APE-QUEST: an MT Quality Gate

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

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

Maxim Khalilov, Christine Maroti, Eduardo Farah*, Artur Ventura
Unbabel
Rua Visconde de Santarém, 67B
1000-286 Lisboa, Portugal
{firstname}@unbabel.com, *{first.lastname}@unbabel.com

Abstract

The APE-QUEST project (2018–2020) sets up a quality gate and crowdsourcing workflow for the eTranslation system of EC's Connecting Europe Facility to improve translation quality in specific domains. It packages these services as a translation portal for machine-to-machine and machine-to-human scenarios.

1 Objectives

The APE-QUEST project (Automated Post-editing and Quality Estimation) is funded by the EC's CEF Telecom programme (Connecting Europe Facility, project 2017-EU-IA-0151) and runs from October 2018 until September 2020. The project provides a quality gate and crowdsourcing workflow for the eTranslation machine translation (MT) system. The latter system is developed by the Directorate-General for Translation, supports all 24 official EU languages, and is provided by the CEF Automated Translation building block[1] 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 APE-QUEST consortium consists of two companies, CrossLang (project coordinator) and Unbabel, and the University of Sheffield.

APE-QUEST provides a quality gate by injecting quality estimation (QE) and automated post-editing (APE) into the translation workflow. QE and APE may be applied to the output of eTranslation or to crowdsourced translation. The main objectives of this injection are (1) to improve MT quality with additional linguistic services and (2) to create data aggregation opportunities by making translations and post-edits "locally owned", in the sense that the data is generated and curated at the end user's site, thus following the similar main principle of the EC's ELRC action.[2]

The APE-QUEST project focuses on integration of mature technologies: systems for MT, QE and APE, and an evironment for secure and reliable exchange of data, i.e. the EC's eDelivery building block. The tests in the project involve four languages, i.e. English, Portuguese, French

---

[1] https://ec.europa.eu/cefdigital
[2] http://lr-coordination.eu (European Language Resources Coordination)
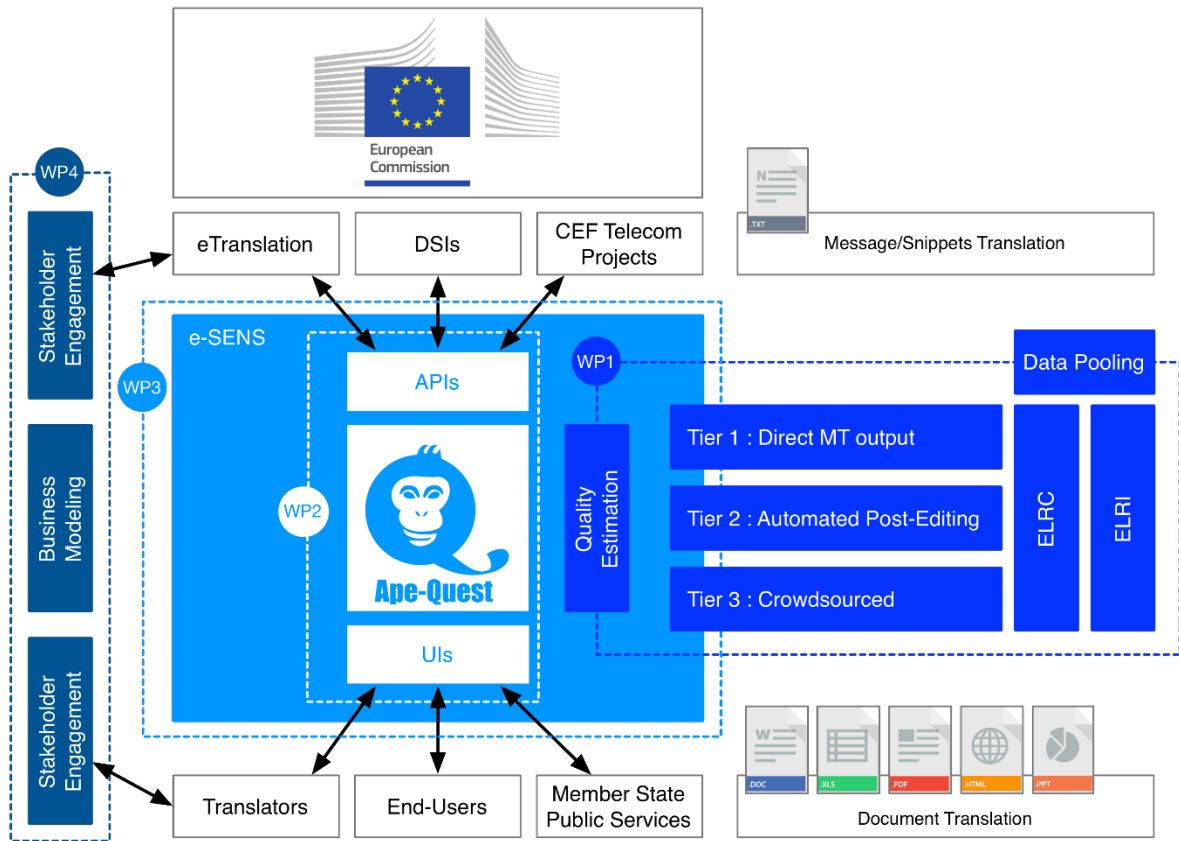
Figure 1 Architecture of APE-QUEST

and Dutch, and several domains, such as e-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 manual PE (crowdsourcing), after which QE and APE can be applied optionally, as an additional quality assurance step.

The input to the workflow consists of text snippets (messages) or full text documents originating from the project's stakeholders, such as DSIs, public services in Member States, and organisations involved in CEF Telecom projects that make use of eTranslation. The input is provided through an API or a user interface (UI) and is segmented into sentences to allow for complex routing. For example, if only one low-quality sentence is detected, only one tier 3 request is issued, while the rest of the sentences is routed to tiers 1 and 2.

The injection of a PE workflow allows for collecting user data for system improvement. These data will be made available to re-train and adapt the eTranslation system, and to re-train the QE and APE systems. The data will also be made available to ELRC, thus extending the latter's resources with postedited data, and to the ELRI project,[3] which collects, prepares and shares language resources.

APE-QUEST's reference implementation will be compliant with the eTranslation system and the eDelivery building block (developed in the framework of the EC's e-SENS[4] project), will contain a portal-style front end, and will be packaged in an easily deployable form for DSIs and public administrations of Member States. The workflow will be backward-compatible for systems that use the current eTranslation interface.

Acknowledgement

---

[3] http://www.elri-project.eu (European Language Resource Infrastructure)

[4] https://www.esens.eu