Proceedings of the International Workshop on Expertise in Translation and Post-editing - Research and Application

Laura Winther Balling, Michael Carl, and Arnt Lykke Jakobsen (eds.)

Center for Research and Innovation in Translation and Translation Technology
Copenhagen Business School
August 17 and 18, 2012

Computer Assisted Translation (CAT) and Machine Translation (MT) technology are modifying the translation profession. However, it is unclear how translation technologies can help translators in the best way to produce high quality translations faster. The workshop Expertise in Translation and Post-editing aims at exploring how we can design advanced editing platforms to deploy translation technology in better ways than merely by post-editing machine generated texts, and how we can assess and compare the human expert behavior during translation and (post-) editing activities.

The workshop brought together 26 interested papers and presentations from the translation industry, from the translation process research community as well as translation system designer to examine Expertise in Translation and Post-editing from different possible angles.

Program, for Friday August 17 and Saturday, August 18

Abstracts:

2. Balling & Carl: The CRITT database of TPR data
3. Blücher: Post-editing lessons learned
4. Cattelan: MateCat - Combining Machine Translation and Human Translation
6. Elming & Bonk: The CASMACAT workbench: a tool for investigating the integration of technology in translation
7. Ferreira: L1 and L2 translation: an investigation of the interface between expertise and recursiveness
8. Goens: Implementing machine translation in the translation process - an LSP point of view
10. Kooponen & Salmi: Does prior use of machine translation systems help in post-editing?
11. Kumar: Quality Control in Translation: Indian Perspective
12. Lehr: The impact of positive and negative affect on performance in translation - a challenge to expertise?
13. Mesa-Lao: Translating vs Post-editing: A pilot study on eye movement behaviour across source texts
14. Moran: Experiences instrumenting and open-source CAT tool to record translator interactions
15. Rico & Diez Orzas: EDI-TA: Addressing Real Challenges in Post-editing
16. Sagara: C3 (Crowd Computer Collaboration) for Translation
18. Sekino: Investigating conceptual and procedural encodings in manual translation and in post-editing processes from Japanese to Portuguese
19. Silva: Post-editing integration in a translation agency workflow
20. Sinha: Smart Post-editing Interface for Hindi and other India Languages
22. Specia & Torres: Post-Editing practice – a survey of translators' preferences and concerns
24. Teixera: Knowledge of provenance and its effects on the translation process
25. Torral & al: Retraining Machine Translation with Post-edits to Increase Post-editing Productivity in Content Management Systems
26. Vieira & Specia: A wiki for translators and LSPs
<table>
<thead>
<tr>
<th>Times</th>
<th>9</th>
<th>FRIDAY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Welcome + Special keynote</td>
<td>Estimating post-editing effort - state-of-the-art systems and open issues</td>
</tr>
<tr>
<td>10</td>
<td>Mesa-Lao</td>
<td>Translating vs Post-editing: A pilot study on eye movement behaviour across source texts</td>
</tr>
<tr>
<td>10.20</td>
<td>Koglin</td>
<td>Cognitive Effort in post-editing and manual translation</td>
</tr>
<tr>
<td>10.40</td>
<td></td>
<td>COFFEE</td>
</tr>
<tr>
<td>11.00</td>
<td>Teixera</td>
<td>Knowledge of provenance and its effects on the translation process</td>
</tr>
<tr>
<td>11.20</td>
<td>Kooponen &amp; Salmi</td>
<td>Does prior use of machine translation systems help in post-editing?</td>
</tr>
<tr>
<td>11.40</td>
<td>Sekino</td>
<td>Investigating conceptual and procedural encodings in manual translation and in post-editing processes from Japanese to Portuguese</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>LUNCH</td>
</tr>
<tr>
<td>13</td>
<td>Alabau &amp; Casacuberta</td>
<td>Study of Electronic Pen Commands for Interactive-Predictive Machine Translation</td>
</tr>
<tr>
<td>13.20</td>
<td>Elming &amp; Bonk</td>
<td>The CASMACAT workbench: a tool for investigating the integration of technology in translation</td>
</tr>
<tr>
<td>13.40</td>
<td>Moran &amp; Lewis</td>
<td>Experiences instrumenting and open-source CAT tool to record translator interactions</td>
</tr>
<tr>
<td>14.00</td>
<td>Torral &amp; al</td>
<td>Retraining Machine Translation with Post-edits to Increase Post-editing Productivity in Content Management Systems</td>
</tr>
<tr>
<td>14.20</td>
<td>Goens</td>
<td>Implementing machine translation in the translation process - an LSP point of view</td>
</tr>
<tr>
<td>14.40</td>
<td></td>
<td>COFFEE &amp; DEMO</td>
</tr>
<tr>
<td>15.40</td>
<td>Lehr</td>
<td>The impact of positive and negative affect on performance in translation - a challenge to expertise?</td>
</tr>
<tr>
<td>16.00</td>
<td>Ferreira</td>
<td>L1 and L2 translation: an investigation of the interface between expertise and recursiveness</td>
</tr>
<tr>
<td>Time</td>
<td>Speaker</td>
<td>Topic</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>16.40</td>
<td>Balling &amp; Carl</td>
<td>The CRITT database of TPR data</td>
</tr>
<tr>
<td>17.00</td>
<td></td>
<td>END</td>
</tr>
<tr>
<td>19.30</td>
<td></td>
<td>WORKSHOP DINNER</td>
</tr>
</tbody>
</table>

**SATURDAY**

<table>
<thead>
<tr>
<th>Times</th>
<th>Speaker</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Mike Dillinger</td>
<td>Expert (Post-)Editing: The Client's Needs, The Researcher's Goals</td>
</tr>
<tr>
<td>10</td>
<td>Specia &amp; Torres</td>
<td>Post-Editing: Translators Preferences and Issues</td>
</tr>
<tr>
<td>10.20</td>
<td>Blücher</td>
<td>Post-editing lessons learned</td>
</tr>
<tr>
<td>10.40</td>
<td></td>
<td>COFFEE</td>
</tr>
<tr>
<td>11.00</td>
<td>Tarvi</td>
<td>From Computer-Assisted Translation to Computer-Assisted Translation Assessment</td>
</tr>
<tr>
<td>11.20</td>
<td>Vieira &amp; Specia</td>
<td>A wiki for translators and LSPs</td>
</tr>
<tr>
<td>11.40</td>
<td>Rico &amp; Diez Orzas</td>
<td>EDI-TA: Addressing Real Challenges in Post-editing</td>
</tr>
<tr>
<td>12</td>
<td>Kumar</td>
<td>Quality Control in Translation: Indian Perspective</td>
</tr>
<tr>
<td>12.20</td>
<td></td>
<td>LUNCH</td>
</tr>
<tr>
<td>13.20</td>
<td>Sagara</td>
<td>C3 (Crowd Computer Collaboration) for Translation</td>
</tr>
<tr>
<td>13.40</td>
<td>Cattelan</td>
<td>MateCat - Combining Machine Translation and Human Translation</td>
</tr>
<tr>
<td>14.00</td>
<td>Silva</td>
<td>Post-editing integration in a translation agency workflow</td>
</tr>
<tr>
<td>14.20</td>
<td></td>
<td>END OF WORKSHOP</td>
</tr>
</tbody>
</table>
Abstracts:

Study of Electronic Pen Commands for Interactive-Predictive Machine Translation

Vicent Alabau (valabau@iti.upv.es)
Francisco Casacuberta (fcn@iti.upv.es)
Institut Tecnològic d’Informatica, Universitat Politècnica de València,
Camino de Vera s/n, 46022 València, Spain

Typically, the post-editing of a machine translation (MT) output consists in performing a series of editing operations (i.e., replace, delete, insert or move pieces of text) in a specific text editor using the keyboard and occasionally the mouse. This approach has been proved to be efficient by the translation industry to the point that [1] proposes post-editing guidelines for translation agencies. However, the user needs to be in front of a desktop computer which imposes some restrictions regarding where and how the work is to be done. Laptop computers can also be used, although arguably performance could be diminished because of the use of uncomfortable laptop keyboards and track pads.

