Chapter submitted for publication

Chapter 35

Editing in Translation Technology

Christophe Declercq

University College London, the United Kingdom

Language and Translation Technology

With the emergence of translation memory technology in the early to mid-1990s,[1] the translation profession underwent a true technological turn that had been eagerly awaited by those working on machine translation systems since the 1950s. At the core of the translation memory systems (TMS) was a database of human translations, aided by the machine: machine-aided human translation, MAHT, or computer-aided translation, CAT. With the segment-based approach of re-use of previously translated material, traditional concepts and workflows changed dramatically too. Other than language skills and writing abilities, translation of texts included an increasing use of computer technology. Processes such as editing, revision and proof-reading should follow suit, but to date translators are struggling to cope with the speed of translation technology uptake.[2]

---

[1] Any historical review of translation memory systems will point at Trados MultiTerm and IBM Translation Manager emerging in 1992, Atril's Déja Vu in 1993 and Trados Workbench in 1994. The concept, however, emerged much earlier, with Peter Arthern already in 1979 stipulating that the use of unrestricted machine translation at the European Commission might very well be too early still, but that there was "scope for post-edited machine translation of a restricted range of texts" (Hutchins 1998: 293).

[2] Crucial in the perception of language and translation technology is Google Translate, which more than a year after it became a paid for service, still had more than 200 million people using it monthly. By April 2012, the daily total number of words equalled that of 1 million books (Kerr 2012). The influence of Google Translate for instance goes well beyond the albeit limited online tool and Google Translator Toolkit. With a free Website Translator

In the last few years, that uptake has assumed the shape of machine translation (MT), especially statistical machine translation (SMT), which has taken the translation industry by storm. Yet, it would not be accurate to brand MT as new. In fact, MT is into its 7th decade. Nonetheless, deeply rooted reservations about quality output of translation engines remain commonplace among translators. Yet, that scepticism among many translators is overcome by the daily dependency of millions of users on Google Translate or Microsoft's Bing Translator and by the incorporation of thousands of customised translation engines the world over.

An in-depth analysis of 'editing and translation today' therefore not only looks into the more traditional workflow of using a translation memory (TM) and its processes of verification such as editing and proof-reading, but also looks into the emerging convergence of translation technology (in particular translation memories) and language technology (especially machine translation).[3] This merging movement becomes clear from any description of the term translation technology. Translation Technology (TT) concerns the applied use of any computer application that supports the translation process as performed by a human translator (HT). But even then the description is flawed as any machine translation output for each no match in a translation project that is performed in a translation memory system (TMS) is in fact language technology (LT), human language generated automatically by a computer system.

As no clear delineation can be established between translation technology and language technology,[4] this contribution therefore aims to include any automated means of facilitating productivity and/or quality of human translation. The following overview will therefore provide an insight into 21st century translation whereby the

---

plugin websites can be made available in more than 60 languages. In fact, Google's Website Translator allows users to edit and improve their websites themselves.

[3]   For an appreciation of the history of post-editing machine translation see Ignacio Garcia 2012.

[4]   For instance, re-use from a large TM on the basis of aligned source and equivalent target texts or MT output from an SMT that has been trained on the same or similar corpus of equivalent texts are perhaps distinctively different in technology but closely related in use.

benefits of technology at the disposal of translators lies in the balance with a further acceptance of post-editing both TM and MT output.

'Traditional'Translation Technology and Editing

Traditional translation technology has at its core the translation memory system (TMS) and is likely to be supported further with a terminology database.[5] The TM application which re-uses segments that matches previously translated material has also been described as machine-aided human translation (MAHT), computer-aided translation (CAT) or translation environment tools (TEnTs). Whatever the acronym and whatever the definition of a TMS, besides re-use of previously translated segments (or translation units), a key feature is often overlooked by translators themselves: TM systems allow translators to deal with complex file formats they not necessarily master themselves.[6] They receive and deliver files in their native formats without interfering with the underlying code such as cross-references or mark-up language.[7] Typical material concerns FrameMaker or InDesign, but also DITA XML or Microsoft .NET files. As such a main benefit of any TMS is that it allows

---

[5]   Although terminology management and the inclusion of term recognition in TM systems is not discussed here, it should be made clear that terminology is not only key to the HT, but also to MT. A combined hybrid TM/MT + terminology management allows for an increased quality assurance and if maintained successful also an increased consistency and thus quality. Also dictionary compilation is a skill crucial in the development of translation engines.

[6]   From here onwards, this contribution does not allocate much space to defining several of its key concepts, let alone analysing the differences between respective definition variants. The applied field of translation technology itself, a world of increased productivity, does not warrant such ponderings.

[7]   Biau Gil, however, attests that translators' perfomance is improved by an environment whereby the non-verbal elements of a text or its native format are visible in an interface that is similar to WYSIWYG, *What You See Is What You Get*: "texts translated using WYSIWYG translation interfaces include fewer errors than those translated using non-WYSIWYG interfaces" and that "when translators use WYSIWYG translation interfaces they work faster than when they use non-WYSIWYG interfaces" (2007:7).

translators to be translating and editing much more material than in any typical word processing environments. With the translation interface, the user interface of any TM environment in which translators visually see the text on screen as they edit it (Biau Gil 2007), translating and editing converge also.

