

PET: A Standalone Tool for Assessing Machine Translation through Post-editing

Wilker Aziz
University of Wolverhampton
Stafford Street, WV1 1SB
Wolverhampton, UK
w.aziz@wlv.ac.uk

Lucia Specia
University of Sheffield
211 Portobello, S1 4DP
Sheffield, UK
l.specia@sheffield.ac.uk

Abstract

Machine translation (MT) post-editing is now a popular practice in the translation industry. It has been shown to allow translations to be produced at lower cost without a decrease in quality. The post-editing of automatic translations can provide useful information for MT researchers and developers to evaluate translations and systems. We describe a standalone tool that has two main purposes: allow the post-editing of translations from any MT system and collect sub-segment and segment level information from the post-editing process, e.g., detailed keystroke and time statistics.

1 Introduction

The editing of semi-automatic translations has been a common practice among users of Translation Memory (TM). TM tools such as SDL Trados¹ and Wordfast² provide user-friendly environments to aid human translators. The post-editing (PE) of Machine Translation (MT) output has only recently started to be more widely adopted as a way of incorporating MT into human translation workflows. Although a number of issues are yet to be addressed, such as an adequate pricing model for PE, this practice has been shown to minimise time and costs. A consequence of the widespread use of MT is the need for PE tools.

Modern TM tools incorporate MT systems with a common PE interface for both MT and TM, e.g., SDL Trados. Some MT systems also incorporate PE facilities, such as Google Translate³ and Systran⁴. However, these and other existing PE tools suffer from one or more of the following limitations, which we aim to address in this work:

- Restricted availability: most of them are proprietary tools only available as part of a major (more expensive) product. These tools generally do not allow the PE of a heterogeneous selection of translations from multiple MT systems.
- Lack of flexibility: these tools do not allow incorporating system- or task-specific functionalities, such as to limit the length of a post-edited segment.
- Limited logging: most tools do not collect explicit assessment nor detailed information about the post-editing process that could be used for measuring translation quality and diagnosing MT systems.

¹<http://www.trados.com/en/>

²<http://www.wordfast.net/>

³<http://translate.google.com/>

⁴<http://www.systran.co.uk/>

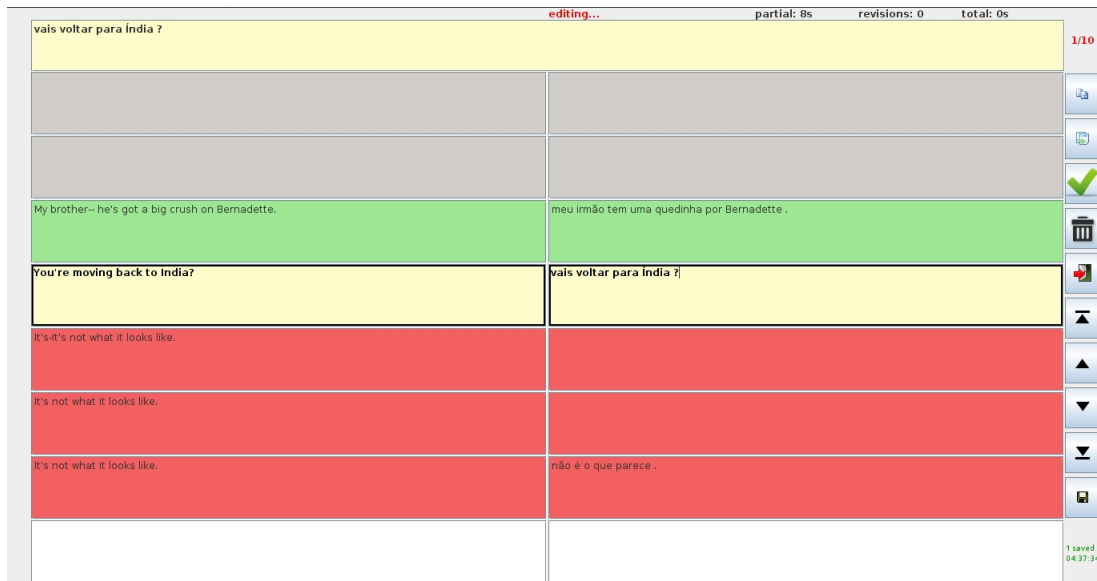


Figure 1: Annotation window

These limitations mostly constrain developers of translation technologies and researchers in machine (or computer-aided) translation. For a detailed study on translation tools that allow post-editing (e.g. Caitra,⁵ Lingotek,⁶ Déjà Vu X2,⁷ and OmegaT⁸) and requirements from the human translator’s perspective, we refer the reader to Vieira and Specia (2011).

