Institution: University of Wolverhampton
Unit of Assessment: 28 Modern Languages and Linguistics
Title of case study: Post-editing effort indicators for estimation of translation quality and productivity
1. Summary of the impact (indicative maximum 100 words)

This case study explores the impact of RGCL's Translation Post-Editing Tool[1] (PET) on *Hermes Traducciones y Servicios Linguisticos (Hermes), NLP Technologies Ltd. (NLPT)* and the *Department of Translation, Interpreting and Communication (DTIC)* at *Ghent University*. *Hermes* and *NLPT* are companies providing translation services in varied domains through a pipeline that combines translation technologies and post-editing. *DTIC* offers postgraduate courses on translation studies and interpreting, including subjects such as post-editing. PET enables the editing of pre-translated text, and the detection of 'effort indicators'. This helps improve assessment of translation systems/approaches, the quality of pre-translated text, and the effort needed to convert it into a publishable form. At *Hermes*, workflows developed using PET have reduced post-editing time by 31-34%; at *NLPT*, workflows optimised using PET have saved an average of 66 seconds of post-editing time per sentence. At *DTIC*, PET has been used to enhance the courses *Computer-Aided Translation* and *Technical Translation*.

2. Underpinning research (indicative maximum 500 words)

Post-editing machine translation (MT) output has been shown to be a successful way of incorporating MT into human translation workflows in order to minimise time and cost in the translation industry. The editing of pre-translated text is a common practice among users of translation memory (TM) tools, which provide user-friendly and functional environments for translators (e.g. *SDL Trados*, *Wordfast* or *Déjà Vu*). Post-editing cost is a function of word count and editing time; that is, the longer it takes, the more it costs.

RGCL developed PET [2], a stand-alone tool that allows the post-editing of output from any MT system and records the following information at the segment level: post-editing time, customisable quality scores, time-stamped events, keystrokes, and the edit distance between the original MT output and the revised version. The collection of this type of information has been largely neglected by other editing tools and enables the assessment of translation quality, diagnosis of translation problems, and estimation of the cost (and remuneration) of post-editing services.

PET can also be used to set fine-grained constraints on a post-editing task, such as limits on the time allowed for editing or on the character/word length of the final document (relevant for subtitling [1], for instance). Independent of any specific MT system, PET makes it possible to collect post-editing/revision information in a controlled way for multiple MT/TM systems. PET was first selectively released under non-disclosure terms in 2011 by Aziz and Specia, and is under continual development. On the 27th of July 2012 the tool was publicly released under LGPL and by November 2013 there were 281 registered downloads (35% from industry) and 2,124 visits to its main page (67.85% new visitors) according to Google Analytics. The tool has also been listed by TAUS Tracker, a free directory for MT, TM and language technology tools. To date, the original paper describing PET has received at least 21 citations according to Google Scholar.

RGCL then proposed the use of post-editing time as an objective quality discriminator [1]. The assumption is that good-quality MT should require little time to be post-edited. That is well aligned to the notion of productivity and how MT is used in practical/commercial scenarios. Moreover, in 2011, RGCL successfully proposed the optimisation of quality predictors towards post-editing time, as opposed to traditional subjective human scoring [5] and, for our findings, we earned the best paper award at EAMT-2011.

Measuring post-editing time may not be sufficient due to the variation in competence and cognitive abilities of different translators. For this reason, since 2011, RGCL has investigated ways in which time- and edit-based effort indicators gathered by PET can be used for the purpose of assessing and predicting translation quality and productivity [3, 4].

---

[1] http://www.clg.wlv.ac.uk/projects/PET/

In previous research, only rudimentary efforts have been made to explore possible effort indicators, these usually being restricted to automatic edit distances that i) do not reflect the real edits performed by users, and ii) do not consider the necessary time and underlying cognitive effort necessary to perform the task.

---

3. References to the research (indicative maximum of six references)

[1] Sousa, S. C. M.; Aziz, W.; Specia, L. (2011). Assessing the post-editing effort for automatic and semi-automatic translations of DVD subtitles. In Recent Advances in Natural Language Processing (RANLP-2011), Hissar, Bulgaria.

[2] Aziz, W.; Sousa, S. C. M.; Specia, L. (2012). PET: a tool for post-editing and assessing machine translation. In The Eighth International Conference on Language Resources and Evaluation, LREC '12, Istanbul, Turkey. May 2012.

[3] Koponen, M.; Aziz, W.; Ramos, L.; Specia, L. (2012). Post-editing Time as a Measure of Cognitive Effort. In the AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012). San Diego, USA.

[4] Aziz, W.; Mitkov, R.; Specia, L. (2013). Ranking Machine Translation Systems via Post-Editing. In Proceedings of Text, Speech and Dialogue (TSD2013). Pilsen, Czech Republic.

[5] Specia, L. (2011). Exploiting Objective Annotations for Measuring Translation Post-editing Effort.15th Annual Conference of the European Association for Machine Translation, pp. 73-80, Leuven, Belgium. Awarded Best Paper.