Not all types of matches from a TM occur in just any translation project. In the screenshot below, the fourth and last segment still need translating from scratch and no source was copied across there. All the 100% matches are re-used from the TM. Whether or not the perfect matches need editing or not, should depend on the quality of the TM results, the formatting and quality assurance settings, the project requirements and the experience of the translator. However, this is often limited to contractual obligations that urge the translator not to alter any perfect match. Note that the named entities make up about 35% of the overall word count. Copying across the source segments with added short cut expertise to be jumping across words in the target segment most certainly constitutes an increase in productivity (especially for this text type, i.e. sports).

Andrew Riley haalde het in 13"39.	100%	Andrew Riley won the event in 13:39.
Adrien Deghelt was zesde in 13"78, Damien Broothaerts zevende in 13"80.	100%	Adrien Deghelt claimed sixth, Damien Broothaerts seventh.
Beiden haalden de limiet voor het WK in Moskou (13"50) niet.	100%	Both Belgian athletes did not make the threshold for the Worlds in Moscow.
Sara Aerts en Eline Berings gingen in de 100 meter horden op zoek naar de WK-limiet van 12"94.		
Aerts strandde op de vierde plaats met 13"02.	100%	Aerts claimed fourth at 13:02.
Anne Zagré, al zeker van een WK-ticket, liep 13"07 en was vijfde.	100%	Anne Zagré, who already secured her ticket, managed 13:07.
Berings werd achtste in 13"54.	100%	Berings came 8th in 13:54.
Kristi Castlin won in 12"77.	100%	Kristi Castlin won the event in 12:77.
Abraham Rotich kon geen twee jaar op rij de 800 meter winnen.		

Picture 1: Detail of the Editor Environment of SDL Trados Studio 2011 (SP1), with 3+1+4+1 Units

However, beyond those stipulations, each degree of matching requires different cognitive processes of the translator. Whereas often minor brief additions or alterations might improve a segment to the level that it is acceptable for the purpose it is used,[8] research of how translators maintain their awareness of possible flaws while

---

[8] One such purpose is to maintain the standard or open format in which the translation memory is contained. XLIFF (XML Localisation Interchange File Format) allows users of translation technology to pass on data between various tools during the translation or localisation process. XLIFF Editors can be found among more familiar providers ot translation technology tools such as MultiTrans as well as through lesser known freeware,

re-using translation units from the TM might be relevant to analyses of editing MT output too.

Cognitive Processes and Editing

Lagoudaki 2006 was a reference work about the translators' perception and use of technology, but translation environments have moved on.[9] Among others, the pervasive use of SMT has effectuated a new paradigm in that perception of language and translation technology. More importantly, in the last few years, translation memory systems have broken away from the – admittedly often preferred by translators – environment of word processors and moved to standalone applications and online software as a service (SaaS). However, what has remained ever since the increased uptake of TM systems in the 1990s, is the widespread concern of translating and editing in a TMS. Based on an empirical study, Dragsted 2008 proved that any TM's segmentation into units, usually sentences, creates a strong focus on those segments, which affects the overall quality of the translation as a final product.

With a text that is presented in a TMS in various segments or units, a sentiment of alienation lies in the balance with a steady pace and a structured approach. In fact, with translation technology as a form of human-computer interaction, it is very difficult to differentiate formal benefits/disadvantages from holistic ones.

Benefits of (segmentation in) a TM	Disadvantages (of segmentation in) a TM

---

such as Transolution. Other file formats that drive the translation editing environment are for instance Poedit, which allows translators and users to edit cross-platform gettext catalogs (PO files). SRT Translator provides a translation memory in which Google Translate produces draft translations of subtitles.

[9] The Copenhagen Business School has been particularly active in researching the cognitive processes while translating using a TM and the effects of segmentation on the productivity and quality of the translator. Dragsted 2004, Dragsted 2008, Jakobsen 2009, Christensen and Schjoldager 2010 and Christensen and Schjoldager 2011 are but a selected few. Other people who have contributed to this field are Bowker 2005, Guerberof 2009, O'Brien 2008, O'Brien 2011 and Pym 2011.

- A sense of control on the segment level - Similar pace - Close reading, no interference of non-verbal elements - Added value of term recognition - No formatting issues - Increased accuracy and consistency - Being able to monitor progress - Auto-propagation - Possible copying across of the source segment	- The layout of the source text is lost - No feeling of overall view and alienation from the context - Lack of non-verbal elements affects quality and productivity (Biau Gil 2007) - Lack of control - Formatting sometimes still requires editing - A tendency to more literal translation

Table 1: Benefits and Disadvantages of Segmentation in Translation Memory Systems

Whether segmentation leads to an increased tendency towards more literal translation or not, remains a matter for scholars to discuss and for further empirical studies. In the debate about the consequences of segmentation, experience and maturity are often overlooked, along with the need for increased productivity. In fact, in his pilot study Biau Gil proves that subject-matter knowledge is more relevant than visual information (2007: 7). Taking this finding across the TM/MT threshold already, this is a further argument that post-editors should above all be knowledgeable about the subject topic.

Forms of Editing, Other Than Translating

Editing in projects that involves translation technology run along two axes. A first axis ranges from TM to MT. A second axis then concerns editing, ranging from pre-editing to post-editing. As pre-editing and controlled language are discussed elsewhere in this encyclopaedia, post-editing is broken down into more sub-concepts. Editing, revision and proof-reading are fundamental elements in translation projects and as a consequence their validity in MAHT projects is equally important.

