

TAUS

REPORT

February

2006

Best practices in post-editing

By Andrew Joscelyne

| T | A | U | S | 
Translation Automation User Society

Table of Contents

What is post-editing	1
Types of post-editing	2
How post-editing fits in the translation workflow	3
Throughput rates	4
Examples of use	5
Key issues in post-editing management	6
Automating the post-editing function	9
Future developments	11
Information sources	12

Summary

This interim report collates findings from various user sources on current practice in post-editing in a translation automation environment.

Post-editing is best understood as an integral part of the automated translation/localization process, rather than a separate stage of editing/ revising/quality assurance. It involves linguistic more than subject area skills and is performed best by alert translators, familiar with machine output, working in a standard translation environment. Practical training is required to spot and correct typical machine mistakes as quickly as possible, and ensure that the automation system receives appropriate feedback to upgrade the dictionary or rule base. Publishable (dissemination) quality post-editing can output around 5,000 translated words a day, whereas lighter editing for gisting (assimilation) can at least double this rate.

Since the overall aim of any translation automation solution is to reduce costs and accelerate throughput at consistent quality levels, future work on post-editing will seek to optimize the post-editing task by both improving input quality to the translation process, and also using emerging tools to automatically correct egregious machine output errors before the actual post-editing begins.

FOUNDING MEMBERS**USERS***IT*

- Autodesk
- EMC Software Group
- FileNET
- Hewlett-Packard ACG
- Intel
- McAfee
- Oracle
- Sun Microsystems
- Symantec
- PTC
- UGS

Telecom

- Avaya
- Cisco
- Lucent

Medical

- Gambro BCT
- MAQUET Critical Care
- Molina
- Philips
- Siemens
- Spacelabs Medical

Food

- McDonald's Corporation

Patents

- Zacco A/S

Institutions

- European Patent Office
- International Monetary Fund
- SWIFT

PRACTITIONERS

- CLS Communications
- Delta International
- Eurotexte
- GrafiData
- Lionbridge
- Logrus
- Merrill Brink
- SDL International
- TOIN
- Transco
- Vistatec
- WH&P
- Yamagata Europe

PRODUCTS

- acrollynx
- AuthorIT
- CCID Transtech
- Cross Language
- DocZone
- Idiom Technologies
- Language Weaver
- Meaningful Machines
- Multicorpora

Best practices in post-editing

by Andrew Joscelyne

What is post-editing

“Post-editing: examination and correction of the text resulting from an automatic or semi-automatic machine system (machine translation, translation memory) to ensure it complies with the natural laws of grammar, punctuation, spelling and meaning.”
Draft of European Standard for Translation Services (Brussels, 2004).

The product of post-editing will be either a publishable document (‘full’ post-editing) comparable to high quality human translation, or an understandable document (‘light’ post-editing), containing correct terminology and names, expressed in unambiguous but not necessarily elegant sentences. The overall aim of the post-editing process is to improve a machine-generated translation with a minimum of extra work.

Post-editing vs. quality assurance

Post-editing should not be considered as or a type of translation quality assurance (QA). As suggested, it is seen as a specific translation practice and therefore part of the normal language-centric translation review process. In a typical localization/translation workflow, QA is a separate step, that will come after the post-editing of TM +/- MT output.

Post-editing vs. pre-editing

Post-editing is so named to contrast with pre-editing - i.e. modifying the input text before translation automation to facilitate machine processing, improve raw output quality, and therefore reduce the post-editing workload, especially if one document is to be machine translated into several languages.

Pre-editing can cover either a broad or a fairly narrow range of preparatory activities. In its very widest sense it can merge with controlled language input, whereby the source text is authored according to fixed rules to make it easier to translate automatically. We can call this linguistic pre-editing. For example, using relatively short sentences, avoiding certain complex or ambiguous syntactic structures, ensuring that the same term is used consistently, without synonyms. But there is considerable evidence that the cost (cognitive, time, and resources) of creating a cost-effective controlled authoring environment, training the authors, and ensuring compliance, is still too high (and the process poorly understood) for most organizations involved in handling large-scale multilingual content flows.