---

4. Details of the impact (indicative maximum 750 words)

Hermes Traducciones y Servicios Lingüísticos, SL, established in 1991, is a leading translation company with 100% Spanish capital, specialising in software and hardware localisation and also undertaking a broad range of other translation projects. At present, it is an assembly member of GALA (Globalisation and Localisation Association), ATA (American Translators Association) and ACT (Spanish Translation Companies Association). Furthermore, *Hermes* maintains a close relationship with Spanish universities by giving lectures in Translation and Interpreting, postgraduate courses, trainee programmes, localisation and applied technology seminars and sponsoring conferences and translation-related events, etc.

Hermes makes use of language technologies to speed up translation and achieves publishable quality via human post-editing. In the last two years *Hermes* has post-edited nearly 1 million words, a post-editing activity worth €90,000. These figures should increase by 15-25% in the next two years. It is therefore clear that reducing the cost of the post-editing practice is a priority for them.

NLP Technologies Inc (NLPT) is a Canadian company specialised in NLP that offers accurate and reliable certified translation services and technologies by employing Computer Aided Translation (CAT). It offers translation services in the legal, governmental, medical and financial domains and it has clients such as Alberta Education, Tribunal Administratif du Québec, Barreau du Québec, Commission des droits de la personne et des droits de la jeunesse and Fisheries and Oceans Canada.

Both companies benefit from general purpose translation technologies and domain adaptation strategies to fit their general purpose translation workflows to specific domains. In the past, *Hermes* and *NLPT* would optimise their translation workflows, including their domain adaptation strategies, towards semi-automatic metrics of translation evaluation such as HTER (an edit distance between a cheaply obtainable draft translation and its human post-edited version). However there is no guarantee that HTER and other state-of-the-art metrics correlate well with the real effort spent on post-editing as they are only an estimate of the real edits made.

In 2012 *Hermes* achieved a 31% reduction in post-editing, whilst 2013 saw a 34% reduction. This was accomplished by combining two uses of PET as follows:

1. optimising domain adaptation strategies towards a combination of post-editing time and edit distances (accurately measured editing time is one of the original outcomes of PET);

2. collecting data from the post-editing practice (PET's novel detailed effort indicators) that allowed them to learn an "aptitude profile" of their post-editors to the different types of texts (e.g. legal documents, movie subtitles, etc.). Such profiles were consulted before tasks were assigned so that the human subjects involved in a task were chosen depending on their aptitude for that task.

In a similar experiment that took place in the last quarter of 2011, *NLPT* used PET to assess their own domain adaptation in order to maximise their own efficiency. They report a study in which four professional post-editors worked on documents translated using different domain adaptation strategies. PET allowed them to objectively tune their strategies towards time savings in a controlled experiment. They have found that an average of 66 seconds could be saved per sentence (for over 78% of the cases) if their models were optimised towards editing time as opposed to HTER. These savings can be extrapolated to future tasks compatible with the domains analysed by matching their goal (time savings) with their evaluation methodology in a short and non-expensive analysis using PET.

The PET tool has also had an impact in education. In 2012-2013 DTIC introduced post-editing as part of their Master's programme and they chose PET due to its unique effort indicators as well as its flexibility in dealing with both human translation and post-editing. In their practical sessions, students participate collectively in experiments aiming at contrasting human translation and post-editing in terms of (i) quality, (ii) productivity, and (iii) typically encountered errors. While students are introduced to state-of-the-art translation and post-editing practices, they are also assessed in terms of (a) how thoroughly they apply guidelines, (b) the quality they achieve, (c) their own productivity, and (d) the errors they identify. PET's effort indicators, explicit assessments and history of edits are used for both purposes, namely, teaching (i-iii) and assessment (a-d).

Since 2012, thanks to PET, *DTIC*'s Master's students have been learning about and experimenting with post-editing, a practice that is of increasing importance in the translation industry and which is becoming indispensable to any modern translation professional. As a result of using PET, *DTIC* has designed novel and modern Translation Quality Assessment approaches which they put into practice in the postgraduate courses they offer.

---

5. Sources to corroborate the impact (indicative maximum of 10 references)

- Letter of support from the Managing Director of Hermes

- Letter of support from the President of NLP Technologies. Some of their findings have been published in the public domain:

  Sankaran, B.; Razmara, M.; Farzindar, A.; Khreich, W.; Popowich, F.; Sarkar, A. (2012). Domain Adaptation Techniques for Machine Translation and their Evaluation in a Real-World Setting. In Proceedings of the 25th Canadian Conference on Artificial Intelligence. http://rali.iro.umontreal.ca/rali/sites/default/files/publis/Farzindar-ai2012.pdf

- Letter of support from Lieve Macken, Lecturer responsible for the courses on computer-aided translation and technical translation at Ghent University. Some of their findings have been published in the public domain:

  Daems, J.; Macken, L.; Vandepitte, S. (2013). Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE. In the MT Summit XIV Workshop on Post-editing Technology and Practice. Nice, France.