In this work, we envision an alternative scenario in which the user can use a touch screen or an electronic pen (e-pen) to perform post-editing tasks. Although e-pen interaction may sound impractical for texts that need a large amount of post-editing, there is a number of circumstances where it can be more comfortable. First, it can be well suited for post-editing sentences with few errors, as it is the case of sentences with high fuzzy matches, or the revision of human post-edited sentences. Second, it would allow to perform such tasks while commuting, traveling or sitting comfortably on the couch in the living room.

There is already a ‘de facto’ standard for gestures for proof reading (cf. Figure 1) from which we have extracted the most promising gestures: substitutions, deletions, insertions and, transpositions. Furthermore, we have added a shift gesture to move phrases to specific places in the text (i.e., the user circles the phrase and draws an arrow to the final destination). Then, we have studied two e-pen post-editing approaches. In the first one, we consider substitutions, deletions, insertions and, shifts. The number of these operations to obtain a reference can be computed with the translation error rate (TER) [2]. In the second approach, we assume that the user is working with an interactive-predictive MT system (IMT) [3]. In IMT, the user and the MT system collaborate to produce a high-quality output. The user locates the first error from left-to-right and amends it. Then, leveraging the recently validated text, the system reformulates (predicts) the continuation of the translation aiming to improve the previous hypothesis. In this case, we have also considered transpositions. To know what gestures could be more useful, we have conducted an experiment on the Xerox corpus [4]. The Xerox corpus consists of a collection of technical manuals. It consists of 56k sentences of training and a development and test sets of 1.1k sentences. Test perplexities for Spanish and English are 35 and 51, respectively. The summary of the edit rate results is displayed in
Table 1. The edit rate is the number of edit operations needed to obtain the reference normalized by the number of words. We can see that the IMT system requires less interactions, especially for en. Next, the number of times a particular edit operation has been applied is shown. We expect the gestures for deletion, insertion, shifting and transposition to be easy to tell apart for a machine learning algorithm. However, this will be the subject of future work. In addition, substitutions or insertions require the user to write the correct word, which can be done with a virtual keyboard or by handwriting [5]. The perplexities for these words is 336 for English and 242 for Spanish, whereas the errors rates for handwriting recognition are 7.4 for English and 8.9 for Spanish.

References
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>🖊️</td>
<td>insert a comma</td>
<td>The mayor's brother, I tell you, is a crock.</td>
</tr>
<tr>
<td>📁</td>
<td>apostrophe or single quotation mark</td>
<td>I wouldn't know where to put this vase.</td>
</tr>
<tr>
<td>🔄</td>
<td>insert something</td>
<td>I know it in fact, everyone knows it.</td>
</tr>
<tr>
<td>🙁 🙁</td>
<td>use double quotation marks</td>
<td>May favorite poem is Design.</td>
</tr>
<tr>
<td>⚪️</td>
<td>use a period here</td>
<td>This is a declarative sentence.</td>
</tr>
<tr>
<td>❌</td>
<td>delete</td>
<td>The elephant's trunk is really its nose.</td>
</tr>
<tr>
<td>⤷</td>
<td>transpose elements</td>
<td>He only picked the one he likes.</td>
</tr>
<tr>
<td>⚪️</td>
<td>close up this space</td>
<td>Jordan lost his favorite basket.</td>
</tr>
<tr>
<td>🌈</td>
<td>a space needed here</td>
<td>I have only three friends: Ted, Raoul, and Alice.</td>
</tr>
<tr>
<td>🏷️</td>
<td>begin new paragraph</td>
<td>&quot;I know it,&quot; I said.  &quot;I thought so,&quot; she replied.</td>
</tr>
<tr>
<td>🎁</td>
<td>no paragraph</td>
<td>&quot;I knew it, she said. No. He's no good.&quot;</td>
</tr>
<tr>
<td>🔃</td>
<td>lowercase</td>
<td>Lunch was delicious.</td>
</tr>
<tr>
<td></td>
<td>=</td>
<td>capitalize</td>
</tr>
</tbody>
</table>

Figure 1: List of ‘de facto’ standard proof reading symbols obtained from [editorwriter.org](http://editorwriter.org)
<table>
<thead>
<tr>
<th></th>
<th>post-editing</th>
<th></th>
<th>IMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>en-es</td>
<td>es-en</td>
<td>en-es</td>
</tr>
<tr>
<td>edit rate (%)</td>
<td>21.3</td>
<td>24.4</td>
<td>21.1</td>
</tr>
<tr>
<td>substitutions</td>
<td>1028</td>
<td>919</td>
<td>1549</td>
</tr>
<tr>
<td>insertions</td>
<td>325</td>
<td>461</td>
<td>190</td>
</tr>
<tr>
<td>deletions</td>
<td>484</td>
<td>302</td>
<td>0</td>
</tr>
<tr>
<td>transpositions</td>
<td>–</td>
<td>–</td>
<td>41</td>
</tr>
<tr>
<td>shifts</td>
<td>319</td>
<td>357</td>
<td>347</td>
</tr>
</tbody>
</table>

Table 1: Summary of number of edit operations needed to obtain the reference for post-editing and interactive-predictive machine translation. The edit rate is the ratio between the number of edit operations and the number of words in the reference. Follows the number of occurrences for each edit operation. Here, we assume a perfect gesture recognizer. The gesture recognizer will be developed in future work.
The CRITT database of TPR data

Laura Winther Balling & Michael Carl
Copenhagen Business School

We introduce the CRITT database of translation process data which we are currently building as a publicly available database for translation process research (TPR). The database contains data from experiments conducted over the last five years at the CRITT centre at Copenhagen Business School and other places, and is used for Translation Process Research.

The data is stored in a consistent format and includes both process and product data. The process data are data about keystrokes obtained with Translog - or other process data collection devices - as well as (where available) eye-tracking data. The product data include tokenized and aligned versions of the source and target texts and their annotations in the Treex format. Additionally, we include a whole range of characteristics of the texts, languages, translation settings, and translators, which may be used as predictors in large-scale statistical investigations of the translation process. The data has the potential to give us tremendous insight into the similarities and differences of translators’ performance, within and between different languages and texts. In addition, a main objective is to be able to investigate different translation settings, comparing from-scratch translation with post-editing (where the source text is available) and editing (where the source text is not available).

In addition to introducing the database, we will outline an analysis of production data from a series of experiments where six English texts were translated into as different languages as Danish, German, Spanish, Hindi and Chinese. The analyses will show how rich a resource a publicly available database will be for the study of the translation process.
Post-editing lessons learned

Mads Blücher
Independent Researcher

The following is based on hands-on experience with early publicly available S-MT technology and online CAT tools within a translation workflow.