In its narrower sense, pre-editing often means cleaning up the source text by spell-checking, format checking, tagging texts such as addresses and proper names to prevent translation. Tools have been developed (e.g. by Terence Hall working with Siemens Netherlands) to facilitate pre-editing, for example by automatically breaking each sentence into brief “natural” phrases which can then be manually translated where necessary, and stored in a repository of multiword expressions. This “human

contribution” can improve translation accuracy by the machine and therefore reduce the post-editing load later.

This suggests there is an obvious trade-off between pre and post-editing. If a single document is to be translated into more than three languages, then it is more cost-effective to invest time in extensive linguistic pre-editing to ensure the best possible MT output, and hence lesser post-editing expenditure in each target language.

Both pre- and post-editing should be viewed as ‘standard’ rather than ‘extra’ editing tasks in a multilingual document production context. What distinguishes post-editing particularly from the more mainstream “editing” tasks of copy-editing, proof-reading, fact-checking or revising is its specifically “linguistic” content. Successful post-editing requires competence in both source and target language. It is this linguistic dimension that will determine the kind of post-editing processes that can be set up to deliver cost-effective, fast and high quality translations.

Types of post-editing

There is some consensus around the idea of two degrees of post-editing: light and full.

Light post-editing

This involves taking the raw MT output and performing as few modifications as possible on the text to make it an accurate reflection of the source text content. Light (sometimes called rapid) post-editing tasks may include:

- replacing unknown words,
- deleting superfluous translation alternatives generated by the machine,
- repairing machine-induced meaning distortion (possibly the chief priority),
- correcting the most salient word and grammatical errors,
- partially or wholly rewriting some sentences.

The focus is on making content acceptably understandable, but ignoring stylistic niceties.

Light post-editing is best applied as gisting support for any translated document from a specified knowledge domain that is needed as a decision aid. Typically these will be short, urgent texts such as emails, financial and health alerts, or outputs from a knowledge-base, to be used for immediate consumption online. For example, the European Commission translation service supplied light post-edited internal documents for meetings or staff requiring access to information. It is suggested that a “rapid post-edited” disclaimer is added to any such translations. Light post-editing is based on the principle that the terminology for the knowledge domain in question is already covered in the MT dictionaries, and that almost no in-depth term checking is required.

One vendor of post-edited MT documents containing rapid financial information calculated that the whole process (including any pre-editing, translation, and other steps) should take about 15 minutes a page, once the process is fully in place. One drawback of rapid post-editing is that it may be hard to capture post-editor

modifications and feed them back into the document/ information lifecycle to upgrade the translation memory and/or MT rule base.

Full post-editing

For ‘publishing quality’ automated translations, extensive post-editing will usually be required. This involves all standard revision practices (checking terminology, proper names) as well as making syntactic/grammatical modifications to adapt the unusual linguistic nature of some MT output to standard stylistic practice in the target language.

Translators who carry out full post-editing often complain that they could have done the job better from scratch. However, the value-add of full post-editing is that the corrections to a machine text are (ideally) recycled into the system as ‘linguistic’ upgrades, thereby improving the system over time and increasing translation automation ROI.

How post-editing fits in the translation workflow

Post-editors usually receive their input files directly from the MT server and then pass them on, either to a quality reviewer or to the next stage in the dissemination process. In time, the post-editing position will become naturally integrated into the ‘translation review’ function, whatever the ‘source’ of the translation (MT, TM + MT, human + MT, statistical MT, rule-based MT).

Post-editors are therefore more like translators working in a bilingual TM environment than fully-fledged reviewers, working in a monolingual ‘copy-editing’ environment. In due course, the development of more automated pre- and post-editing tools will reduce the time spent on post-editing, but in the foreseeable future, machine generated text will still need the services of multi-skilled post-editors to deliver the quality that customers pay for.

In most cases, post-editors work on hybrid TM + MT documents, in which a certain percentage of the TM-translated version will already be close to the desired (i.e. publishable) quality.

