APE-QUEST: a Quality Gate for Routing MT

Joachim Van den Bogaert, Heidi Depraetere, Sara Szoc, Tom Vanallemeersch
CrossLang, Kerkstraat 106, 9050 Gentbrugge, Belgium
`{first.lastname}@crosslang.com`

Lucia Specia, Julia Ive
University of Sheffield
`{initial.lastname}@sheffield.ac.uk`

Maxim Khalilov
Unbabel, Rua Visconde de Santarém, 67B, 1000-286 Lisboa, Portugal
`maxim@unbabel.com`

Abstract

The APE-QUEST project (2018–2020) provides a quality gate connected to the eTranslation system of EC's Connecting Europe Facility (CEF). The quality gate supports translation in specific domains and involves quality estimation (QE), automatic post-edition (APE) of machine translation (MT) output, human post-editing (PE) and secure data transfer. Public PE datasets are provided. Evaluations involving three language pairs are ongoing.

1 Overview

The APE-QUEST project (Automated Post-editing and Quality Estimation) is funded by the EC's CEF Telecom programme (project 2017-EU-IA-0151) which started in October 2018 and runs until September 2020. The project offers a quality gate connected to the eTranslation MT system, developed by the Directorate-General for Translation and provided by the CEF Automated Translation building block of the Directorate-General for Communications Networks, Content and Technology (DG CNECT) as a service to Digital Service Infrastructures (DSIs) of the EC and to public administrations of Member States. The project consortium consists of two companies, CrossLang (coordinator) and Unbabel, and the University of Sheffield.

APE-QUEST consists of a quality gate which injects QE and APE into the translation workflow. The main goals of this injection are (1) to reach the desired translation quality in an efficient and reliable way using MT and PE and (2) to create data aggregation opportunities by making translations and post-edits "locally owned", as data is generated and curated at the end user's site, in accordance with the similar main principle of the EC's ELRC action.[1]

The APE-QUEST project focuses on mature technologies by integrating systems for MT, QE and APE. Tests involve three language pairs (English to Portuguese, French and Dutch) and three domains (legal-domain text, procurement and online dispute resolution).