Translation Service Providers (TSPs, sometimes also referred to as LSPs, Language Service Providers) adhere to the TEP model (translation / editing / proof-reading). However, in marketing their services the added value, especially of proof-reading, is often sold as a separate service. In the next section, the differentiation between the various forms of going over a text other than translating is effectuated in

a sense of best practice, not in an academic overanalysing of terminological diffusion. [10] Publications and/or guidelines on editing, revision and proof-reading often concern a mere modal framework, 'how revisers *ought* to go about their jobs or what jobs they *could* use' (Mossop 2007, online), and eventually best practices or workflows for revisers are often based on experience anyway.

Comparing the translation with the original text and ensuring that there are no errors left such as spelling mistakes, grammatical errors, omissions or ambiguities, is a well-established practice by the Translation Bureau of the Public Works and Government Services Canada. In their style guide long lists of possible errors in both writing and editing are produced. However, much of this list is aimed at text-production and not necessarily at translation projects in computerised setting. The error categorisation by the Canadian Translation Bureau proves that translation technology increased the speed of how editing (of errors) and translation merged: translation memory tools started to elaborate on their proprietary quality assurance functionalities (such as verification in SDL Trados Studio 2011). Companies have been working towards this trend too, as can be seen with Yamagata Europe's QA Distiller.

Whether in QA Distiller, in Studio or in any other TMS, detection of possible errors has become very much an automated feature of translation projects too. This greatly enhances the consistency of translator's output as well as his/her ability to be submitting a formally flawless target file, but it also provides a learning curve for translators to become more experienced in translation quality assurance and as such set themselves apart for those who do not.

In order to distinguish between the various forms of editing and the various identities editing can assume, a practical overview is reproduced below, whereby the various forms of editing are in fact allocated a position in the workflow.

---

[10]   In analogy to doctors being the worst patients, translators have a similar ailment: perennial analysis of concepts, their definitions and denotations, and a subsequent ongoing debate about the slight differences.

Source files		Translated files		Final files	
Formal QA	Non-formal QA	Formal QA	Non-formal QA	Formal QA	Non-formal QA
Manual check QA software	Editing Controlled language Authoring memory, etc.	Manual check QA software	Proofreading Editing	Manual check QA software	In-country review

Figure 2: Editing Stages in an Overall Quality Assurance Approach (Makoushina and Kockaert 2008: 3)

Makoushina and Kockaert 2008 place editing of the translated files along proof-reading and deem it a non-formal form of quality assurance. With this approach, editing 'after' the translation (either HT or MT), 'post-editing', and editing of source files, 'pre-editing', are differentiated clearly as stages in the translation workflow. As mentioned earlier, pre-editing and controlled language are not the scope of this article, as they are dealt with elsewhere, but (post-)editing still needs to be set apart from proof-reading.
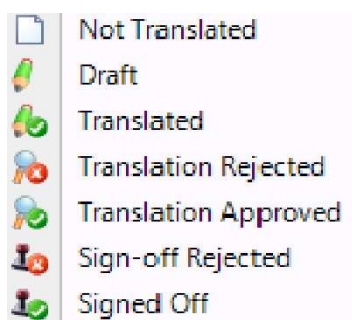
Editing, Revision and Proof-reading

In 2006, the European Committee for Standardisation (CEN) published the EN15038 standard,[11] developed for Translation Service Providers. The standard aimed to cover the entire translation process, including quality assurance. The standard offered TSPs and their clients a breakdown of the entire translation provision in accurate definitions and standard description. Most importantly, the European standard required both a translator and a reviewer for each translation and differentiated between the two. Under EN15038 only translators with the appropriate background and competences can translate documents and it is the task of that

---

[11]   Even though it was published in 2006 already EN 15038 has been gaining acceptance. It was accepted by 28 nations (all EU member states, except Bulgaria and Croatia, but it was accepted by non-EU Iceland, Norway and Switzerland) after its inception and acted as a benchmark in the European Union.

translator to check the translation themselves already. [12] A reviewer then is a subsequent person in the translation workflow who examines 'a translation for its suitability for the agreed purpose, and respect for the conventions of the domain to which it belongs' and who recommends corrective measures, if necessary. A review can be distinguished from a revision in that in the case of the latter, a translation is examined with both source and target texts compared. According to the European standard, proof-reading is limited to checking of proofs.[13]

These concepts and their allocated positions in the translation workflow are often mimicked by the translation tools themselves. In the Editor window of SDL Trados Studio 2011, the status of each translated segment can be altered, including translated and reviewed. This is similar to what XTM Cloud offers. Across Systems takes this even a step further and includes buttons for the various steps in the translation process and aligns them with the EN15038 standard.[14]

| Picture 3. Segment status in SDL Trados Studio 2011. | Picture 4. Various translation workflows possible in XTM Cloud. |

---

[12]   This check by the translator is also called self-editing.

[13]   The Language Resource Centre of the Aalborg University refers to proof-reading as follows the process where 'we focus exclusively on orthography, typing errors, grammar and punctuation'. Vocabulary and spelling are proof-read so as to make them consistent. For English-language texts 'either British or American spelling is used, and not a mixture of the two varieties of English'. In the case of an ambiguous translation a comment is inserted explaining the problem, but the text itself will never be re-phrased. (LRC 2009, online)

[14]   Across Systems uses a slightly different terminology: the corrector and reviewer ensure checking, revision, reviewing and verification.