Provided the requisite terminology has been integrated into the MT dictionaries and selected for the project in question, the TM and MT dictionary databases will contain identical vocabulary, and the relevant terminology can be considered accurate for any output in that project.

Post-editors should be provided with a feedback function that ensures that the edited texts are entered into the TM database, and informs MT system developers of any systematic structural errors that can be corrected in the rule system. As mentioned above, it may prove hard to feed back these errors if light post-editing is being used in a rapid translation workflow context.

Throughput rates

Post-editing is essentially a rapid quality booster. Like the translation process itself, the main aim is to do as much post-editing as quickly and therefore as cheaply as possible to deliver the desired quality level.

Obviously speed depends on the quality of raw TM/MT output, and the scope of the post-editing instructions. The table shows throughput rates that suggest plausible practical benchmarks for light and full post-editing:

Organization	Activity	Quality level	Post-editing rate (pages/ day)
Capstan (BE)	Vendor, working for EC	Understandable	30
CLS (CH)	Vendor, working for a bank (no longer using PE)	Understandable	30
Christian Boitet (FR)	Academic experiment	Understandable	30
Symantec (USA)	User, alerts	Understandable	
Jeff Allen (FR)	Experiment	Understandable ++	24
SDL (UK)	Vendor	Publishable	20
DELTA (DE)	Vendor	Publishable	20
PAHO (USA)	User organization	Publishable	14-16
LSI (USA)	Vendor advertising a service	Publishable	4-5

As well as being job and text-specific, these rates for translation automation plus post-editing translate very generally into the following ball-park figures for savings on:

Time: 25% reduction in time compared to human translation” (Christian Boitet). This does not include TM leverage, so total time reduction can go as high as 40% and more.

Cost: MT and post-editing reduces the cost of a human translation by 30% (PAHO).

Examples of use

The standard metric is that an MT solution delivers productivity in a world-class workflow context for projects with over 500,000 translated words a year into a target language set.

Organization	Documents	Practitioners	Priorities / results
Symantec (user)	Technical support documents and virus write-ups	In-house translators using Trados 6.5 and Systran 5.0 to post-edit that content.	Information accuracy
Pan-American Health Organization (user)	Manuals Reports Proposals Summaries Scientific articles	Outsourced to experienced professional translators	Grammaticality to commercial quality output levels Publishable translations Post-editing not costed separately Cut cost of translation. MT post-editing rates = c. 70% of corresponding human rate
Siemens Nederland (user)	Tech doc, office docs	Developer pre-edits and post-edits	Charges 40% of price of human translation
European Commission Translation service (user / vendor)	Internal reports, memos, minutes, annexes to contracts, proposals, etc	Freelancers hired via tenders Translation Service also offers rapid post-editing for Commission Administrators	Rapid post-editing of internal documents only Avoid meaning distortion Users do QA As few changes as possible from raw output Grammatically correct, faithful reproduction of original message
DELTA (vendor)	Service manuals	Post-editing service. Post-editing done in TagEditor, IBM Translation Manager, etc	Full post-editing to human quality Costed separately Post-editing skills more important than automating "text repair tools"

Key issues in post-editing management

Selecting post-editors

Translators are to be preferred over monolingual target language editors since they know the source language. Since post-editing continues to be a linguistic rather than a purely proof-reading/ copy-editing type of activity, bilinguals will prevent errors in apparently "accurate" MT output. This is in line with the general understanding of post-editing as a 'translation' component in the workflow.

However, many practitioners observe that translators tend to worry too much about "style" when post-editing, thereby increasing the time and cost of post-editing. They therefore need to be trained to focus on "communicative" rather than purely "linguistic" accuracy or relevance. In addition, they should be motivated to maintain and improve the MT system as part of their responsibilities.

A distinction is sometimes made between "translators" and "linguists" (professional language experts who know about linguistic structure). Some "light" users consider

linguists to be more useful for post-editing since, again they will worry less about stylistic quality, and where possible will tend to give more structured, easier-to-encode feedback to the MT development team. Once a user can accurately estimate translation loads, it should be possible to build effective post-editing teams and train them properly.

