The ACCEPT Academic Portal: A User-centred Online Platform for Pre-editing and Post-editing

Asheesh Gulati **Pierrette Bouillon** **Johanna Gerlach**
Victoria Porro **Violeta Seretan**

Université de Genève FTI/TIM
40 Bvd. Du Pont-d'Arve, CH-1211 Genève 4, Suisse
{Asheesh.Gulatti, Pierrette.Bouillon, Johanna.Gerlach, Victoria.Porro, Violeta.Seretan}@unige.ch

Keywords: pre-editing, post-editing, machine translation, software tool, usability, academic purpose

Abstract

The advance of machine translation in the last years is placing new demands on professional translators. This entails new requirements on translation educational curricula at the university level and exacerbates the need for dedicated software for teaching students how to leverage the technologies involved in a machine translation workflow. In this paper, we introduce the ACCEPT Academic Portal, a user-centred online platform which implements a complete machine translation workflow and is specifically designed for teaching purposes. Its ultimate objective is to increase the understanding of pre-editing, post-editing and evaluation of machine translation. The platform is built around three main modules, the Pre-editing, Translation and Post-editing modules, and currently supports three language combinations: French > English, English > French and English > German. The pre-editing module provides checking resources to verify the compliance of the input data with automatic and interactive pre-editing rules. The translation module translates the raw and pre-edited version of the input text using a phrase-based Moses system, and highlights the differences between the two translations for easy identification of the impact of pre-editing on translation. The post-editing module allows users to improve translations by freely post-editing the text with the help of interactive and automatic rules. Finally, at the end of the workflow, a summary and statistics on the whole process are made available to users for evaluation and description purposes. Through its simple and user-friendly interface, as well as its pedagogically-motivated functionalities that enable experimentation, visual comparison, and documentation, this academic platform provides a unique tool to study the interactions between processes, and to assess the contribution of new technologies to translation.

1. INTRODUCTION

The increasing use of machine translation (MT) in the translation industry is placing new demands on professional translators and translation students. Understanding the advantages and limitations of the technologies involved in the machine translation workflow is an essential part of the translators' skill set. This entails new requirements on translation educational curricula, and, consequently, exacerbates the need for dedicated software for teaching students how these technologies may interact to offer a better final output.

In this paper, we introduce the ACCEPT Academic Portal (henceforth, AAP), a user-centred online platform which implements a complete machine translation workflow and is specifically designed for teaching purposes. This platform was developed in the framework of the ACCEPT European project, which was devoted to improving the automatic translation of user-generated content (www.accept-project.eu). The platform, publicly available at www.accept-portal.unige.ch, allows users to pre-edit a selected text with different types of Controlled Language rules (O'Brien, 2003) and directly evaluate the impact on translation quality and post-editing using metrics such as time and keystrokes. The ultimate objective of the AAP is to increase the understanding of the importance of pre-editing, post-editing and evaluation of MT. In the next section, we describe the context in which the AAP was developed and its specificities. In Section 3 and 4, we detail the design choices and the main functionalities of the platform.

2. BACKGROUND AND MOTIVATION

The AAP is an output of the ACCEPT European project, aimed at improving statistical machine translation (SMT) of user-generated content through minimally-intrusive pre-editing techniques, SMT improvement methods and post-editing strategies. The ACCEPT technology was originally developed as a series of demonstrators, plug-ins and APIs allowing integration into different web-based environments such as forums, portals or crowdsourcing platforms (Roturier et al., 2013; Seretan et al., 2014). The various software components are available on the freely-accessible online portal, www.accept-portal.eu.

To make the ACCEPT technology accessible to a wider public and, in particular, to teachers and students, we undertook the task of transforming the existing demo portal into an easy-to-use, fully-integrated platform combining pre-editing, MT and post-editing in a single workflow. While tools exist for each of these individual processes (e.g., PET [Aziz, et al., 2012], Casmacat [Alabau et al., 2013] and MateCat [Federico et al., 2014] for post-editing), to the best of our knowledge they have never been combined into a single platform. By enabling experimentation with the multiple processes involved in the MT workflow, the AAP provides a unique tool to study interactions between these processes. The next sections focus on the design choices and functionalities of the portal.

3. DESIGN CHOICES

The AAP is built using the JavaScript framework AngularJS. The platform offers a minimalistic and user-friendly interface that integrates and regroups the original plug-ins into a complete MT workflow, allowing users to subject a text to a sequence of processes until the desired output is reached.