In EN15038, editing in any form (copy-editing, pre-editing, post-editing) is included in appendix only, as an added value service, but just how editing differs from review, revision and proof-reading is not very clear.[15] It can, however, be easily deduced from the descriptions what editing is and what it is not:

Elements editing shares with EN15038 stipulations of review, revision and proof-reading	Elements editing does not share with EN15038 stipulations of review, revision and proof-reading
• Altering a translation for its suitability for the agreed purpose • Matching the translation to the conventions of the domain to which it belongs • A level of comparing source and target text is involved	• Checking of proofs (even though it can be argued editing shares elements of checking of proofs on screen) • Recommendation of corrective measures (even though it can be argued editing pro-actively ensures these measures)

Table 2. How EN15038 could possibly set editing apart from review, revision and proof-reading.

Still, as already indicated by the various forms of editing, the above stipulations do not entail a set of practical guidelines on how editing is used in translation projects,

---

[15] According to Mossop 2007, in editing a translation project, corrections and improvements are made whereby the purpose and the given readership of the text are prioritised. Revising is a very similar task, but this is then applied to draft translations. Trying to rename all the PEMT, post-editing machine translation, as PRMT (post-revision?), seems not immediately feasible. In light of Mossop 2007, it could be argued that post-editors first revise the MT segments first and edit the text in its entirety next. In practice, this would hardly happen and texts are translated and subsequently edited on a segment-by-segment basis. These corrections to a translation in order to increase its quality is also known as Quality Assurance (QA), whereas any correction round to detect flaws in a translation after it has been submitted is often referred to as Quality Control (QC). For an appreciation of QA and QC, see Makoushina and Kockaert 2008, Rasmussen and Schjoldager 2011 and European Union 2012.

be it in the strictest sense by means of a translation memory system or in a broader interpretation of translation technology. This then not only includes machine translation, but also social media (crowd-sourced translations or community translations), sometimes both are combined even (as if often the case with projects posted on platforms such as Transifex). But most importantly, editing alongside translation and/or language technology takes the shape of post-editing machine translation.

Post-editing and Machine Translation

Post-editing machine translation concerns the practical answer to the longstanding quest for the Holy Grail: machine translated material that is substantially good enough for communication and/or dissemination[16]. A valid example of how practical post-editing MT can be, is Jeff Allen's *Creole MT*,[17] a publicly available MT system for the purpose of relief during the 2010 Haiti earthquake and its aftermath.

Understanding the choices translators make while working with translation technology such as translation memories can be of significant relevance on how to approach the influence of translation provided by machine translation. Even when translation scholars have considered the 'black box' of machine translation in the past, it was in opposition to Holmes's 'little black box of the translator's mind' (Holmes 1972: 72). However, especially when post-editing machine translation (PEMT) is concerned, the two in fact are more in juxtaposition and will be converging more in the future still. Above all, post-editing should be seen as a process of improving through modification (rather than revision) a machine-generated translation, often eyeing a minimum of effort on behalf of the post-editor.[18] The quicker the turn-around

---

[16]   John Hutchins differentiates between MT for the purpose of communication (light post-editing required only) and dissemination (full post-editing required). (Hutchins 2013)

[17]   For an appreciation of the language technology effort for distress relief in Haiti, see Munro 2010.

[18]   The description of post-editing is a combination of two definitions: post-editing is 'the process of improving a machine-generated translation with a minimum of manual labour'

needs of a translation, the more likely the PEMT effort will be a fast one, also known as 'light post-editing.' More thorough modifications, with less of a time urgency, eyes a better quality and is often known as 'full post-editing'. The latter category is the more common one, not least because it aims to obtain a quality level that is the same as if the entire text would have been translated from scratch by the human translator.

Source Text	Raw MT
Un vaste réseau qui piratait les codes de déverrouillage des téléphones portables a été démantelé, ont annoncé, dimanche 26 septembre, les enquêteurs.	A vast network hacked unlock codes for mobile phones has been dismantled, announced Sunday, Sept. 26, investigators.
Example of Light Post-Edit	A vast network which hacked unlock codes for mobile phones has been dismantled, it was announced Sunday, Sept. 26, by investigators.
Example of Full Post-Edit	A vast network which hacked security codes for mobile phones has been dismantled, according to an announcement by investigators on Sunday, Sept. 26.

Picture 5: Light and Full Post-editing of Raw MT Output (O'Brien 2010: 5)

The quality of a translation is a hotly debated issue, let alone the quality of a translation in which MT played a part, and subsequent post-editing. O'Brien 2010 rightly argues that the quality expectations differ depending on where a particular person is involved. Developers are very interested in automated quality metrics such as BLEU (Bilingual Evaluation Understudy), TER (Translation Edit Rate) or WER (Word Error Rate).[19] They are also very keen on getting usage feedback from the translator, improving the system they have developed with valuable input.[20] Buyers

---

(TAUS 2010) and 'a process of modification rather than revision' (Loffler-Laurian 1985 in O'Brien 2010).

[19]   For an appreciation of machine translation evaluation metrics see Snover 2006. Users can compare users Google Translate or Bing Translator through iBLEU.