A final distinction can be made between general translators and subject area specialists/translators, who will deliver optimum post-editing quality in their field, and even act as QA reviewers in the more conventional sense of the term.

Training on the job

There is consensus about the need for training, and above all that it should be practical and conducted in-house on the job in the context of the specific translation workflow. A very few university courses for translators attempt to give students a taste of machine output and some post-editing practice. But since most publishable automated translation will be based on a combination of MT and TM, training is best carried out under real-world constraints.

Emphasis should be put on:

- understanding the machine “mindset” (the typical linguistic behavior of the MT system)
- identifying and typologizing systematic, recurring machine errors
- ramping up progressively to full productivity (rather than expecting it immediately)
- providing the most familiar productivity environment (e.g. using Trados TagEditor or any other standard editor, where possible).

Clear post-editing guidelines

Most organizations that practice post-editing have produced a set of guidelines to aid post-editors focus on priorities in the process. Although they differ in details, the guidelines collected from Symantec, PAHO, SDL, Lionbridge, and SAP share the following points:

1. For light post-editing

Style is not a primary consideration; but information accuracy and basic readability/‘scannability’ of the text is.

All the words will probably present in the MT output, but possibly in the wrong order.

Therefore:

- Make sure that all information is transferred accurately.
- Modify what is grammatically deviant from an output of commercial quality.
- Modify what is lexically essential for the understanding of the target text (wrong or nonsensical words and phrases).
- Do not use synonyms for the sake of originality.
- Minimize the search for improvements if they are not obvious immediately.

2. For fully publishable quality

In the case of complex documents with graphs, figures, tables etc, there may be a need for extensive guidelines to deliver high quality that meets corporate publishing standards.

- Look at source in its original format.

- Identify terminology sources (references).
- Check target for completion.
- Identify and correct words and expressions left in source language.
- Identify and correct format code problems (e.g. figures in tables, numbering systems, references, citations, indications of where graphics are on page may change).
- Identify and correct use of capital letters.

The post-editing work environment

The guidelines specify that the post-editing environment should be as similar to the translator standard work space as possible, to ensure quicker, easier productivity levels. Adding too many tools will slow down the learning process and create additional training time.

The only vital tool is a workflow that sends various types of raw or edited content back into the automation process (TM segments, MT dictionary, MT rule base). For example, in the Lionbridge environment, by pressing a key combination, problem translations are recorded to the MT system, together with the post-edited versions, classifying them to specific error categories.

Practice seems to focus around the Trados TM environment, especially TagEditor. It is suggested that a secondary TM is set up to avoid any corruption of the main TM database, until the segments to be later used by the TM are considered to be up to human quality standard.

Interface and productivity tools

Typically, post-editors should use side by side screens, one with the raw output and one with the edited version. There should also be a feedback column for post-editors to send information back to the system.

In the PAHO post-editing environment, post-editors have access to statistics on the translation process:

- % of sentences completely OK, partially analyzed, not analyzed, and not translated.
- number of sentences over 70 words.
- number of not found words (could be spelling problem in the source text).

The purpose of these statistics is to roughly size the post-editing job ahead. Also, if there are less than 50% completely analyzed source sentences, this indicates that there are major grammar problems in the source, and the job should be resubmitted, possibly with some pre-editing.

PAHO also provides a dedicated post-editing tool bar for typical Word, PowerPoint, and Excel documents. These macros provide easy access to such functionality as:

- A draft View (to check readability)
- Page Size
- Search and Replace
- Browse Dictionaries

- Switch Right and Left
- Delete
- Lower / Upper case change
- Remove highlight

PAHO also has language specific productivity tools. For example, English post-editors have access to the following macros among others:

- Delete next *the*
- Change *Its* to *their*
- Delete and switch (book about science => science book)
- Serial comma X, Y and Z => X, Y, and Z.

In other languages, there may be a need for systematically switching plurals and singular forms, or adding articles, or reversing nouns and adjectives.