Post-Editing
At LanguageWire we have only worked with MT offered by Google. This was done in early stages with a simple test on a word document run through their website and afterwards post-edited. Naturally, our translators did not consider this to be an efficient solution. When we tried the same test within our own web-based CAT tool based on single segments, we saw an enormous change in use. As long as the user could start with an empty segment and actively ask for the MT result, users were very happy with the Google MT. The first reason was that a lot of the simple sentences could easily be reused without modification, but as sentence complexity increased, MT was not useful for the complete sentence but was still very helpful with terminology. One comment we heard a lot was that post-editing was not an option when working with Google, as too many segments had very bad structure. Translators commented that they could write a correct sentence faster than it took them to correct a bad MT result. We called it “use or throw away”.

When we activated the service the first reports were very negative, but over time the criticism died out. After a year with Google connected to our CAT tool, we had to remove it due to legal issues with some clients. The day after we removed the Google MT-API my inbox was filled with messages from frustrated translators, all asking where the MT results had gone.

We realized then that the service has a massive impact on the way that the translators worked and therefore also on the price that we could pay for the translation.

General vs. domain
So, based on these lessons learned, our thoughts on MT are that it offers advantages as long as it is easy-to-use for all clients and easily administrated. Building domain-specific MT makes sense if you have one big client, but as we specialize in SME and have to handle several hundred clients with different domains on a daily basis, there is not enough data to train MT for one client.
Achievements with the so-called statistical machine translation approach have increased expectations and created new opportunities in the translation market. On the one hand, machine translation is seen as an opportunity because it has the potential to open up new markets for language service providers (translation of user generated content, low value translations, etc.). On the other hand, it can help in reducing the costs of the services offered to translation buyers. However, machine translation has not yet proven to be a disruptive factor for the professional translation industry. Research on machine translation has mostly focused on delivering ready-to-use translations rather than supporting and improving the professional translation workflow. So far, statistical MT is mainly trained with the objective of creating the most comprehensible output for a final user, rather than outputs that minimize the effort of a human translator.

Over the past 20 years, the translation industry has adopted and continued to use the same processes based on Computer Assisted Translation tools which make use of translation memory technology extending the capabilities of professional translators (memory) and allowing them to reuse previously translated content. It has proven useful in reducing the costs and the turnaround times and, arguably, in improving the translation quality and consistency. Such a technology, however, is not of much use when it comes to translating new content nor can it further speed up the translation process. Translators still struggle to meet the generally approved turnaround of 2,500 words per day, translation memories are still only based on manually translated content and translation buyers are still missing out on opportunities to increase the value for money they can get from translation.

MateCat sets out to fill the gap between machine translation technology and the translation industry’s needs and processes. It integrates statistical machine translation and collaborative translation memories, within the human translation workflow. MateCat’s goal is to define new operating conditions for statistical MT in order to better match the typical workflow and functional requirements of those using CAT tools. It aims at defining an MT system whose output minimises the time the translator needs to post-edit it. It will advance the state-of-the-art by making MT technology aware of how it is used, attuning itself to the domain, adapting to the corrections and implicit feedback from the user, and providing useful information to the user.
Expert (Post-)Editing: The Client's Needs, The Researcher's Goals

Mike Dillinger

Clients in global enterprises have a quickly growing need for expertise in "post-editing". But what exactly - which expertise - do they need? How will they know that they got what they paid for? These clients are not at all sure of the answers, and industry players have not helped. In this talk, I'll synthesize the point of view of clients who need help with large-scale translation based on MT and the point of view of researchers in cognitive science who want to understand the nature of this newly-important skill. There are interesting similarities and differences in how these two disparate groups define, measure, train, deliver, and support expert (post-)editing.
We present the CASMACAT workbench which is a tool for analyzing the process of translation. The tool is especially useful for examining the effects of integrating technology in the translation process, especially how it affects the behavior of the human translator. Key features of the tool are: web-based technology, extensive logging of user behavior including eye tracking, and exact replay of the translation session.

The CASMACAT Workbench builds on experience from the Translog tool (Lykke Jakobsen, 1999), which is a tool for studying reading and writing processes e.g. in translation. The main advances of the tool are 1) that it is using web-based technology which allows for easier portability across different machine platforms and versions, 2) it gives a much more realistic translation session by both visually and functionally resembling commercial translation tools, and 3) it allows for direct integration of technologies such as (interactive) machine translation. 1) and 3) are properties it shares with e.g. the Caitra tool (Koehn, 2009), but this does not allow for the same detailed analysis of the user behavior.

The CASMACAT Workbench is basically a web page where the subject logs in using an assigned user name and password. This allows for easy control over subjects, their profiles, the translation tasks they are assigned, and records of their translation sessions. The tool provides different layout options such as two columns with source segments on the left aligned to translation segments on the right, or one column with already translated segments above the current segment and future source segments below. Shortcut keys are used for functions such as navigating between segments. The translation field can be pre-filled by machine translation through a server connection and also automatically updated online from an interactive machine translation server. Another possibility is to display special assistance features e.g. information extracted from machine translation.

The CASMACAT Workbench opens for the possibility of both qualitative and quantitative analysis of translator behavior. The replay function allows the researcher to visually gain insight into the choices made by the translator during the session, and the extensive log file containing entries of all events that have occurred in the session, can be used to analyze and model the translation process at a higher level.

References
This paper aims at examining the extent to which expertise in translation can be modeled in terms of directionality – L1 (from English into Brazilian Portuguese) and L2 (from Brazilian Portuguese into English). More specifically, it investigates how expertise in L1 and L2 translation can be explained by means of Translog 2006 linear representations using recursiveness as a discrete variable, i.e., as an empirical indicator of adaptive behavior when a translator faces a translation difficulty (Buchweitz & Alves, 2006; Ferreira, 2010). During the production of the target text, the translator goes back numerous times to specific parts of the text in an attempt to come up with a more concise and coherent target text. This type of online revision is registered by means of keylogged files as the participant moves through the text and the 'revision keystrokes' (e.g., delete, backspace) are produced. In the present study, eight professional translators carried out L1 and L2 translation tasks. Data were collected at two different experiments. In data collection 1 (DC1), participants translated two subject-related scientific texts in English and in Portuguese, both about sickle cell disease. The task order was shifted randomly among subjects to control for a likely facilitating effect. In data collection 2 (DC2), participants carried out L1 and L2 translations of two source texts about different topics but with similar textual structures mapped by Rhetorical Structure Theory (Mann & Thompson 1987). Source texts pairs in DC1 and DC2 had roughly the same number of words. Data collection was based on the triangulation paradigm used in translation process research (Jakobsen, 1999; Alves, 2001; 2003), and data analysis drew on the linear representations of recursive movements generated by Translog 2006. These recursive movements were calculated and classified in terms of task time duration, pause values and linguistic features observed during task segmentation. The results were correlated across the four tasks and suggest that professional translators adapted their behavior according to the task in hand. When comparing the number of recursive movements in L1 and L2 translation tasks, there was an increase in the number of online revisions in L1 translations regardless of task order in DC1, which might be related to the subjects’ greater capacity to monitor target text production in L1. In cases where the sickle cell text had first been read in the L1 and then translated into L2, a facilitating effect might be expected, which might induce more critical monitoring of the L1 target text. In DC2, on the other hand, where the two texts were on different topics, recursiveness increased significantly in L2 translations, corroborating Buchweitz & Alves’s (2006) results and suggesting that, in the absence of a facilitating effect, L2 translations required a higher number of recursive movements. The differences between subjects’ performance in DC1 and DC2 point to an interesting interpretation in terms of translation expertise. Results show that in DC1 task order and a likely facilitating effect increased the number of recursive movements in L1 translations in comparison to L2. As a whole, the group produced more recursive movements in L1 translation. Three of the four translators that carried out the L1 first produced more recursive movements in L1 than in L2 translation, and three...
of the four translators that carried out the L2 translation presented more recursiveness in L2 translation task, comparing to the L1 translation task. On the other hand, in DC2 all the eight participants carried out the L1 translation first and most of translators (six) produced more recursive movements in L2 in comparison to L1 translation. Together, they produced more recursive movements in L2 than in L1 translation in DC2. These results indicate that task conditions (DC1 vs. DC2) do have an impact on subjects’ performance and suggest that translation expertise and directionality may be closely related to how L1 and L2 translation tasks are carried out.

