COPECO: a Collaborative Post-Editing Corpus in Pedagogical Context

Jonathan Mutal[1], **Pierrette Bouillon**[1], **Perrine Schumacher**[2], and **Johanna Gerlach**[1]

[1]FTI/TIM, University of Geneva, Switzerland
[2]CIRTI, Liège University, Belgium
{Jonathan.Mutal, Pierrette.Bouillon, Johanna.Gerlach}@unige.ch
p.schumacher@uliege.be

1 Motivation and objective

Learner corpora are important resources to help translation teachers understand student mistakes and improve their course content[1].

COPECO is a joint project between Geneva University and Liège University, with three main objectives: 1) to collect post-edits produced by students and teacher corrections, 2) to build an open-source student post-editing corpus and 3) to help systematise the task of translation error annotation (O'Brien, 2011). It provides translation teachers with an online post-editing platform, designed to help them to annotate student post-editing tasks using a shared or personalised annotation scheme.

2 COPECO platform

The platform[2] has six main features.

Annotation schemes: importing standardised annotation schemes (for example MQM[3]) and personalising them to suit the teachers needs i.e. adding, removing or editing error annotation categories; including explanations for students.

Post-editing tasks: defining tasks with list of students, source text, machine translated text, optional translation reference and metadata (text reference, MT tool, translation date, etc.); selecting the type of post-editing task (monolingual vs bilingual; sentence by sentence vs in block; showing source or target text first); assigning on-line post-editing tasks to students.

Post-editing statistics: recording post-editing activity data such as time, keystrokes and percentage of MT content edited.

Annotation: annotating errors in the text by appending tags from the selected annotation scheme and optionally adding comments for the students; comparing post-editing results with reference translations (if available) by highlighting differences.

Reports: displaying corrected texts with annotations, comments and error statistics; additionally, in the teacher view, displaying statistics for all post-editing tasks (number of keystrokes, time, annotations and percentage of MT content edited).

Collaborative corpus: sharing the anonymised annotated texts, with different access mode depending on student authorisation; making the corpus available for download in different formats.

3 Demonstration

The demonstration will present the platform and the different functionalities, taking as example a specific learner corpus (Casas, 2020).

References

Casas, S. (2020). Intégration de la post-édition dans un cours de traduction et révision : Étude de cas d'étudiant·e·s à la faculté de traduction et d'interprétation. Master's thesis. ID: unige:131300.

O'Brien, S. (2011). Towards a dynamic quality evaluation model for translation. *In Journal of Specialized Translation*, 17:1–2.

---

[1]https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpus-bibliography.html
[2]Link to the platform https://copeco.unige.ch:9996/
[3]http://www.qt21.eu/mqm-definition