2 Architecture

The workflow consists of three tiers: (1) MT output with acceptable quality flows directly to the end user or connected system, (2) moderate-quality MT is enhanced through APE, and (3) low-quality MT is sent to a workflow for human PE. The input consists of text snippets (messages) or full text documents from the project's stakeholders, such as DSIs, public services in Member States, and organisations involved in CEF Telecom projects adopting eTranslation. The input is provided through an API or user interface and is segmented into sentences to allow for routing sentences to different tiers based on the detected quality.
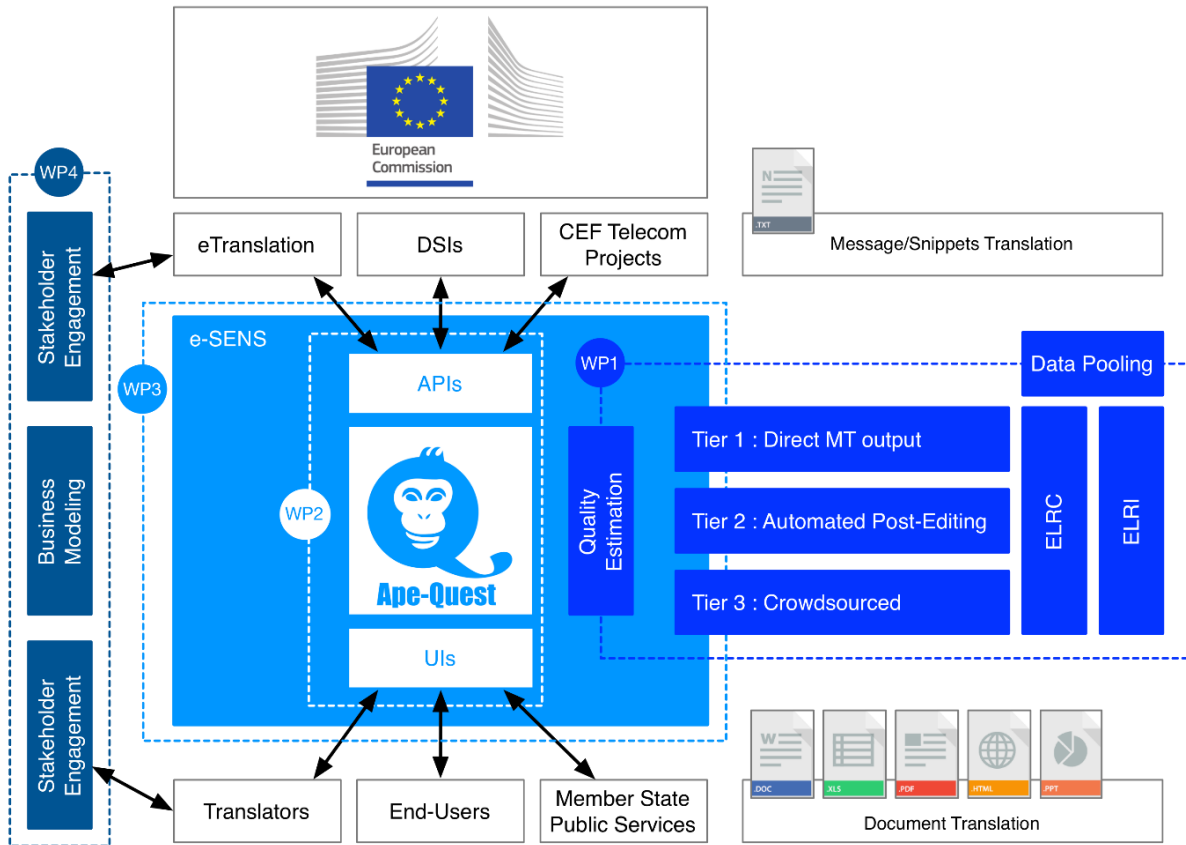
[1] http://lr-coordination.eu (European Language Resources Coordination)

Figure 1 Architecture of APE-QUEST

The PE workflow allows for collecting user data for system improvement, i.e. data for adapting the MT system and re-training QE and APE systems.

The components of the workflow have been successfully integrated and it is fully operational. It is interoperable with the eTranslation system, it is conformant with the EC's eDelivery building block in order to ensure secure and reliable data transfer, and it contains a portal-style front end (an adaptation of MateCat).[2]

3 Outcome

The human PE data created in the framework of the project is available in the ELRC-SHARE repository,[3] which collects, prepares and shares language resources. This data consists of around 30K tuples including a source sentence, a corresponding neural MT output, a post-edited version created by a professional translator, and the original reference translation crawled from parallel language websites in the legal domain. All data were anonymized.

Tests with the quality gate are ongoing in order to determine QE thresholds. The APE component for these tests is based on neural *copycat* networks, while the open-source framework *OpenKiwi* is used for QE. Both are described in Ive et al. (2020). The thresholds aim at establishing a trade-off between automatic processing (MT and APE) and human PE, in terms of translation speed, quality and cost. The tests involve two use cases, i.e. dissemination and assimilation of translated texts.

.

Acknowledgement

References

Ive, Julia, Lucia Specia, Sara Szoc, Tom Vanallemeersch, Joachim Van den Bogaert, Eduardo Farah, Christine Maroti, Artur Ventura, and Maxim Khalilov. 2020. A Post-Editing Dataset in the Legal Domain: Do we Underestimate Neural Machine Translation Quality? To appear in Proceedings of LREC 2020.

---

[2]   http://www.matecat.com

[3]   Name of the resource: "Post-editing corpus English to Dutch/French/Portuguese, legal domain"