Keywords: Directionality in translation, Recursiveness, Professional translators

References


Implementing machine translation in the translation process
– an LSP point of view

Jurgen Goens
Technology & Operations Manager of the Xplanation group

Who has not at least once used Google Translate? Statistical machine translation services like Google Translate make intelligent guesses as to what an appropriate translation should be. But because the translations are generated by machines, the translations are far from perfect. Still, Google Translate has made people expect ‘usable’ translations to be produced in no time by these machine translation engines – with no human translators involved. In addition to raised expectations of quality and lead time, people are getting used to paying nothing or at least to paying less for translation services. But is open source machine translation truly a threat to human translators and language service providers?

Integration of machine translation systems in the translation process can actually improve both translation quality and turnaround times. As far back as 1993, Xplanation started developing rule-based machine translation engines. A couple of years ago, Xplanation started analysing the available open source statistical machine translation engines and comparing their quality. At the Expertise in Translation and Post-editing workshop, Jurgen Goens, Technology & Operations Manager of the Xplanation group, will give an insight into how Xplanation has successfully implemented machine translation systems in their translation workflow called Tstream®.
Cognitive effort in post-editing and manual translation:
a pilot study on metaphor interpretation

Arlene Koglin
Federal University of Minas Gerais

This study draws on Relevance Theory (Sperber & Wilson, 1986/1995) and its application to translation (Gutt, 2000) and to metaphor interpretation (Gibbs & Tendahl, 2006, 2008) to compare cognitive effort allocated to two tasks of metaphor post-editing and one task of metaphor translation. Specifically, we aim at investigating the impact of raw machine translation output on post-editing effort in two tasks of metaphor post-editing by analyzing differences between fixation count and total fixation duration in areas of interest (AOIs) in source and target texts. We hypothesize that a) the raw machine translation output will have a positive effect on cognitive effort and b) manual translation will require more cognitive effort than post-editing. In order to test these hypotheses, an experiment was carried out at the Laboratory for Experimentation in Translation (LETRA) using eye-tracking and key-logged data and retrospective think-aloud protocols. The study analyses process-driven data collected from four subjects while post-editing and from two subjects while translating. Both tasks were performed using the same source text, i.e., a short extract of a newspaper text about the Tea Party movement. To analyze the impact of raw machine translation output, the four subjects were asked to post-edit a Google machine translated output in Task 1 (T1) and to post-edit a Systran machine translated output in Task 2 (T2). To compare cognitive effort between post-editing and manual translation, the two subjects were asked to translate the same source text. For the purposes of this study, eye-tracking data related to total fixation duration and fixation count on the metaphor “The Tea Party Pork Binge” were analyzed. Data analysis shows that the cognitive effort allocated to metaphor post-editing is lower in T2 than in T1. These results provide indications that cognitive environment (Sperber & Wilson, 1986/1995) may be shown to have a positive effect on reducing cognitive effort when post-editing metaphors. However, contrary to what Krings (1994/2001), O’Brien (2006) and Carl et al (2011) have found, our results show that cognitive effort in post-editing is higher than in manual translation. Statistical analysis was performed using Mann-Whitney Test in order to ascertain significant differences between the two groups but no significance was found probably due to the small sample size. Subsequent analysis should be carried out to investigate whether the results will be the same when analyzing a larger number of metaphors and participants.

References


Does prior use of machine translation systems help in post-editing?

Maarit Koponen
University of Helsinki

&

Leena Salmi
University of Turku

The use of machine translations for post-editing by a translator is becoming increasingly common, and such post-editing work is likely to form an even larger share of many translators' tasks in the future (see e.g. Garcia 2010, Guerberof 2010). Some have suggested that post-editing may move in a direction similar to copyediting, where the post-editor would mostly make linguistic corrections and other improvements only consulting the source text if necessary – or even without the source text. In such a situation, the question becomes whether the post-editor is able to gather the correct meaning based on the machine translation alone. Prior studies on post-editing without source text (Koehn 2010, Callison-Burch et al. 2010) suggest that performance varies but some editors are able to achieve surprisingly good results. In addition to features related to the texts and MT systems, post-editing may also require specific post-editing skills. These skills, in turn, may increase with exposure to machine translated texts and familiarity with typical MT errors. Translator students, for example, may benefit from exercises related to machine translation post-editing.
To examine the question of whether prior exposure to machine-translated texts can help with post-editing, we present a post-editing task carried out during an introductory course on translation technology. In this task, students post-edited magazine articles machine translated from English into Finnish. To assess how well they would be able to gather the meaning based on the machine translation alone, they were shown only the machine translation without the source text. Additionally, they were given the option to not edit segments they considered incomprehensible. The students were also asked about how frequently they had used machine translations. The machine translated and the post-edited versions were then evaluated for correctness. In this presentation, we present results from the correctness evaluation and discuss connections observed between the frequency of prior use of MT systems, the perception of MT fluency and clarity as well as the success rate in the post-editing task. The results suggest that prior exposure to machine translated texts may affect perception of the usability of machine translation and also lead to better performance in post-editing.

References
Quality Control in Translation: Indian Perspective

Ravi Kumar
President, Indian Translators Association, New Delhi
ravi@modlingua.com

The communication explosion of the 90’s has helped in the rapid growth of translation and localization activities giving rise to the translation service industry where the expectations of consumers play a major role. It was noticed that in the absence of consumer oriented guidelines for translation services, on many occasions there had been a mismatch of assumptions and goals between the people requesting a translation and the people supplying the translation and finally affecting the project quality. Many times such mismatches had surprised the payer (consumers) and shocked the service provider (translators). This situation created the need to stipulate translation standards to protect and educate translation consumers and help the translator develop professionalism.

Exemplified by the German standard DIN 2345 on translation adopted in 1998, the development of national standards in different parts of the world gained importance specially in Europe, USA and Canada. On the other hand a highly diversified and multilingual country like India has failed to catch up with the growing importance of quality standards in Translation.

While, especially for technical text in Europe, translators and reviewers pay special attention to maintain word to word meaning of the translation (off course with due importance to adaptation and localization) to give high quality output through second review or even third party review or even reverse translation. On the other hand, In India, most of the translators (especially those from academic background) fail to deliver translation as per the international standards as they are tuned towards ruptantar or transcreation whereby they take liberty to mould the translated text as per the convenience and understanding/ misunderstanding of the text.

This creates new hurdle for the Project Manager as he / she fails to make the translator understand the requirement of the client, and he/she remains vulnerable to mismatch of expectations. My presentation will explain in detail the factors that many times create hurdles in understanding the concept of translation in India. Perhaps these factors are embedded in Indian society since ages, and careful understanding of the background can help get the best translation out of Indian diversity.
The impact of positive and negative affect on performance in translation – a challenge to expertise?