The pre-editing module of the AAP relies on a jQuery plug-in that implements a lightweight version of the original ACCEPT "real-time" pre-editing plug-in (designed to function without an external dialog). The checking process uses the Acrolinx technology (Bredenkamp et al., 2000) and makes use of the pre-editing rules developed within the ACCEPT European project to improve the translatability of technical forum data (Gerlach et al., 2013; Gerlach, 2015; ACCEPT D2.2).

The translation module uses a phrase-based Moses system built specifically to deal with technical forum data in the framework of the ACCEPT European project. Training data includes translation memories supplied by the project partners, Europarl and news-commentary data, and a small corpus of forum data. More information about this system can be found in ACCEPT deliverables D4.1 and D4.2.

Finally, the post-editing module implements the ACCEPT European project post-editing client, described in detail in Roturier et al. (2013). This module integrates post-editing rules developed with the same technology as pre-editing rules.

4. FUNCTIONALITIES

The AAP is built around three main modules, the Pre-editing, Translation, and Post-editing modules, which can be activated individually. There are two additional components, the Start and Statistics pages. A help button is available in each of the modules and pages, giving access to online help which provides information about usage of the platform as well as underlying resources, such as the available pre-editing and post-editing rules. In this section, we briefly describe the components and modules that make up the AAP.

4.1. Start

The start page (Figure 1) allows users to select data and define the translation workflow. The workflow can be applied to different types of data:

- sample data (text from a technical forum post)
- text inserted in the inline editor
- uploaded text files

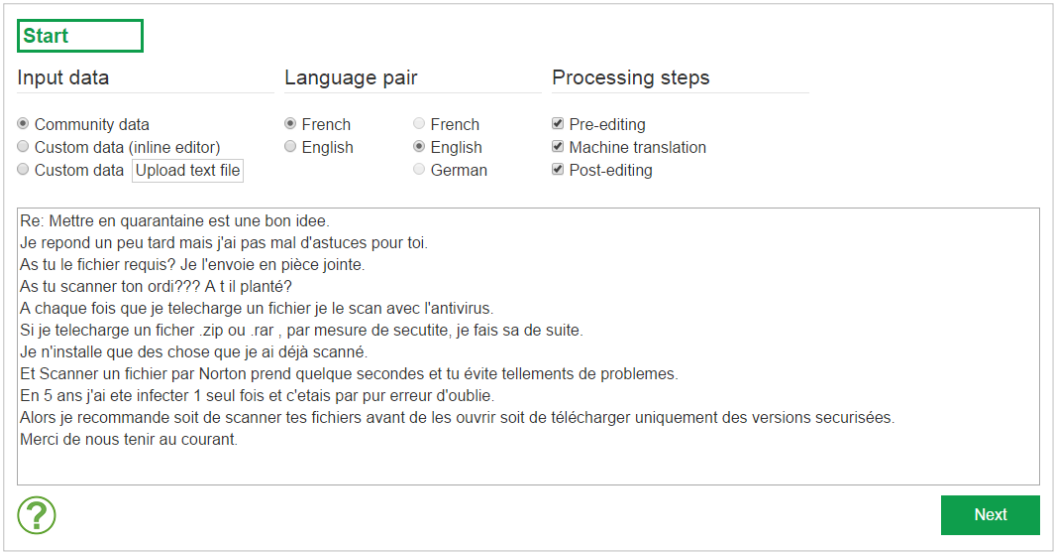

Figure 1: The ACCEPT Academic Portal – Screen capture of the Start page

The selection of the modules needs to be done at this stage. Different scenarios are possible:

- Pre-editing only
- Pre-editing and Machine Translation
- Pre-editing, Machine Translation and Post-editing
- Machine Translation
- Machine Translation and Post-editing
- Post-editing

Currently, three language pairs are available: French > English, English > French and English > German.

4.2. Pre-editing module

The pre-editing module (Figure 2) provides checking resources to verify the compliance of the input data with pre-editing rules. It allows the user to test interactive and automatic rules. Automatic rules may be either machine-oriented (called "silent" rules) or human-oriented. All rules can be activated individually. Besides applying rules, users can also edit the text manually.