[20]   This is where pre-editing re-emerges: by comparing the raw MT output with the source text, errors can be found and arguably a system behind types of errors too. Other than leaving things as they are, developers have two options: boost the engine by training it on new

allocate PEMT projects to translators or TSP's because they hope for a faster turnaround. The overall translation cost might be similar to HT, if the PEMT approach saves time, then that is a major benefit for the buyer already. The translators or TSP's hope that by increasing their productivity, they can also increase their client portfolio and/or market share. Two categories that are often overlooked are the project managers[21] and the account executives or sales. These people do not necessarily carry the need to be included in the list just now, but they are very crucial in the communication chain with the client and its subsequent users and as such cannot afford themselves to be creating false expectations. In the end, much of the success of post-edited machine translations depends on how the users have perceived the quality of what was disseminated or communicated.

In the entire debate of considering raw MT output as fuzzy matches so as to gauge the probable workload for post-editors properly, Guerberof 2009 analysed findings of a small-scale research project that are very interesting. Translators were asked to post-edit TM segments of 80-90% fuzzy matching on the one hand and SMT output on the other hand, as well as translate anew. In an analysis of all the errors produced in each of the three categories, new segments accounted for roughly 1 error in 5. Intriguingly, a similar amount of words to be post-edited triggered not many more errors. In fact, the errors in the final translation produced with the aid of a translation memory accounted for half of all the errors, i.e. editing fuzzy matches in a TMS triggers double the amount of errors compared to post-editing raw MT output.[22]

---

data or allowing document authors to pre-edit their source material so as to have an increased raw MT output quality.

[21]   For an appreciation of machine translation and project management, see Guerberof 2010.

[22]   When Guerberof categorised the errors according to five types (mistranslation, accuracy, terminology, language and consistency), post-editing raw MT output produced very similar numbers of errors for language and consistency as the new segments did. With double the errors for mistranslation and accuracy, it should then come as no surprise that re-using and editing fuzzy matches from the TM in fact landed more than half the errors for the three approaches together, whereas MT only did for a quarter.

Similarly, using the TM even slowed down productivity by 2.5%, whereas MT increased this by 24.5%, a combined difference of 27% or nearly a third.

The re-usable nature of raw MT output has been confirmed by Fontes 2013, chair of the European Commission's MTUG (machine translation user group). In a survey across the Directorate-General for Translation experienced translation were asked to rate MT output quality. Of the 643 ratings of language pair combinations, 200 ratings confirmed that they had used MT for more than 75% of their translation jobs. Asked to rate the output of the respective engines on a 0-4 scale, 726 ratings were delivered. 185 people rated the MT quality as 4 or 3, in which most segments were considered re-usable. Asked for the reasons why MT should be used 3 of the 5 answers[23] (MT is a typing aid, MT is a source of inspiration for alternative translations available in the translation memory and a quick draft) imply subsequent use of post-editing.

Post-editing Guidelines

TAUS, the Translation Automation Society is one of the most authoritative source on post-editing machine translation. Crucial to raising the awareness with users of PEMT about the various issues involved, they have highlighted recommendations and post-editing guidelines.

On the recommendation of tuning your engine appropriately TAUS 2010 distinguishes between rule-based or statistical engines, whereby a high-level dictionary and linguistic coding is crucial for RBMT and clean, high-quality, domain-specific data are key to data-driven systems. The second TAUS recommendation is to ensure that the source text is written well, preferably written with later MT in mind even. As mentioned earlier: there is no post-editing machine translation without including pre-editing the source material.

One of the most obvious recommendations by TAUS 2010 is to train post-editors in advance. However, there is a major difference between training people to act as post-editors for a specific job with project-specific data and guidelines on the

---

[23]   Other responses referred to an increase in productivity and a gain in time for more thorough research.

one hand and linguists on the other hand who receive more basic training because they work across projects and therefore need to adhere more to a common denominator. Moreover, including post-editing into the curriculum of higher education has proven a difficult feature.[24]

Providing generic guidelines for achieving quality that is in line with the project stipulations and the agreed expectations is not easy, as TAUS 2010 proves. Most guidelines, twice half a dozen of bullet points only, remain very tentative and do not immediately constitute a checklist. However, in line to the quality assurance capacities of translation memories mentioned earlier, several guidelines can in fact be dealt with in the automated environment of a TMS:
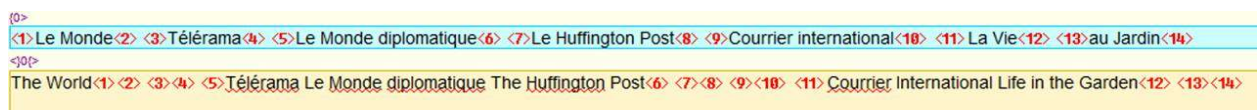
Selected guidelines for post-editing (TAUS 2010)	Quality Assurance in SDL Trados Studio 2011
'Ensure that no information has been accidentally added or omitted.'	QA Checker 3.0: Segment verification - Check for forgotten and empty translations - Check for segments where source and target are identical - Check for segments which are x% shorter / longer - Segments to exclude
'Basic rules about spelling, punctuation and hyphenation apply.'	QA Checker 3.0: - Inconsistencies (repeated words in target, unedited fuzzy matches) - Punctuation - Numbers, times, dates, measurements
'Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of 'Do Not Translate' terms.'	QA Checker 3.0: Word List and Regular Expression Terminology Verifier (with a term base open)

---

[24]   For an appreciation of teaching post-editing, see Allen 2001, Kenny and Way 2001, O'Brien 2002, Belam 2003 and Kliffer 2008.

