the acceptance and use of half- or semi-finished texts.

This issue has not necessarily been a concern before within the field of traditional Human Translation (HT) because human translators normally do not produce a partially completed translated text. And once the idea of MT and post-editing is taken on by an organization, it is important to determine to what extent MT output texts are acceptable, and how much human effort is necessary to improve such imperfect texts. This human effort can be determined by the cognitive effort exerted to identify the corrections (especially since PE is a task different from translating or revising), as well as the manual effort to make the corrections on paper and/or on-line. With constant forces in the push for globalisation, SIMship strategies, process streamlining, faster turn-around of information, etc, MT and PE are gaining more importance than ever before.

Yet, some people might think that PE is a one-size-fits-all technique. This is quite far from true as described below. I would like to demonstrate that there are in fact different levels of post-editing and that they serve different purposes. Organisations that consider implementing MT systems into their translation workflow processes should be aware of these PE levels during their discovery and decision-making stages. The first level of type of PE is actually using MT without post-editing. There are two reasons for employing this strategy. One such case is for information translation gisting (also called content browsing) which is currently generating up to 1 million translation requests per month at various major Internet MT portals (Bennett, 2000).

This approach bypasses human intervention by providing an understandable, base-level MT output translation (despite the errors) of a foreign language text to readers in their mother tongue or in a language that they are proficient in. The second typical case for using MT without PE fits within the purposes of creating, publishing and disseminating information. The only domain which up to this point has had very consistent and published results for non post-edited or limited post-edited information is for weather bulletins. The METEO weather bulletin MT system has consistently demonstrated for many years that it is possible to reach 90-95% MT accuracy. Yet, this is understandable since weather bulletins constitute a sublanguage which presents a favourable grammatical structure for MT processing. In general, the notion of 100% MT accuracy without PE was publicised in the 1980s but then dwindled off during the 1990s once developers and implementers were faced with the incredibly complex issues involved in knowledge management, document processing, and authoring / translation / localisation within industrial and corporate contexts. In all other cases, to my knowledge, especially for cases where documentation is published or used by third party users, a minimal level of PE remains an important element.

A second type of PE, referred to as Rapid PE, is destined to provide a strictly minimal amount of corrections on documents that usually contain perishable information (i.e., having a very short life span). Such documents, in essence, are not necessarily intended for public use, nor for wide circulation, but are rather urgent texts intended merely for information purposes or for restricted circulation, such as working papers for internal meetings, meeting minutes, technical reports or annexes, etc. Rapid PE emphasises strict minimal editing of texts in order to remove only the most blatant and significant errors; stylistic issues are not even considered. A good case for this is the European Commission (EC) Translation Service which noticed a striking increase in MT usage by the EC's operating departments at the beginning of the 1990s with regard to dealing with urgent translations that could not be met by traditional translation channels. The EC's post-editing service was thus created as a response by providing rapid translation revisions of MT output.

The EC Translation Service current post-edits approximately 600,000 pages of MT raw output per year. A third type of PE, commonly referred to as "minimal PE", came into play in the 1990s within several industrial and corporate sectors. The main issue for partial PE is how to quantify the amount of PE changes that must be made to a raw MT output text. Minimal PE is a fuzzy, wide-range category because it often depends on how the post-editors define and implement the "minimum" amount of changes to be made in view of the client / reader audience (Allen, in preparation). It is important to keep in mind that when the resulting documents are destined for distribution, whether this be internal or external circulation, the level of interpretation regarding

Jeff Allen
postediting@aol.com

Jeff Allen
postediting@aol.com

"minimum" PE often seems to vary from one post-editor to another, from one manager to another, from one reviser to another. Hence the need for PE guidelines. As for the full, or complete, post-editing of texts, this notion has been debated for many years because full PE implies a high level of quality of the resulting texts. The main issue is whether or not it would be faster to post-edit the raw MT output or simply translate the document from scratch. It has been shown by specific industrial projects that post-editing on documents written according to controlled language principles take less time to PE than translating the entire document without any computer-aided translation assistance.

The use of full PE on uncontrolled input language texts has generally been avoided in the past. However, recent activities by localisation and translation agencies (e.g., ABLE International) that implement MT systems for translating texts that are not written according to any specific controlled grammar or writing guidelines, indicate that a new market for full PE may in fact be underway. A new initiative known as the Post-Editing Special Interest Group was set up by a few members of the Association for MT in the Americas (AMTA).

This group met at the AMTA98 meeting in Philadelphia (USA), then at CLAW2000 in Seattle (USA), and again recently at AMTA2000 in Cuernavaca (Mexico). The primary thrusts of this SIG are to develop specifications for what would be an optimum PE environment, to educate the various audiences which need to know more about PE, and to develop PE courseware for translation programs.

And some of us ask if even PE processing can be automated. The Pan-American Health Organization (PAHO) has been providing special post-editing macros to its post-editing staff for well over a decade. Also, in 1999 the EC Translation Service provided Christopher Hogan and I with training and testing texts from its database in order for us to develop an automated PE (APE) module which is described in Allen & Hogan (2000). This module basically allows for the semi-automatic correction of the most common repetitive errors in MT raw output, thus letting the post-editors to focus on the more critical changes. So, is it better to pursue "Post-editing" or "No Post-editing"?

There simply is not a single answer to this question. Once your organisation has chosen to take the route of MT integration, then it is important to consider issues such as the control placed on the input text, the required translation turn-around time, the life expectancy of the target texts, the customer's needs, etc. The level of post-editing is another one of the factors in the decision-making formula and needs to be based on your own needs and expectations.

References: Allen, Jeffrey. (in preparation). Post-editing. In Computers and Translation: A Handbook for Translators, edited by Harold Somers. Amsterdam: Benjamins. (information available at: http://www.ccl-.umist.ac.uk/staff/harold/C&T.html) Allen, Jeffrey and Christopher Hogan. (2000). Toward the development of a post-editing module for MT raw output: a new productivity tool for processing controlled language. Presented at the Third International Controlled Language Applications Workshop (CLAW2000), Seattle, Washington, 29-30 April 2000. Bennett, Scott. (2000). Taking the Babble out of Babel Fish. In Language International magazine, Vol. 12, No. 3, pp. 20-21.