Caroline Lehr
University of Geneva

In the last decades research has increasingly come to acknowledge the importance of affect for any kind of cognitive functioning and human interaction. In translation studies and the investigation of the activity of translating, however, affective aspects have so far scarcely been addressed. Several translation scholars have claimed that affect and its role in translation deserve particular attention and that translation theories should not reduce human behavior to rational dimensions only (see Risku 2004:43). Nevertheless, the question if and to what extent positive and negative affect influence translation processes and translators’ performance still lacks empirical investigation. Against a background of psychological research suggesting that positive affect increases creativity, whereas negative affect increases accuracy (“Affect-as-information-model” Clore & Storbeck 2006, “Broaden and Build Model of Positive Emotions” Fredrickson & Branigan 2005), an online experiment with expert translators was conducted. In this experiment, affect was manipulated by bogus performance feedback on a prior translation before a second text was translated. Affective responses to the feedback, as well as self-evaluations and time spent translating were assessed. The translations were evaluated by expert graders for creativity and accuracy. Results show significant differences between the groups in subjective performance as well as in expert evaluations and partly confirm effects that were found in a pilot experiment with students. The paper will discuss the results of this study with regard to the concept of self-monitoring that has been held to be linked to expertise. The paper will also raise future research questions that should address how the influence of affect on translation performance changes or persists during the different stages of expertise and how affect may influence the acquisition of expertise.

References
Translating vs Post-Editing:
A pilot study on eye movement behaviour across source texts

Bartolomé Mesa-Lao
Universitat Autònoma de Barcelona, Spain
barto.mesa@uab.cat

New technologies are creating new translation workflows as well as new professional profiles. Post-editing is gradually becoming one of the most requested services in localisation as opposed to full human translation. Major language service providers now pre-translate source texts using existing translation memories and then automatically translate the remaining text using a machine-translation engine. This “hybrid” pre-translated text is then given to human translators to post-edit. Following guidelines the post-editors correct the output from machine translation to produce a target text with different levels of quality.

The main purpose of this pilot study is to explore the differences between translation and post-editing of texts through the analysis of user activity data. A group of ten professional translators translated and post-edited four different texts from English into Spanish while their eye movements were being tracked. Each participant translated two texts from scratch and post-edited two further texts using a first machine translation draft. Our aim and interest when comparing these two different modalities was ultimately to study the effects on eye movements when reading the same text for two different purposes, i.e. translation vs. post-editing. Research was devised so as to find out to what extent reading a source text while translating results in different degrees of visual attention in comparison with the attention devoted to it by the translator while post-editing a machine-generated translation of the same text.

Four different measures were registered during the translation process in order to make comparisons between reading for translation and reading for post-editing: 1) task time, 2) fixation frequency, 3) total gaze time duration, and 4) transitions across source and target areas on the monitor screen.

If differences were found between reading for translation and reading for post-editing, we would certainly have empirical data to start thinking about what the actual role played by the source text is in post-editing. Similarly, we could evaluate how much attention it deserves when designing computer-aided translation interfaces which integrate post-editing tasks as part of their translation workflow.

Preliminary results show significant differences in the way translators approach the source text when it comes to translating or post-editing it.
Experiences instrumenting an open-source CAT tool to record translator interactions

John Moran
Trinity College Dublin / Centre for Next Generation Localization
moranj3@cs.tcd.ie

iOmegaT is a customised version of OmegaT which has been instrumented in a manner which can be used to unobtrusively gather observational data on how a translator interacts with a computer-aided-translation (CAT) tool. The customisation logs a number of user interaction events in the CAT tool like post-editing time in a segment, keystrokes and concordancer use as well as context data, for example, terminology matches and spelling errors. Segment editing sessions and file editing sessions are recorded separately so translators may move and revisit segments and files at will. We have also reduced the number of menu items available in OmegaT to pre-configure and simplify OmegaT so that settings cannot be changed by a translator during a project and implemented a project download feature so that translators can get started more easily.

The XML format used to store the translator triggered events and translation context data can also be used to replay events and context data in a manner similar to the replay function in TransLog (http://www.translog.dk). However, unlike TransLog, the tool is suitable for use in large scale commercial translation projects involving translators located in different countries, where commercial CAT tools currently dominate.

As part of a commercial pilot project we have used the tool to gather machine translation post-editing temporal data in 12 European and Asian languages in a two-day translation task with two professional translators per language. The study was carried out to compare machine translation post-editing speed with human translation speed. The study is similar to one carried out by AutoDesk described by Plitt and Masselot in their paper entitled “A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context” published in “The Prague Bulletin of Mathematical Linguistics”, January 2010. Our study differed from theirs in that in our study translators worked in a CAT tool as opposed to a specially developed web page. Thus, along with accounting for inline formatting tags, translators in our study also had to repair fuzzy matches and take terminology matches into account. Our translators also had CAT tool features at their disposal like spell-checking and concordancing.

In this presentation I will describe iOmegaT and discuss in general terms why this approach to gathering translation process data is a promising means of evaluating the results of computational linguistic research and improving CAT tools for translators. I will also discuss the format we used to store translation process data using iOmegaT, which we intend to publish at a later date in an attempt to encourage commercial CAT tool vendors to enable the recording of similar translation process data in commercial translation projects.
EDI-TA: Addressing Real Challenges in Post-editing

Celia Rico (PhD.)
Universidad Europea de Madrid

&

Pedro L. Díez Orzas (PhD.)
Linguaserve

How do you go about implementing Post-editing (PE) processes in your company as an LSP? How does PE differ from reviewing TM fuzzy matches? What is the post-editor’s role and how can it fit in the company’s workflow? How is quality to be assessed? And productivity? Is it true that PE contributes to reducing costs?

These are some of the questions whose answers involve a challenge when implementing PE in a real scenario. Questions that Linguaserve, as a Language Service Provider, raised when setting up the context for PE in its translation workflow. In order to come up with adequate answers, the EDI-TA project was launched in March 2012, in agreement with Universidad Europea de Madrid, with three main objectives: (a) to define the functionalities for a post-editing tool, (b) to design a methodology for training post-editors, (c) to analyze the economic impact of implementing post-editing processes. EDI-TA has, then, a practical orientation, as a business oriented R&D project, and takes as its starting point TAUS/CNGL definition of PE as “the correction of machine-generated translation output to ensure it meets a level of quality negotiated in advance between client and post-editor”.

The project setting was then defined taking into account the company’s resources and translation workflow, along the following lines:

* MT output was produced by a rule-based system (Lucy Software)
* Language pairs: EN-ES, ES-EN, ES-CAT, ES-EU
* Text typology: Administrative and Financial
* PE environment: Transit

EDI-TA laid its groundwork following TAUS/CNGL guidelines (2012) and Allen’s proposals (2003). The work that derived from here is the object of the present paper, which will describe EDI-TA’s main features, empirical methodology and major findings. It will also report on its connection to the MultilingualWeb-LT Working Group (which receives funding from the European Commission --project name LT-Web-- through the Seventh Framework Programme (FP7) Grant Agreement No. 287815), serving the purpose of a test-bed of how metadata contributes to improving manual intervention in PE, and contributing to defining relevant metadata for PE purposes in a real scenario.
References
MultilingualWeb-LT Working Group: http://www.w3.org/International/multilingualweb/lt/
http://www.cngl.ie/node/2542
This presentation is about "Ryugakusei Network @ Minna no Hon'yaku," a crowd-source translation platform run by a start-up, Baobab, and its technology partner, NICT, which has successfully provided the best solution in practice in the translation market, i.e., quick and low-priced (one quarter the price compared with conventional translation suppliers). It is enabled by the concept of C³ (Crowd Computer Collaboration) for Translation. Our crowd builds the initial translations, which are fed to a statistical machine translation (SMT) system for translating new sentences. The SMT is embedded in the platform so the crowd can exploit the computer-made translation as suggestions. This collaboration has a synergetic effect which improves both the quality and efficiency. A flagship example is that the biggest apparel e-commerce site has adopted Baobab’s translation not only by crowd but also computer.