'Ensure that formatting is correct.'	Any TMS strives towards maintaining exactly the same formatting between source and target. Most TMS also include warning messages in case where there are differences.

Table 3: TAUS Post-editing Guidelines versus Quality Assurance in SDL Trados Studio 2011

In a combined approach of the above, the text segment represented below, which could have been reproduced in many other TMS too, requires actions on both levels: in the TMS of Wordfast Anywhere (WFA) formatting has not been reproduced appropriately by Google Translate. A post-editor would need to restore the tags. However, this would be picked up on already by the verification feature of WFA. The post-editor would have to restore some cultural elements back to the source text and this example indeed triggers the copying across of the source segment.

Picture 6: Tag Differences Returned by Google Translate in Wordfast Anywhere (Text by Le Monde)

So far, no proprietary environment for post-editing alone has been mentioned and even though they are around (such as PET by Wilker Aziz and Lucia Specia), it should be clear that post-editing can happen very well in the environment of a TMS. It should be noted that post-editing is also required in platforms for crowd-sourced translations such as Transifex, live subtitling with speech recognition or subtitling editors such as dotsub and Youtube Subtitler.

With post-editing material that has been provided by a translation memory, machine translation or even speech recognition, pricing methods are a tricky business. Three common options apply. Other than having a linguist available in-house (for public broadcasting and live captioning for instance), either a nominal fee is paid based on the time spent or a word rate is agreed, differentiation between re-use from the TM (see earlier categories of matches) and machine translation (which differs based on the training data and the input). Eventually PEMT is paid along the lines of

fuzzy matching.

Conclusion

While on the Eurostar into London, the author wanted to joke with friends who also use *Road Bike*, a cycling app. After travelling at about 285km/h on average for 5 minutes, the live tracking was stopped and as the 20660 kcal were about to be sent via Gmail, the app, which had been installed in Dutch along with the operating language of its Android 4.1 system, neatly indicated 'U gaat wel erg snel. Wellicht heeft u de verkeerde sport gekozen' [*You are going very fast. Perhaps you have chosen the wrong sport.*, MT by Google Translate]. It would be very difficult to find out whether this segment had been localised into Dutch by a translator (who might have used machine translation for draft output and treat it as fuzzy matches), by machine translation *tout court* or by a community of users that master Dutch. Such a community can use a platform such as Transifex, which in its turn can have community members that base their work on machine translation. Although this anecdotal instance does not prove much, it will be recognised by millions of users, 99.9% of whom are not translators nor linguists. The world of translation technology, language technology, mobile technology and social media (the people networks, the cloud and the crowd, and subsequently the feed of social data too) are converging.

With that rapid uptake of machine translation at a low entry level, but also on mobile phones and on tablets, the perception of translation from the global user's perspective is changing dramatically. The main problem in overcoming that threshold fear by translators to be incorporating machine translation in their workflow, and therefore post-editing, is that translators deem the process of translation sacred, whereas eventually the target text is only a product with a purpose that is relevant to a world outside their own. If the wider translation profession does not see the opportunity to still be maintaining a much cherished art and profession, too many users will discard the human translator and resort to MT output that has been post-edited by either a native speaker or someone who knows the subject really well. The latter can very well be someone who is trusted within the (online) user community.

Editing in translation technology applications is an elementary step in the well-sought increase in productivity. Any target text that is the product of a translation process should be considered complete only after careful revision and editing. Reviewing segment after segment whereby that process has been produced by a

computer application can indeed be more cumbersome than to edit a human translation. However, if translation as a process and the means to an end product, whether by a human, a machine or hybrid, needs post-edition and this is not mastered by the human translators themselves, then who will fight the corner of the added value of humans here?

Arguably most clashes between quality expectations and deliverables can be overcome beforehand. By examining raw MT output quality an appropriate price needs to be negotiated and an agreement needs to be reached about the final quality of the information to be post-edited. Even though these two recommendations are including in those by TAUS 2010, they in fact constitute common practice in projects that involve HT only or HT+TM. However, it is ever so true that the ongoing new paradigm of pervasive use of MT can indeed act as a technological turn that triggers an awareness HT has not been able to do for decades. Including MT output in translation projects offers an opportunity to start negotiating this awareness anew. It would be lethal to miss out on that.

References

Across Systems (2011) 'The EN 15038 Standard Workflow'. Available at: www.across.net/documentation/onlinehelp/across_en/acrossHaupt648.htm.

Allen, Jeff (2001) 'Post-editing: An Integrated Part of a Translation Software Program', *Language International* April, 26-29.

Aziz, Wilker and Lucia Specia (2012) 'PET: A Tool for Post-editing and Assessing Machine Translation', in *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, 28-30 May 2012, Trento, Italy, 99. Available at: www.lrec-conf.org/proceedings/lrec2012/pdf/985_Paper.pdf.

Belam, Judith (2003) 'Buying up to Falling Down: A Deductive Approach to Teaching Post-editing', in *Proceedings of MT Summit IX*, 23-27 September 2003, New Orleans, Louisiana, the United States of America. Available at: http://www.dlsi.ua.es/t4/proceedings.html.