Automating the post-editing function

Organizations committed to the translation automation agenda are beginning to realize that there may be innovative ways to speed up the post-editing process by

- a) automatically improving input (pre-editing), as is well known, but also
- b) developing tools that can rapidly perform automatic edits on the MT/TM output.

Wherever a purely formal edit procedure can be modeled as an algorithm and run through the output, this will allow human post-editors to concentrate on the more appropriate task of repairing meaning failures.

Symantec is exploring the use of pattern-matching tools to catch frequent mistakes and speed up the post-editing process. This is basically a post-processing stage that extends the capabilities of certain tools such as Systran's normalization dictionaries. During this automatic post-processing step, the raw MT output is not tokenized or tagged, so there is no linguistic analysis of the MT output. Rather pattern matching tools over "regular expressions" can handle such features as:

- Capitalization
- Incorrect spellings
- Missing contractions
- Extra words (fichier de .bmp vs. fichier .bmp)
- Word order
- Formatting: trailing spaces after symbols (backslashes)
- Punctuation inconsistent with style guide: e.g. inverted commas for German

Lingtech (a Danish vendor providing automatic translation services for patents in Denmark) has developed a tool called APE to handle 'failsofted' output (i.e. when the machine forces out a sentence without being able to parse it) that delivers seriously garbled word orders. APE indexes the source text word order and uses this as a knowledge source, and after the MT has been output, runs through the text identifying failsofted sentences and reorders them in line with the source content word order.

The user community will naturally need more information on whether these tools work, and how they and any other tools can be integrated in a broader tool platform that might produce relevant support for post-editors in any natural translation automation setting.

Can post-editing benefit from MT evaluation techniques?

The ‘technical’ evaluation of MT output quality is developing into a separate discipline, focused exclusively on raw output quality. It basically enables different MT systems to be compared against an identical task, using automatable measurements that deliver a ranking. The best known recent example of a large-scale MT evaluation competition was the 2005 NIST event run by the US government, with Arabic and Chinese as the key test languages.

Underlying most of the available techniques is an N-gram analysis (comparing the co-occurrence of words in a text, not just the statistics of single words) that compares the raw output quality against a reference translation, previously made by humans.

These statistical techniques include:

- WER: evaluates the percentage of words that need to be inserted, deleted or replaced in the target output to obtain the same sentence as in the reference.
- Bleu: (IBM) which evaluates only the target language to no reference to the source as and appears to needs a lot of reference text to make it work effectively.
- NIST: an upgraded version of Bleu.
- GTM (NYU): counts matches of precision and recall of words from source and target
- Meteor (CMU): calculates N-gram overlaps in source and target. A Reordering penalty module will evaluate how many chunks need to be moved around in target text to get back to reference text.

Can any of these fully automated evaluation techniques be of help to the practical translation automation community to develop better post-editing tools? The algorithms are cheap and quick, although they obviously depend on a set of well-maintained reference translations for their effectiveness. Translation memories maintained in a project database provide one form of reference translation, and could possibly be used in various ways as forming the baseline against which new MT output could be compared. But there may be good reasons to maintain reference sets at a more sophisticated level than as plain text translation memories. This will require considerable research before it reaches the translation business environment.

Transco's MAT

One interesting project called MAT is being undertaken by localization vendor Transco to evaluate how far a given MT system can boost post-editing productivity. Transco reckons that standard evaluation tests of MT are either time-consuming or are only run over deceptively small samples, and so only measure very vaguely how much post-editing effort is required to meet specific quality standards.

MAT therefore uses TM tools to measure the “similarity” of MT output to the human output in the TMs and compare it with the productivity of using TMs. The full method has not yet been worked out or applied, but one of the aims is to find out whether there is a linear relationship between MT/post-editing productivity and TM productivity.

Statistics snapshots

In the case of bulk post-editing tasks, a quick statistical snapshot of the evaluated output quality using one of the MT evaluation techniques might provide a post-editor with a more accurate idea of how long or complex their task be. But this again is an empirical question and needs experimentation and testing. PAHO (see above) already makes basic machine output (but not evaluation) statistics available for its post-editors. And the Meteor method offers a ‘moving chunks’ criterion - basically how much time a post-editor would need to press the keyboard to improve the text.