We present PET (Post-Editing Tool) (Aziz et al., 2012a), a simple, freely available open-source standalone tool that allows the PE of any MT system and records various segment-level information. While PET is not yet a full-fledge tool for post-editing, offering limited built-in functionalities (dictionaries, etc.), it offers the flexibility that other tools lack (i) to enable easy design of post-editing tasks with specific requirements (such as constraints on the revisions produced in terms of length, word use, etc.), and (ii) to collect a number of (customisable) effort indicators and statistics on post-editing tasks.

2 Basic functioning

PET was developed using Java-6 libraries, hence it works on any platform running a Java Virtual Machine.

The interface displays source and target texts in two columns. Figure 1 shows the annotation window, where the left-hand side column is for the visualization of the source text and the right-hand side column enables the editing of its translation. In addition to post-editing, PET can be used to collect information about translation from scratch.

The unit of text to translate can be a sentence or a text of any length. Units to be translated/edited are grouped in “jobs”. A job may contain a mixture of units to translate or post-edit from one or more MT systems. Each unit is translated/edited at a time and navigation is achieved through the navigation bar on the right-hand side, or keyboard shortcuts. For the active unit, an extra text box at the top of the window can display alternative source sentences, alternative translations (produced by different MT/TM systems), past revisions, or reference (human) translations, where available.

⁵<http://www.caitra.org>

⁶<http://www.lingotek.com>

⁷<http://www.atril.com/en/software/deja-vu-x-professional>

⁸<http://www.omegat.org>

Units can also contain attributes that may or may not be made visible and may or may not add behaviour to the tool. For example, one can choose to show the number of characters used in the post-edited unit and even block edits above a certain length. New attributes can be provided as input files and their behaviour programmed through PET's API.

Using optional text boxes at the bottom of the window, the tool can also display external information for each active unit, for example, translation options that match words or phrases in such a unit (for source or translation sentences) from monolingual and bilingual dictionaries.

3 Translation quality indicators

Once a unit is completed, an (optional) assessment window can be displayed to collect explicit feedback about that unit (e.g., quality/effort scores), as the example shown in Figure 2. The number and type of assessments are configurable.

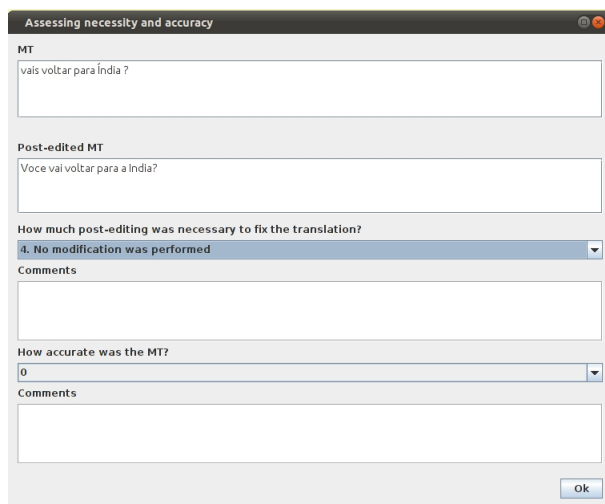


Figure 2: Assessment window

Besides these explicit quality indicators, PET provides a number of built-in implicit indicators, e.g.:

- Time spent editing a unit;
- Time spent assessing a unit;
- Counts of specific groups of keys, such as white keys, non-printable keys and non-white/printable keys;
- Timestamped edits (deletion, insertion, substitution, shift), i.e., words or phrases edited in the sentence and how much time each of these edits required;
- Number of revisions: how many times the annotator edited the unit;
- Edit distance between the draft translation and its post-edited version.