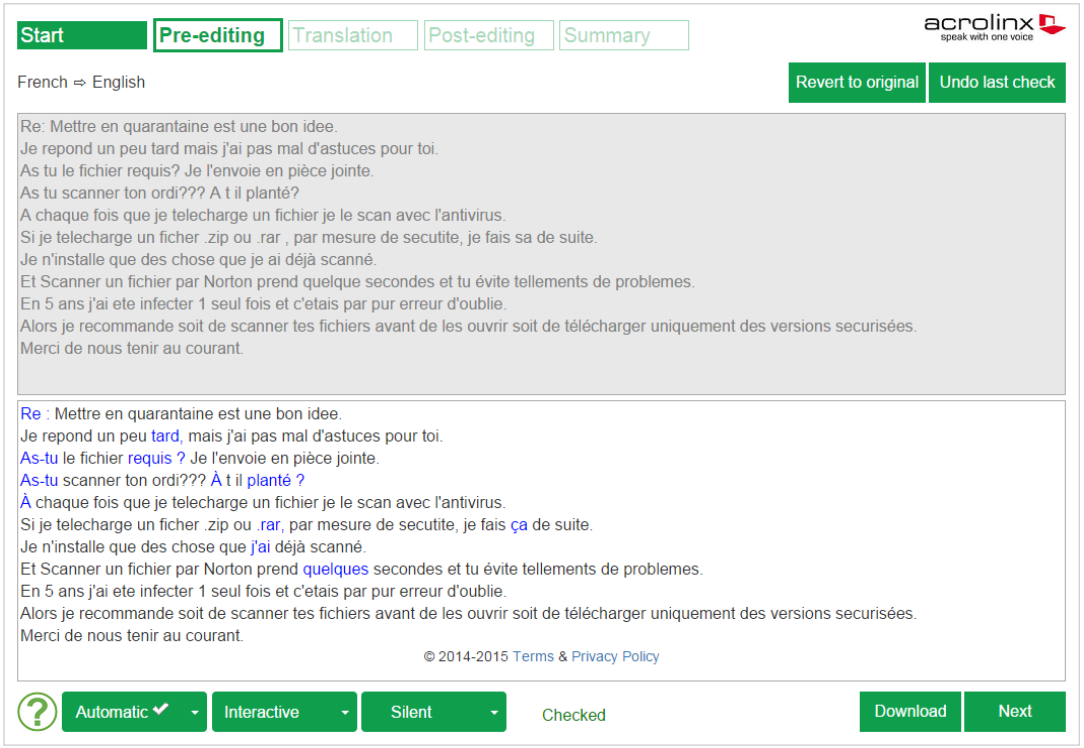

Figure 2: The ACCEPT Academic Portal – Screen capture of the Pre-editing module

4.3. Translation module

The translation module (Figure 3) translates the raw and pre-edited version of the input text using the ACCEPT SMT system developed during the European project. This module highlights the differences between the two translations for easy identification of the impact of pre-editing on translation. The user can select the version (raw or pre-edited) to be sent to the post-editing module.

