

Application of an industry practical human MT output Quality Evaluation Metric in the EMT classroom

Parthena Charalampidou, Aristotle University of Thessaloniki, Greece
Serge Gladkoff, Logrus Global, USA

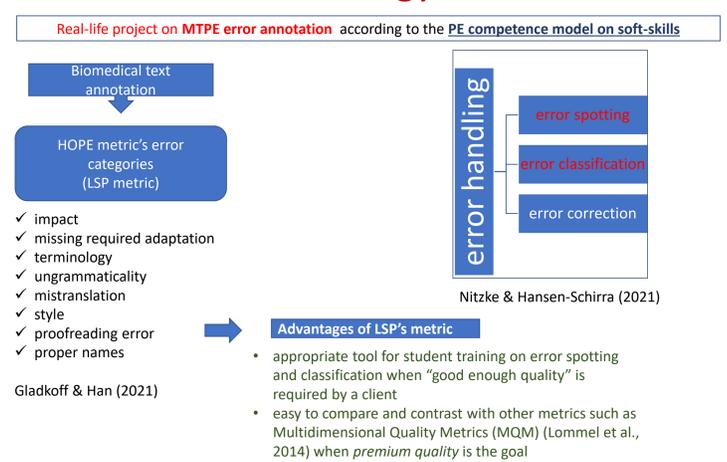
Introduction

Employability is one of the educational goals set in the Bologna Declaration for Higher Education Institutions (HEI) and as such its attainment should be a priority in the translation classroom as well. Considering the existing gap between academic studies and the challenges of the professional market instead of completely redesigning the academic curricula, an efficient approach would be to have industry practitioners collaborating in activities organized in higher education (Jacobs et al., 2022). Ankras and Al-Tabbaa (2015:387) provide a broad definition, defining University-Industry Collaboration (UIC) as “the interaction between any parts of the higher educational system and industry aiming mainly to encourage knowledge and technology exchange”. Apart from offering opportunities for internship, organizing collaborative conferences and developing joint research projects, connection with the industry is often attained through integration of industry practitioners in the academic teaching and even more incorporation of industry practitioners to co-teach courses with the instructor of record (Teaching With the Industry (TWI)) (ibid:2). The attempt to bridge the gap between academia and industry in the translation sector as well, is evident in new collaborations and initiatives such as ELIA Exchange Network, European Master's in Translation and EUATC or Lind-Web (Begoña, 2017) as well as in the discussion that has opened through scientific publications on the need for HEIs' adjustment to constantly changing professional demands and job requirements in the language industry (Schnell and Rodríguez, 2017; Álvarez-Álvarez and Amáiz-Uzquiza, 2017). The present poster attempts to describe a case of collaboration between industry and academia as a method of translator training enhancement.

Industry – Academia collaboration



Methodology/Tools

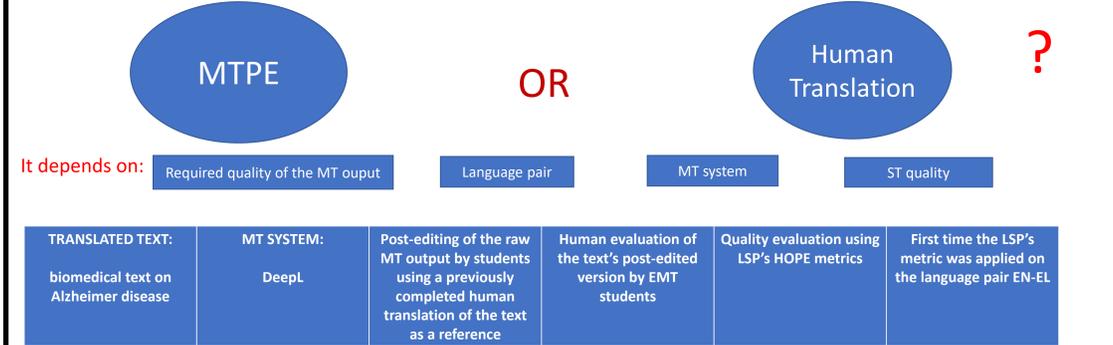


BENEFITS FOR BOTH COLLABORATORS

Diverging from a reference metric such as MQM in order to adopt a metric in the industry and actually work for a real translation assignment increased the degree of simulation for students and helped the trainer highlight the need for adaptability, flexibility and positive attitude towards alternative versions of the theory taught in the classroom.

Considering the risks that an organization has to manage every day and the decisions that have to be made, minimizing risks on a strategic level, such as deciding whether to use MT or not (Nitzke and Hansen-Schirra, 2021) is of utmost importance. According to Wallberg [...] there is [...] the question of whether the source text is actually suitable for MT” (2017: 150) and testing a tool such as HOPE in as many language pairs as possible can contribute to minimizing translation risks.