Biau Gil, José Ramon (2007) 'What You See Is What You Get? A Pilot Experiment on Access to Visual Information in Translation Interfaces', Paper presented at the 12th Annual Internationalisation and Localisation Conference, 26-28 September 2007, Dublin, Ireland. Available at: www.localisation.ie/resources/conferences/2007/presentations/Biau_LRC_XII/LRC_XII_slides.pps.

Bowker, Lynn (2005) 'Productivity vs. Quality: A Pilot Study on the Impact of Translation Memory Systems', *Localisation Focus* 4(1): 13-20.

Canada Translation Bureau (1997) 'The Canadian Style: A Guide to Writing and Editing'. Available at: http://bt-tb.tpsgc-pwgsc.gc.ca/btb.php?lang=eng&cont=791.

Chan, Sin-wai (2004) *A Dictionary of Translation Technology*, Hong Kong: The Chinese University Press.

Christensen, Tina Paulsen, and Anne Schjoldage (2010) 'Translation-Memory (TM) Research: What Do We Know and How Do We Know It?' *Journal of Language and Communication Studies* 44: 89-101.

Christensen, Tina Paulsen and Anne Schjoldage (2011) 'The Impact of Translation-Memory (TM) Technology on Cognitive Processes: Student-translators' Retrospective Comments in an Online Questionnaire', in Bernadette Sharp, Michael Zock, Michael Carl, and Arnt Lykke Jakobsen (eds.) *Proceedings of the 8th International NLPCS Workshop: Special Theme: Human-Machine Interaction in Translation*, 20-21 August 2011, Copenhagen, Denmark, 119-130.

Crabbe, Stephen (2010) 'Controlled Languages for Technical Writing and Translation', in Ian Kemble (ed.) *The Changing Face of Translation: Proceedings of the 9th Annual Portsmouth Translation Conference 2009*, 7 November 2009, Portsmouth, the United Kingdom, Portsmouth: University of Portsmouth, 48-62. Available at: http://tinyurl.com/b8truqo.

Data Translations (2013) Services. Available at: www.datatra.be/en-services-revision-proof-reading.

Dotsub, http://dotsub.com.

Dragsted, Barbara (2004) 'Segmentation in Translation and Translation Memory Systems: An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM System into the Translation Process', Copenhagen: Copenhagan Business School.

Dragsted, Barbara (2008) 'Computer-aided Translation as a Distributed Cognitive Task', in Itiel Dror and Stevan Harnad (eds.) *Cognition Distributed: How Cognitive Technology Extends Our Minds*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 237–256.

European Quality Standard EN-15038: 2006. Available at: http://qualitystandard.bs.en-15038.com/ (last accessed 1 February 2013).

European Union (2012) 'Quantifying Quality Costs and the Cost of Poor Quality in Translation: Quality Efforts and the Consequences of Poor Quality in the European Commission's Directorate-General for Translation', Luxembourg: Publications Office of the European Union.

Fontes, Hilàrio Leal Fontes (2013) 'Evaluating Machine Translation: Preliminary Findings from the First DGT-wide Translators' Survey', in *Language and Translation: Machine Translation*, European Commission. Available at: http://goo.gl/66H3a.

Garcia, Ignacio (2012) 'A Brief History of Postediting and of Research on Postediting', *Revista Anglo Saxonica* 3(3): 291-310.

Google Translator Toolkit. Available at: translate.google.com/toolkit.

Google's Website Translator. Available at: http://translate.google.com/manager/website/?hl=en.

Guerberof, Ana (2009) 'Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation', *Localisation Focus* 7(1): 11-21.

Guerberof, Ana (2010) 'Project Management and Machine Translation', *Multilingual* April/May. Available at: http://isg.urv.es/library/papers/2010_guerberof.pdf.

Holmes, James S. (1972) 'The Name and Nature of Translation Studies', Third International Congress of Applied Linguistics, Copenhagen. Available at: www.universita-mediazione.com/wp-content/uploads/2012/02/Materiale_Prof_Donadio_31_01_2012.pdf.

Hutchins, W. John (1998) 'The Origins of the Translator's Workstation', *Machine Translation* 13(4): 287-307.

Hutchins, W. John (2013) 'History and Methods of MT', Visiting Seminar, Imperial College London's Translation Unit, 27 February 2013.

ITR (2013) 'Document Translation'. Available at: www.itr.co.uk/documentation-translation.

Jakobsen, Arnt Lykke (2009) 'Instances of Peak Performance in Translation', *Lebende Sprachen* 50(3): 111-116.

Kenny, Dorothy and Andy Way (2001) 'Teaching Machine Translation and Translation Technology: A Contrastive Study', in *Proceedings of the Machine Translation Summit VII, Teaching MT Workshop*. Available at: http://doras.dcu.ie/15830/1/Teaching_Machine_Translation_%26_Translation_Technology.pdf.

Kerr, Dara (2012) 'Google Translate Boasts 64 Languages and 200M Users, *CNET Internet and Media News*, 26 April 2012. Available at: http://news.cnet.com/8301-1023_3-57422613-93/google-translate-boasts-64-languages-and-200m-users.

Kliffer, Michael (2008) 'Post-editing Machine Translation as an FSL Exercise', *Porta Linguarum* 9: 53-67.

Lagoudaki, Elina (2006) 'Translation Memory Systems: Enlightening Users' Perspective. Key Finding of the TM Survey 2006 Carried out during July and August 2006', unpublished doctoral dissertation, Imperial College London.