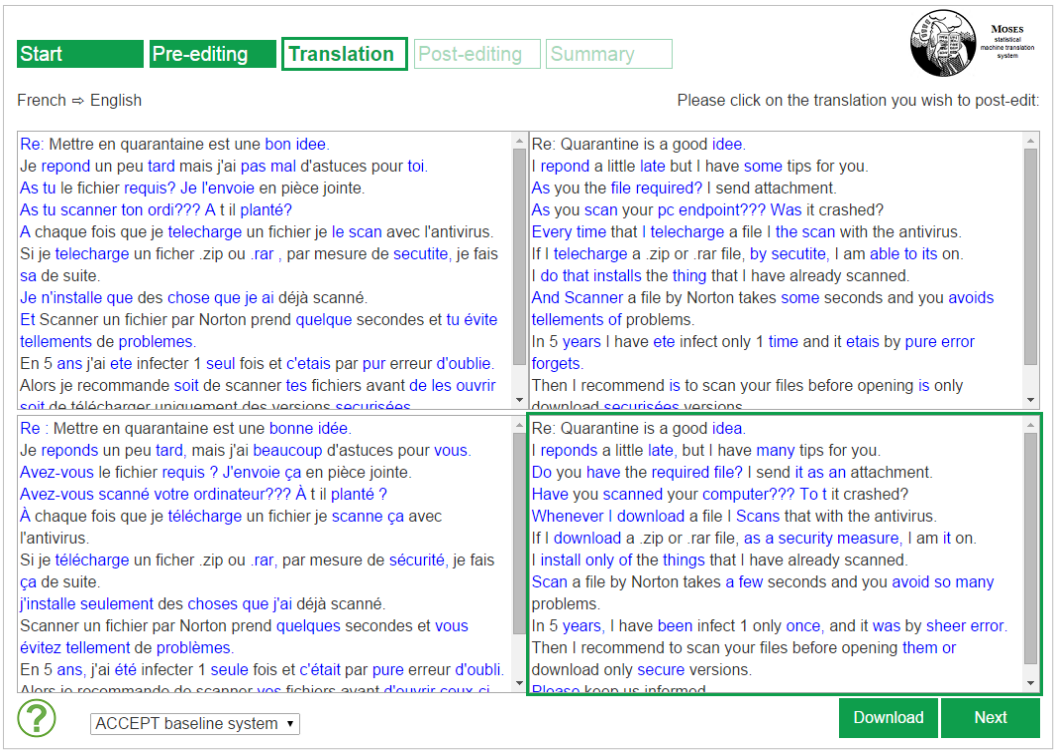

Figure 3: The ACCEPT Academic Portal – Screen capture of the Translation module

4.4. Post-editing module

The post-editing module (Figure 4) allows users to improve translations by freely post-editing the text. It also allows interactive checking with Acrolinx post-editing rules, specifically designed for correcting MT output (Porro et al., 2014; ACCEPT 2.4). The interface shows the source (for bilingual scenarios), the MT output, and the sentence currently being edited. When the main post-editing task is complete, the text undergoes a final revision phase. More precisely, a final check with spelling and grammar rules is performed, to ensure that no errors are left in the text.

The post-editing activity is recorded in an XLIFF file, for maximal interoperability (Roturier et al., 2013). This file contains detailed segment-level information on the actions performed during the post-editing process: for each revision, it reports keystrokes, editing time, number of accepted rule suggestions (if any), and user comments. The report can be exported at this stage or at the end, together with all session data in the Statistics page.

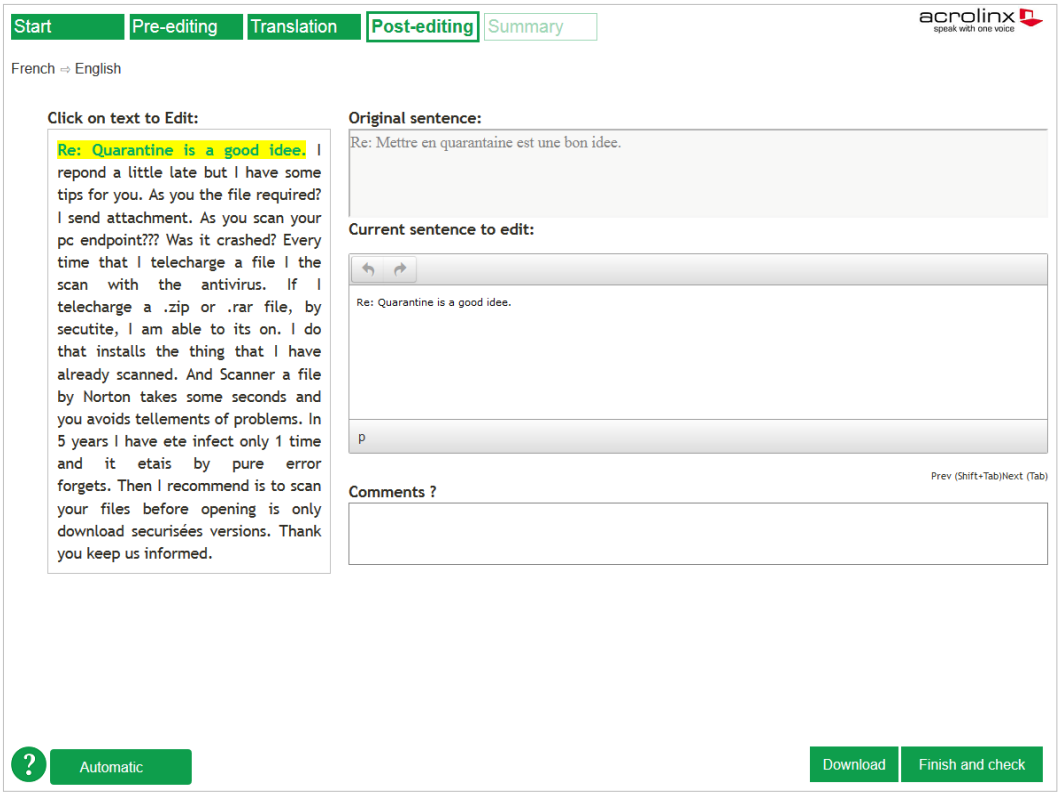

Figure 4: The ACCEPT Academic Portal – Screen capture of the Post-editing module

4.1. Statistics

At the end of the workflow, the Statistics page presents a summary of the entire process for evaluation or description purposes. This includes the individual steps that were performed and statistics about the post-editing activity. All versions of the text produced in each step (pre-edited version, chosen translation, final output) as well as the XLIFF report can be downloaded at this stage.