First, the crowd side of C³ has the following unique features:

(1) **Gamification functions.** Gamification improves the translator motivation by using computer game-style reward methods. For example, when a translator answers another translator's question, they receive a "point." These points accumulate into badges, then medals and then crowns. Higher-ranking translators are rewarded in the community month by month. This mechanism allows successful retention of translators in the pool.

(2) **Skippable job assignments** allows translators to skip sentences which are difficult for them to translate.

(3) Social communication functions, such as BBS between translators, provides translators with a cozy environment for communicating each other on translations, daily life and so on, and rescues translators from the loneliness of the translation process.

(4) **Powerful editing supports** such as an editor tailored for translation with a stress-free dictionary lookup and other reference materials, an easy-to-use terminology bank, and retrieval of parallel texts by approximate matching.

(5) **Additional computer supports**, such as sentences pre-translated by humans are shown with the translation, and an automatic translated result based on the accumulated parallel corpus is also provided.

(6) **"Big Brother" Quality control functions.** Translation quality is checked on-the-spot by senior translators. If a translator goes through three consecutive “BAD” marks, they will lose their job. If a translator shows proof of high-quality translations, however, they will get a higher salary.

Second, the computer side of C³ also has the following unique features:

(1) NICT provides Baobab with **Japanese data and processing for translation.**

(2) NICT provides a **top-notch statistical machine translation technology.**

The marriage of these two parts led the success of this crowd-source translation platform.
This is an ongoing study of the AuTema-PostEd\(^1\) project at University of Macau and reports one empirical investigation in which manual translation is compared with a process in which automatically translated texts are post-edited by twenty four translators from English (L2) into Chinese (L1) inspired by Carl et al (2011). The translations were performed using Translog, a tool for monitoring and collecting keystroke and gaze data, and without any kind of translation support. The data were analyzed through quantify translation process in an empirical user activity data (UAD) (Carl et al 2011). It is presented tools developed to investigate the translation process of the language pair Chinese and English: the Chinese Segmentator was utilized to tokenize the Chinese target texts, and the LexAligner to align the source and target texts, both tools developed by NLP2CT\(^2\) (University of Macau) in the environment of JDTAG (Carl 2009), and manually corrected. For the analysis of triangulation of data were utilized the translation progression graph (Carl 2009, Carl & Jakobsen 2009, Carl & Kay forthcoming) to visualize and describe patterns of keystrokes (target text production units) and patterns of gaze fixations (source and target texts fixation units), as well as the relationships and overlap between the two types of units. It is described the translation behavior with respect to the type of attention dispensed in each translation mode. The preliminary results indicate that the translation times were lower for the post-editing. Eyes and keyboards' movements also differ in these two modes of translations. It was found that almost all of the hand-operated translated texts were shorter than the source texts, while half of the post-edited translation texts were longer than the machine translation version. Although the individual differences of allocation of attention, the preliminary results suggest more alternating attention in post-editing mode than the manual translation mode. Due to the Chinese input system, it is suggested the optimal time to visualize the Chinese characters as 2ms. Forthcoming studies will regard to the quality of the translations produced. (332 words)

**Key words:** English-Chinese Translation Process, Manual Human Translation, Post-editing, Machine Translation.

**References:**
Carl, M., Kay, M. (Forthcoming) "Gazing and Typing activities during Translation: A comparative Study of Translation Units of Professional and Students Translators", in Meta.
Investigating conceptual and procedural encoding in manual translation from Japanese to Portuguese and in post-editing processes

Kyoko Sekino, PhD Student
(Supervisor: Dr. Fabio Alves, PhD)
Federal University of Minas Gerais, Brazil

This ongoing exploratory study aims at analyzing cognitive processes when subjects perform two comparable tasks, i.e., when translating an original text from Japanese into Portuguese and when post-editing machine translation output produced by Google Translate. Drawing on Wilson (2011), we assume that the conceptual-procedural distinction postulated by relevance theory entails a cognitive commitment. To the extent that it does, Alves & Gonçalves (in progress) have hypothesized that if most conceptual words also carry some procedural meaning, effort in translation should be greater when processing procedural encodings (transformation into linguistic form). Their results confirm their hypothesis that procedural encodings demand more processing effort from translators. In this paper, we test Alves & Gonçalves’s hypothesis in manual translations and in post-editing of machine translated output from Japanese into Portuguese. We expected to identify distinctive characteristics in the process of translation of two typologically distant languages. It is well known that Japanese has three writing systems, namely, Hiragana, Katakana, and Kanji (Chinese characters). Each system has its own function: Hiragana is simple syllabic writing for grammatical and conceptual words and Katakana, specifically for borrowings from foreign languages. Kanji, or Chinese characters, exclusively for conceptual words. Considering the three distinct writing systems and a marked structural difference (SOV in Japanese versus SVO in Portuguese), we assume that the processing of procedural encoding in the source texts will entail more effort by manual translators and post-editors alike. To assess subjects’ performance we used the key-logging software Translog and retrospective think-aloud protocols to investigate the performance of 6 subjects, namely 2 professional translators, 2 fluent Portuguese-Japanese bilinguals and 2 Japanese language students who have been familiarized with post-editing tasks. We based our analysis on the taxonomy proposed by Alves & Gonçalves, that is to mark; when (which stage of translation), what (lexical units which involve concepts or not) and what distance between micro TUs, observing the unfolding of micro and macro translation units. Our results for the Japanese-Portuguese language pair corroborate Alves & Gonçalves’s findings, showing that processing effort is greater when dealing with procedural encodings in both manual translations and post-editing tasks. The results also suggest that there may be more processing effort when subjects deal with Chinese characters, especially when they have to recognize the procedural characteristics of such characters even if they are content words.
Post-editing integration in a translation agency workflow

Roberto Silva
Information Security and Technology Manager

Machine translation is in good shape nowadays. Research activity is booming. There are many open source tools and a good number of free or paid applications available for non-professional end users and freelancers, and for all kind of devices. However, transfer from the research community to the language service provider sector is not happening at the same speed. Besides technical issues like implementation, integration and information security, machine translation post-editing is not well accepted or understood by the professional translation community. This is not something new: MT as a process and post-editing as a task have been around for quite a long time, and a large number of pages have been written about it. The new fact is that more and more companies and freelancers are adopting both in order to cut costs, meet tight deadlines, cope with the growing demand and new language pairs. Recent initiatives, developments and findings, along with today’s turbulent financial situation, have surely helped.

We present here the way a small sized language service provider has successfully integrated MT into its workflow, incorporating human translators who initially were not particularly keen on post-editing. We take a look at the post-editing task, both from the agency and the translator point of view. We will describe how MT has helped us to meet our clients’ requirements, the complexity of integrating it, its advantages and drawbacks, quality control, why it is so difficult to find or train good post-editors and analyze the effect on productivity and the translation cycle.