Lagoudaki, Elina (2009) 'Translation Editing Environments', in *Proceedings of the 12th MT Summit*, 26-30 August 2009, Ottawa, Ontario, Canada. Available at: http://goo.gl/sQHxS.

LionBridge (2013) Language Services. Available at: https://en-gb.lionbridge.com/translation-localization/language-quality.htm.

Madnani, Nitin (2013) iBLEU. Available at: https://code.google.com/p/ibleu.

Makoushina, Julia and Hendrik Kockaert (2008) 'Zen and the Art of Quality Assurance: Quality Assurance Automation in Translation: Needs, Reality and Expectations', in *Proceedings of the 13th International Conference on Translating and the Computer*, 27-28 November 2008, London ASLIB. Available at: http://goo.gl/l08EJ.

Mendez, José (1986) 'Machine Translation in Bureau Service', *Terminologie et Traduction* 1: 48-53.

Mossop, Brian (2006) 'Has Computerization Changed Translation?' *Meta* 51(4): 787-793. Available at: www.erudit.org/revue/meta/2006/v51/n4/index.html.

Mossop, Brian (2007) 'Empirical Studies of Revision: What We Know and Need to Know', *Journal of Specialised Translation* 8. Available at: www.jostrans.org/issue08/art_mossop.pdf.

Mossop, Brian (2007) *Revising and Editing for Translators*, 2nd edition, Manchester: St. Jerome Publishing.

Multitrans XLIFF Editor. Available at: http://multicorpora.com/products-services/options-and-add-ons/xliff-editor.

Munro, Robert (2010) 'Crowdsourced Translation for Emergency Response in Haiti: The Global Collaboration of Local Knowledge', in *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas*, 31 October – 4 November 2010, Denver, Colorado, the United States of America. Available at: http://amta2010.amtaweb.org/AMTA/papers/7-01-01-Munro.pdf.

O'Brien, Sharon (2002) 'Teaching Post-editing: A Proposal for Course Content', in *Proceedings of 6th EAMT Conference Workshop: Teaching Machine Translation*, 14-15 November 2002, Manchester, the United Kingdom, 99-106. Available at: http://mt-archive.info/EAMT-2002-OBrien.pdf.

O'Brien, Sharon (2008) 'Processing Fuzzy Matches in Translation Memory Tools: An Eye-tracking Analysis', in Susanne Göpferich, Arnt Lykke Jakobsen, and Inger M. Mees (eds.) *Looking at Eyes: Eye Tracking Studies of Reading and Translation Processing*, Copenhagen: Copenhagan Business School, 79-102.

O'Brien, Sharon (ed.) (2011) *Cognitive Explorations of Translation*, London: Bloomsbury Publishing.

Och, Franz (2012) 'Breaking down the Language Barrier — Six Years in', Online blog post, Google official blog. Available at:

http://googleblog.blogspot.be/2012/04/breaking-down-language-barriersix-years.html.

PoEdit. Available at: http://sourceforge.net/projects/poedit.

Pym, Anthony (2011) 'What Technology Does to Translating', *The International Journal for Translation and Interpreting Research* 3(1): 1-9. Available at: www.trans-int.org/index.php/transint/article/viewFile/121/81.

QA Distiller. Available at: www.qa-distiller.com.

Rasmussen, Kirsten Wölch and Anne Schjoldager (2011) 'A Survey of Revision Policies in Danish Translation Companies', *Journal of Specialised Translation* 15: 87-120. Available at: http://www.jostrans.org/issue15/art_rasmussen.pdf.

Snover, Matthew and Bonnie J. Dorr (2006) 'A Study of Translation Edit Rate with Targeted Human Annotation', in *Proceedings of the 7th Biennial AMTA Conference*, 8-12 August 2006, Cambridge, Massachusetts, the United States of America, 223-231. Available at: http://goo.gl/33Et1.

SDL Trados Studio 2011. Available at: www.sdl.com/products/sdl-trados-studio.

Somers, Harold L. (2003) *Computers and Translation: A Translator's Guide*, Amsterdam and Philadelphia: John Benjamins Publishing Company.

Somers, Nick (2001) 'Revision – Food for Thought', *Translation Journal Online* 5:1. Available at: https://sites.google.com/site/penidea/revision%E2%80%94foodforthought.

SRT Subtitler. Available at: http://sourceforge.net/projects/srt-tran/?source=directory.

TAUS (2010) 'MT Post-editing Guidelines'. Available at: www.translationautomation.com/images/stories/guidelines/taus-cngl-machine-translation-postediting-guidelines.pdf.

TAUS. Available at: www.translationautomation.com.

Transifex. Available at: www.transifex.com.

Transolution XLIFF Editor. Available at: http://sourceforge.net/projects/eviltrans.

Transperfect (2013) 'Human Translation Services'. Available at: www.transperfect.com/services/translation.html.

van de Poel, Kris, W.A.M. Carstens, and John Linnegar (2012) *Text Editing: A Handbook for Students and Practitioners*, New York: Academic and Scientific Publishing.

Verbeke, Charles A. (1973) 'Caterpillar Fundamental English', *Training and Development Journal* 27(2): 36-40.

Yamagata Europe (2013) *Quality Assurance*. Available at: www.yamagata-europe.com/en-gb/page/520/iso-certification.

Youtube Subtitler. Available at: http://yt-subs.appspot.com.