5. CONCLUSION

The ACCEPT Academic Portal integrates into a single platform the multiple processes involved in the MT workflow, from pre-editing to post-editing. Through its simple and user-friendly interface, as well as its pedagogically-motivated functionalities that enable experimentation, visual comparison, and documentation, the AAP provides a unique tool to study the interactions between processes, and to assess the contribution of new technologies to translation.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288769.

References

ACCEPT D 2.2 (2013). *Definition of pre-editing rules for English and French (final version)*, [online]. Available at:

<http://www.accept.unige.ch/Products/D2_2_Definition_of_Pre-Editing_Rules_for_English_and_French_with_appendixes.pdf>.

[Accessed 20 October 2014].

ACCEPT D 2.4 (2014). *Definition of post-editing rules for English, French, German and Japanese*, [online]. Available at:

<http://www.accept.unige.ch/Products/D-2-4-Definition-of-Post-editing-Rules.pdf>.
[Accessed 20 October 2014].

ACCEPT D 4.1 (2012). *Baseline machine translation system*, [online]. Available at:

<http://www.accept.unige.ch/Products/D_4_1_Baseline_MT_systems.pdf>.

[Accessed 20 October 2014].

ACCEPT D 4.2 (2013). *Report on robust machine translation: domain adaptation and linguistic back-off*, [online]. Available at:

<http://www.accept.unige.ch/Products/D_4_2_Report_on_robust_machine_translation_domain_adaptation_and_linguistic_back-off.pdf >.

[Accessed 20 October 2014].

ACCEPT D 5.6 (2013). *Browser-based client demonstrator and adapted post-editing environment and evaluation portal prototypes*, [online]. Available at:

<http://www.accept.unige.ch/Products/D_5_6_Browser-based_client_demonstrator_and_adapted_post-editing_environment_and_evaluation_portal_prototypes.pdf>.

[Accessed 20 October 2014].

Alabau, V., Bonk, R., Buck, C., Carl, M., Casacuberta, F., Garcia-Martinez, M., Gonzalez, J., Koehn, P., Leiva, L., Mesa-Lao, B., Ortiz, D., Saint-Amand, H., Sanchis, G. and Tsoukala, C., 2013. Casmacat: An open source workbench for advanced computer aided translation. *The Prague Bulletin of Mathematical Linguistics*, Volume 100, Pages 101–112.

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.

Bredenkamp, A., Crysmann, B. and Petrea, M., 2000. Looking for errors: A declarative formalism for resource-adaptive language checking. In: *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens, Greece.

Federico, M., Bertoldi, N., Cettolo, M., Negri, M., Turchi, M., Trombetti, M., Cattelan, A., Farina, A., Lupinetti, D., Martines, A., Massidda, A., Schwenk, H., Barrault, L., Blain, F., Koehn, P., Christian, B., and Germann, U., 2014. The Matecat tool. In: *Proceedings of COLING 2014*, Dublin, Ireland.

Gerlach, J., Porro, V., Bouillon, P. and Lehmann, S., 2013. La préédition avec des règles peu coûteuses, utile pour la TA statistique des forums ? In: *Actes de la 20e conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, Sables d'Olonne, France.

Gerlach, J., 2015. Improving Statistical Machine Translation of Informal Language: A Rule-based Pre-editing Approach for French Forums. PhD thesis, University of Geneva.

O'Brien, S., 2003. Controlling Controlled English. An analysis of several controlled language rule sets. In: Proceedings of EAMT-CLAW, Dublin, Ireland.

Porro, V., Gerlach, J., Bouillon, P. and Seretan, V., 2014. Rule-based automatic post-processing of SMT output to reduce human post-editing effort: a case study. In: *Proceedings of Translating and the Computer 36*, London, England.

Roturier, J., Mitchell, L., Silva, D., 2013. The ACCEPT post-editing environment: A flexible and customisable online tool to perform and analyse machine translation post-editing. In: *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice*, Nice, France.

Seretan, V., Roturier, J., Silva, D. and Bouillon, P., 2014. The ACCEPT Portal: An online framework for the pre-editing and post-editing of user-generated content. In: *Proceedings of the Workshop on Humans and Computer-Assisted Translation*, Gothenburg, Sweden.