References
http://www.rafaelguzman.ie/articles/guzman/advancedautomaticMTpost-editing.pdf (This article was originally published in Multilingual Computing, #95, vol. 19, issue 3, pp. 52-57, April 2008)
GUZMAN, Rafael. 2008. Manual MT Post-editing: "if it's not broken, don't fix it!".
GUERBEROF, Ana. 2012. Productivity and quality in MT post-editing


Smart Post-editing Interface for Hindi and other India Languages

R. Mahesh K. Sinha
sinharmk@gmail.com

The paper presents details of a smart interface for post-editing Indian language texts as generated by AnglaBharati machine translation system. AnglaBharati MT system generates multiple translations as its possible output. It is based on the number of nested rules that get fired, the inherent unresolved polysemous lexicons, multiple interpretations of articles and quantifiers, unresolved insertion of post-positions, ambiguities in number, gender and person and variations in transliteration of named entities. A human post-editor has to make a selection out of the multiple choices generated by the system.

The smart interface designed for the human post-editing provides support for quick navigation of through possible errors. The human post-editor need not read the sentence word by word but is required to concentrate only on highlighted parts and the mouse over cursor gives a pull down menu of possible choices. Different colours are used to highlight different kinds of possible errors. In many of the Indian languages the verb form is as per the number, gender and person of the subject and sometimes dependent upon the direct object. Many a time with unknown subjects it remains unresolved. Plural forms are also used for honorific subjects that may not be apriori marked. Such navigation yields a very fast post-editing. Moreover, the system is made to learn about the corrections/choices made that improves the future translations of similar texts. Transliteration errors for named entities that get corrected are internalized and do not appear in future texts. Further there are some particles that are used in Hindi and other Indian languages whose position in the target text is not appropriated determined by the MT system. A different colour highlight on such particles and corresponding associated action of movements to the left or right positions provide another quick help to the post-editor.
Estimating post-editing effort – state-of-the-art systems and open issues

Lucia Specia
University of Sheffield, UK

Despite the arguable increasing adoption of Machine Translation (MT) by the translation industry, many factors prevent the use of MT to spread at a much faster pace. One such factor is the fact that very few translation solutions incorporate a mechanism to inform translators about the quality of the automatically translated text. It is well known that the quality of an MT system can vary significantly on a per-sentence basis and that this variation is dictated by complex conditions that go beyond simple and commonly used indicators such as sentence length. Confidence Estimation methods, and recently the more general Quality Estimation methods have been proposed to address this limitation. They aim at providing an estimate of the quality of translated documents, sentences, or even sub-sentential fragments. These methods have achieved satisfactory performance levels for different applications, such as selecting documents for publishing without any post-editing; highlighting translation errors to facilitate their correction; and filtering out cases that are not good enough for post-editing and that should be translated from scratch.

Focusing on the latter application -- the most popular way of using quality estimates --, in this talk I will present the main findings of a recent shared task for estimating post-editing effort organized at WMT-2012 (http://www.statmt.org/wmt12/quality-estimation-task.html). I will provide an overview of the state-of-the-art systems and discuss a few open issues in the area, including the complexity of the task of measuring post-editing effort (observed even for humans), and a series of “design decisions” yet to be agreed upon with respect to how quality estimates should be incorporated into computer aided translation workflows, such as:

• Whether (supposedly) bad quality translations should be ruled out or shown to translators with different scores or color codes (as it is the case with translations with low fuzzy match scores in translation memories);
• How to define a threshold on the estimated translation quality to decide what should be filtered out; and
• Whether translators prefer detailed estimates (sub-sentence level, similar to error detection) or an overall estimate for the complete sentence.
We recently conducted an online survey with 489 Translation students and professionals from 57 countries on the topic of translation technologies. This survey seeks to shed some light on the actual opinion of these users and to update the findings from previous studies, such as the 2009 SDL survey and the 2010 TAUS post-editing survey. In this talk, we will summarise some interesting findings on the respondents' preferences and concerns with respect to machine translation (MT) post-editing, including the following:

- 15% of 289 respondents reported being currently post-editing MT output, whereas 25% admitted that they had considered doing it. 32% do not post-edit due to quality concerns or because they do not consider it appropriate, necessary or worthwhile. 13% were unfamiliar with human post-editing of MT.

- When asked about post-editing tools, 47% of only 55 respondents who reported being currently post-editing use SDL Trados, 27% use the Google Translator Toolkit, 12% use Wordfast, 12% use OmegaT and 5% use Déjà Vu X2. 51% of 489 participants reported having heard of one or more post-editing tools, even though they were not using any tool.

- Only 39 participants who claim to practise post-editing indicated that they post-edit more statistical MT output (66%) than rule-based MT output (43%). 450 participants skipped this question. It is worth noticing that 50% of 67 participants claimed to use statistical MT, followed by those who used hybrid MT systems (32%), example-based MT (28%) and rule-based MT (19%).

- 55 participants said that this practice allows them to deliver translations in a faster way (47%) and at reduced costs (34%). 25% said that the main benefit of post-editing was the possibility of producing high-quality translations in a cheaper way.

- 46% of 171 respondents specified that post-editing accounted for less than 10% of their overall translation work, while 18% of respondents said that post-editing accounted for 11-25% of their total work.

- 70% of 182 informants prefer post-editing segments to be paid according to the time taken to edit a segment. 41% would like to have it paid according to the number of edits in the segment, and 21%
consider that the cognitive load involved in post-editing should be used as a criterion. With regard to the pricing rates, an hourly rate (61% of 176 informants) is preferred rather than a match rate (30%).

The majority of our participants were professional translators. In comparison with previous surveys, where participants were mainly language service providers, the use of MT and MT post-editing was found to be less prominent. Furthermore, the relatively low number of answers for the MT post-editing questions may show: a) a gap in the translation professionals’ technical knowledge on MT and MT post-editing, and b) the reluctance of professionals to adopt MT and, by extension, MT post-editing. According to some open-ended responses, translation professionals still see MT as a threat to their profession rather than helpful technology. Practical and up-to-date training on such MT and post-editing is still missing in most Translation courses. In addition, more user-friendly, efficient, and affordable translation technologies should be developed, taking into closer consideration the individuals' needs.
Translation product assessment is a major problem at any level of production since quality judgment is by nature subjective. The idea of quantifying assessment seems unattainable but, nonetheless, it has been used in comparative literary assessment (Tarvi 2004), developed later for the use in translation class (Tarvi 2011), and is now suggested as a wide-mesh tool in post-editing. The idea of computerizing assessment seems even more chimerical but the suggested quantitative method is, as remarked by some reviewers (Lengyel 2006), a ‘computer-predisposed’ ST-TT correlating technique.

The approach is rooted in the popular idea branded in various ways (e.g., Honkela et al, 2010), when translations are viewed as mappings between two languages through emergent conceptual spaces, which form a medium based on an intermediate level of representation. In its literary and teaching modes, the method is based on consecutive numbering of ST tokens followed by ST-TT token alignment or matching in the vein of, for instance, Dan Melamed (2001) or Lars Ahrenberg (2011). Both in human and computer revision, the Token Equivalence Method (TEM) allows one to obtain certain quantitative ‘markers,’ which seem to correlate well with the quality of translation products.

The method originated as a manual quantitative ST-TT comparative technique designed for assessment of the nineteen English translations of the classical Russian novel in verse Eugene Onegin by Pushkin in the PhD thesis “Comparative Translation Assessment: Quantifying Quality”. The figures obtained with the TEM were shown to totally agree with the available reported results obtained on the same material by other methods.