To return anecdotally to the 2005 NIST evaluation, the competition generated unexpected supporting evidence that post-editing is needed in business-critical translation environments. One of the MT engines competing was Linear B, a statistical system developed by a small vendor in Scotland. Curiously, Linear B was not submitted as a fully automatic MT system for the evaluation of the Arabic to English test. Here is the report:

“Rather, it was a human-aided statistical MT system that used non-Arabic speakers to correct the English fluency by selecting optional English phrases from the system’s lattices. Search engines were used to look up the spelling of proper names.”

Apart from the fact that Linear B used monolingual English post-editors, the environment they worked in looks reasonably close to current practices we have examined in this report. Linear B naturally failed to win the evaluation competition, but it got by far the highest score! It also suggested that triggering automatic searches for proper names and similar data on the Web might speed up one aspect of the post-editors task.

Future developments

In general, academic developers of MT have tended to view post-editing as the weak link in the MT value chain. One of their research targets is to remove the need for post-editing as an unfortunate human intrusion into a fully-automatic process.

However, as has been repeated throughout this report, the actual act of post-editing raw texts is itself a source of useful data for improving the translation automation system. The goal of a post-editing function should not simply be to quantify performance, but to boost performance quality of the entire system. Post-editing is more than a stage in a process; under the right conditions it is a vital source of process transformation.

This suggests that the MT user community needs to develop simple, clear and scalable methods to pool quality evaluation ratings, to agree on standards for such evaluations, and work hard to implement the best and most efficient methods into a shareable tool kit.

Information sources

Jeff Allen, Website devoted to Machine post-editing (<http://www.geocities.com/mtpostediting/>) A post-editing evangelist, with a long list of author's papers, but less focused on industrial-strength MT solutions.

Catherine Hyland, "Testing "Prompt"; The Development of a Rapid Post-editing Service at CLS Corporate Language Services AG, Switzerland" in MT Summit 2003, available at: <http://www.amtaweb.org/summit/MTSummit/FinalPapers/100-Hyland-final.pdf>. An apparently discontinued service in light post-editing for a financial customer.

PAHO Machine Translation System

Document Center: http://www.paho.org/english/am/gsp/tr/mt_docs.htm Useful papers on an comprehensive organizational approach to MT practice and post-editing

Claus Povlsen and Anneliese Bech, "Ape: Reducing the Monkey Business in Post-editing by Automating the Task Intelligently" Proceedings of MT Summit VIII, 2002 <http://www.eamt.org/summit-VIII/papers/povlsen.pdf>

Falko Schäfer, "MT post-editing: How to shed light on the "unknown task". Experiences (sic) made at SAP (from the EAMT archive)

Dorothy Senez, "The Machine Translation Help Desk and the Post-Editing Service" Terminologie & Traduction 1-1998, OPOCE, European Commission, pp. 289-295.

Many thanks to TAUS members and many others cited here who provided input to this survey.

About TAUS:

TAUS is a networking community for corporate and institutional users, developers and practitioners of authoring, translation and localization technologies and services. By sharing best practices and intelligence in cross-industry meetings and online forums the members aim at advancing the adoption of translation automation technologies.

TAUS Reports cover:

- *Different approaches*: Introductions to the key areas of translation automation.
- *Best practices*: Overview of best practices in applying technologies. Best practice reports are regularly updated.
- *User cases*: Analyses of processes in member and non-member companies.

For more information on TAUS, see: www.translationautomation.com.

Replies, questions and observations can also be sent to:

Email: jvdm@translationautomation.com

Director: Jaap van der Meer

Address: Oosteinde 9-11, 1483 AB De Rijp, Netherlands, tel. +31 299 672028

©2006 by TAUS B.V. All rights reserved.

TAUS Reports are published by TAUS B.V. exclusively for members. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means.