Input and output are human-readable XML-formatted files. PET comes with scripts to help generate and interpret them.

4 Input and output files

The input format for the tool is XML, which facilitates establishing new attributes. For example, a post-editing job can be defined by the following basic elements for each unit, as shown in Figure 3: type of job, unit identifier, input files with source (S) and reference (R) texts, and MT system that produced the translation and the actual translation.

```

<task type="pe" id="3">
  <S producer="xfiles.en">Excuse me.</S>
  <R producer="xfiles.pt">- Com licença,</R>
  <MT producer="google">Desculpe-me.</MT>
</task>

```

Figure 3: Extract of input file

The outcome of a task is organized as an annotation object per unit. This object contains the final translation, the implicit indicators obtained during the translation/editing and any explicit assessment. If PET is set to allow multiple edits of the same unit, every unit can have a list of annotation objects marked with revision stamps. This output is also stored in an XML file, such as the one in Figure 4. In this case, only the post-edited segment and time indicators are logged and the unit was edited only once.

```

- <task id="3" status="FINISHED" type="pe">
  <S producer="xfiles.en">Excuse me.</S>
  <R producer="xfiles.pt">- Com licença,</R>
  <MT producer="google">Desculpe-me.</MT>
- <annotations revisions="1">
  - <annotation r="1">
    <PE producer="pet">Desculpe-me. </PE>
    <indicator id="editing">3s</indicator>
    <indicator id="assessing">0s</indicator>
    <comment/>
  </annotation>
</annotations>
</task>

```

Figure 4: Extract of output file

5 Conclusions

We have presented a simple tool for post-editing and assessing translations that is MT system-independent and allows customisation at various levels, including the types of assessments that can be collected and restrictions on the post-editing process. The tool has already been used in different experiments, including (i) comparing different translation systems (Sankaran et al., 2012), (ii) contrasting post-editing and translation from scratch (Sousa et al., 2011), (iii) collecting information to build and compare quality estimation models (Specia, 2011), and (iv) measuring translation quality through post-editing for subtitles, where the tool dynamically restricts the length of each post-edited translation based on the length of the source segment and general time and space constraints for units Aziz et al. (2012b). A more detailed analysis on the use of information collected by the tool for measuring post-editing effort is presented in (Koponen et al., 2012).

PET is freely available for download at <http://pers-www.wlv.ac.uk/~in1676/pet>. Documentation, examples of input and output files, configuration files and additional code to produce input and process output files are also provided with the distribution.

References

- Aziz, W., de Sousa, S. C. M., and Specia, L. (2012a). PET: a Tool for Post-editing and Assessing Machine Translation. In *8th International Conference on Language Resources and Evaluation*, Istanbul, Turkey.
- Aziz, W., Sousa, S. C. M., and Specia, L. (2012b). Cross-lingual Sentence Compression for Subtitles. In *16th Annual Conference of the European Association for Machine Translation*, Trento, Italy.

- Koponen, M., Aziz, W., Ramos, L., and Specia, L. (2012). Post-editing time as a measure of cognitive effort . In *AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012)*, pages 11–20, San Diego, USA.
- Sankaran, B., Razmara, M., Farzindar, A., Khreich, W., Popowich, F., and Sarkar, A. (2012). Domain Adaptation Techniques for Machine Translation and their Evaluation in a Real-World Setting. In *25th Canadian Conference on Artificial Intelligence*, Toronto, Ontario.
- Sousa, S. C. M., Aziz, W., and Specia, L. (2011). Assessing the Post-editing Effort for Automatic and Semi-automatic Translations of DVD Subtitles. In *Recent Advances in Natural Language Processing Conference*, Hissar, Bulgaria.
- Specia, L. (2011). Exploiting Objective Annotations for Measuring Translation Post-editing Effort. In *15th Conference of the European Association for Machine Translation*, pages 73–80, Leuven.
- Vieira, L. and Specia, L. (2011). A Review of Machine Translation Tools from a Post-Editing Perspective. In *3rd Joint EM+/CNGL Workshop Bringing MT to the User: Research Meets Translators*, Luxembourg.