In its ‘teaching mode’, the TEM has been further elaborated to account for various textual parameters. Analysis can be carried out in a number (from one to six) of content and formal ‘frames’ – some obligatory, some optional, depending on the goals of instruction. The general quantitative result, Translation Quotient (TQ), may be viewed as the first computer-generated assessment grade.

A simplified version of the TEM, based on the use on Machine Readable Bilingual Dictionaries (MRBD) is offered for discussion as a post-editing assessment tool. In this modification, the tokens are counted but not numbered. As a result of source-target token alignment via MRBDs, the translation in question is characterized by certain quantitative parameters, like the number of translated tokens, grammar tokens and added tokens, which allow post-editors to pinpoint the strategies of revision. After post-editing, the same program is run on the revised version. The supposedly increased figures can be expected to reflect a better fit between the original and its translation.
Knowledge of provenance and its effects on the translation process in an integrated TM/MT environment

Carlos S. C. Teixeira
Intercultural Studies Group, Universitat Rovira i Virgili
Avda. Catalunya 35 – 43002 Tarragona, Spain
carlostx@linguanativa.com.br

Abstract. This paper will present preliminary empirical results gathered on the translation process of ten English-to-Spanish professional translators working in two different environments. The second aim of this paper is to discuss the different methods of data collection that have been used and the kinds of information they can provide on the translation process. The motivation behind the research design is the assumption that the way translation tools present metadata to translators may have an impact on performance indicators such as speed and effort. Building on the results of a previous pilot experiment, two competing environments that are often found in the translation industry have been reproduced: a visual or interactive environment vs. a blind or pre-translation environment. In both cases, translation proposals come from a translation memory or from a machine translation engine. Both environments were created in IBM TranslationManager. In the first environment, the software was configured so that translators could see the source text and a translation proposal for each segment, and the provenance information about each proposal (whether it comes from MT, TM, and at which match percentage) was displayed. In the second environment, all segments were populated with the best available proposal. Although translators could also see the source text of the current segment within the tool, they could not see any information about the provenance of each proposal. For the purposes of the experiment, two texts were taken from the same IBM software manual, each containing about 500 words. Machine translation proposals were generated by a customised, Moses-based engine. Translation-memory proposals were based on client-approved memories, and the non-exact matches were produced by editing either the translation-memory source/target segments or the to-be-translated source segments. Each translator was asked to translate the first text in one of the environments and the second text in the other environment, in order to allow for intra-subject comparisons. Translation proposals were equally divided into four types: exact matches, fuzzy matches between 85 and 99 percent, fuzzy matches between 70 and 84 percent, and MT matches. Each document had 28 segments in total, i.e. seven segments of each kind. For each segment, editing time and effort (the number of keystrokes and mouse clicks) were measured in relation with the total length of the segment. For each kind of translation proposal, results were compared between the two environments. Data was collected using screen recording, keystroke logging, eye-tracking and retrospective interviews. Special emphasis was placed on the ecological validity of the experimental set-up. To this end, a stand-alone eye tracker was used, allowing translators to work on their computer of habitual use, including their choice of screen monitor. Preliminary results indicate a great intersubject variation. Although data processing is still underway, it is visible that most translators processed exact matches significantly faster in the visual environment. However, it is still not possible to make definite conclusions on speed for the other kinds of proposals.
Interestingly, typing effort tends to be higher in the visual environment than in the blind environment.

References
Guerberof, Ana. 2012. Productivity and quality in MT post-editing. Available at:
Retraining Machine Translation with Post-edits to Increase Post-editing Productivity in Content Management Systems

Antonio Toral  
School of Computing, Dublin City University. Dublin, Ireland  
Leroy Finn & Dominic Jones  
School of Computer Science and Statistics, Trinity College. Dublin, Ireland  
Pavel Pecina  
Faculty of Mathematics and Physics, Charles University in Prague. Czech Republic  
David Lewis  
School of Computer Science and Statistics, Trinity College. Dublin, Ireland  
Declan Groves  
School of Computing, Dublin City University. Dublin, Ireland

The post-editing of machine translated content has proved to be more productive than translating from scratch in the localisation industry. A study carried out at Autodesk (http://translate.autodesk.com/productivity.html) shows that Machine Translation (MT) and post-editing of technical documentation by professional translators provides a sustained productivity increase across different languages (from 42% to 131%).

In this paper we combine MT, post-editing and MT retraining capabilities in a Drupal-based Content Management System (CMS). We aim to establish whether the post-editing of content can be used to improve the performance of MT systems and whether productivity increases with non-professional translators.

The work arises as an outcome of collaboration between the PANACEA project and the Centre for Next Generation Localization (CNGL). CNGL provides a Drupal-based CAT tool while PANACEA provides a web service for MT and a workflow for re-training MT systems, which is fed by the post-edits created in CMS-L10N (see Figure 1).
Figure 1. System architecture

Content is machine translated between Spanish and English using the MT web service developed by PANACEA. This MT system has been trained using general domain data (Europarl). Subsequently, non-expert translators post-edit the output. Translators are sourced using Crowdflower, a global paid collaboration platform. These translators are presented with the source sentence, translated sentence and asked to provide a third post-edited translation. Content is sourced from pages of Wikipedia related to the European Soccer Championships 2012.

The retraining service is invoked periodically, using the post-edited content to retrain and/or retune the MT system. This allows for the continual adaptation of the MT system. We evaluate the improvement provided by the MT system through this retraining. Our strategy for retraining and retuning follows that proposed by Pecina et al., (2012), which carries out domain-adaptation using data crawled from web, obtaining a substantial improvement of 49.5% in terms of relative BLEU over a baseline trained on generic domain data. Our experiment involves three cycles of post-editing and retraining (2,000 sentences each).

Translation Post-Editing Tools: A Wiki for Translation Professionals

Lucas N. Vieira
Newcastle University, UK
&
Lucia Specia
University of Sheffield, UK

The availability of translation tools aimed at aiding translators and increasing translation productivity has been constantly growing. From tools based solely on Translation Memory technology to tools that also allow the post-editing of Machine Translation, significant effort has been put into the field of Computer Assisted Translation in the past years. This is arguably of interest to translators and providers of translation services. However, as a result of the wide array of options currently available in the market, translators are often faced with the task of electing the tool or type of technology that best suits them. This also brings about the question of what features offered in these tools tend to be more useful and desirable to be incorporated into the translation process.

Having as background a critical review of translation tools from the perspective of a translator (Vieira and Specia, 2011), in this talk we will discuss the initiative of creating a “wiki” environment on the web as a platform where translation professionals can discuss and exchange information on the suitability of different tools and features for the task they have at hand, be it traditional translation revision, or Machine Translation post-editing.

A platform of this kind will serve not only as a constantly updated database on translation revision and post-editing tools, but also as a space where translators can point to current gaps in the market of language technologies. As such, we hope that it will be a valuable source of information to technology providers, who can build on translators’ feedback to (further) develop their products. While a number of forums and mailing lists dedicated to this type of discussion are already available, specific information on tools is scattered in messages posted by the users of these various spaces. A more general comparison between tools and overview on advantages and issues with such tools is still missing. We expect that the proposed wiki will address this gap by having structured information on tools gathered together in a single place, and available to all interested parties.

As a starting point, findings of the aforementioned critical review will be published for further discussion and updating. The review covers nine different tools, which are analysed based on features and functions that are currently offered and the ones that seem to be lacking. Special emphasis is given to functions that would be potentially desirable and that have not been found in any of the tools analysed, which provides a preliminary picture of aspects that still need to be improved.
References