The experiment



Four classes devoted to post-editing and only two to MTPE quality evaluation

STEPS OF THE EXPERIMENT

- students were introduced to HOPE metric and were given a broad description of the error categories it distinguishes
- students completed a pilot annotation of a 336-word raw MT output of a specialized text so as to become acquainted with the specific evaluation metric before the actual experiment
- students were assigned the annotation of a 100-segment (2422 words) text which had to be completed within two hours
- a trained and more experienced annotator evaluated each segment's annotations in order to both assess the quality of students' annotations and the MT's output
- students analyzed the results of the annotation and decided on the translation policy the company should adopt:



- The analysis revealed that more than half of the students' annotations were similar to the reference one classifying the MT output in the category of minor mistakes and showed that MTPE could be applied for the specific type of specialized text from English into Greek without pre-training the MT system

BENEFIT:

- students got the opportunity to act as project-managers and take on the responsibility that such every-day decisions include

DRAWBACKS:

- the step of measuring inter-annotator agreement was omitted due to the fact that all students had to take part in the annotation exercise
- This created discrepancies between annotators which could have been avoided in real production scenario
- the study revealed that even in the case of a simplified evaluation metric such as HOPE there is a need for more distinct description of error categories to potential evaluators which would allow them to make connections with the main evaluation criteria that are used such as *accuracy (fidelity), fluency and fitness for purpose* (Koponen, 2010).

Implications



Conclusions

Ultimately, the application of an industry-based metric in the EMT classroom provided a real-life setting for students' training on quality assessment issues giving them the opportunity to come in contact with the translation market in the supervised context of the university. They were given the chance to get acquainted with MT quality evaluation combining theory with post-edited text annotation for an international company and get an insight into the priorities and the practical needs of the translation market.

ACADEMIA

For LSP, the results provided valuable insight into the IRR (Inter-Rater Reliability) and highlighted the factors to improve consistency of evaluations via validation of reviewers themselves. Data science methods of validation of reviewers, however, are very strict and cannot be applied due to the wider variety of human judgments about text. Therefore, additional research is needed to come up with proper validation process, as well as better distinction and clearer instructions on error categories and scoring model application. From practical point of view one of the recommendations is to use the same reviewer for similar type of content and samples over time.

INDUSTRY

References

Álvarez-Álvarez S. and Arnáiz-Uzquiza V. (2017). Translation and interpreting graduates under construction: do Spanish translation and interpreting studies curricula answer the challenges of employability?, In *The Interpreter and Translator Trainer* 11:2-3, (pp.139-159). DOI: 10.1080/1750399X.2017.1344812

Ankras, S. and Al-Tabbaa, O. (2015). Universities-industry collaboration: A systematic review. In *Scandinavian Journal of Management* 31 (pp. 387-408).

Begoña R. de C. (2017). Addressing employability and enterprise responsibilities in the translation curriculum. In *The Interpreter and Translator Trainer* 11:2-3, (pp. 107-122), DOI: 10.1080/1750399X.2017.1344816

Gladkoff, S. and L. Han. (2021). HOPE: A Task-Oriented and Human-Centric Evaluation Framework Using Professional Post-Editing Towards More Effective MT Evaluation. arXiv preprint arXiv:2112.13833.

Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.

Koponen, M. (2010). Assessing Machine Translation Quality with Error Analysis. In *MikaEL: Electronic proceedings of the KäTu symposium on translation and interpreting studies 4*. Suomen kääntäjien ja tulkkien liitto. https://sktifi.directo.fi/@Bin/40701/Koponen_MikaEL2010.pdf (consulted 22.02.2022)

Lommel, A. et al. (ed.) (2014). *Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics*. In *Revista Tradumática* 12 (pp. 455–463).

MQM. (2015) *Multidimensional Quality Metrics (MQM) Definition*. <http://www.qt21.eu/mqmdefinition/definition-2015-12-30.html> (consulted 24.02.2022).

Nitzke, J. and S. Hansen-Schirra. (2021). A short guide to post-editing. *Translation and Multilingual Natural Language Processing* 16. Berlin: Language Science Press.

Schnell B. and Rodríguez N. (2017) Ivory tower vs. workplace reality. In *The Interpreter and Translator Trainer* 11:2-3 (pp. 160-186). DOI: 10.1080/1750399X.2017.1344920

Valentin, V. and Shane, J.S. (2014). Advantages and challenges of research programs with extensive industry involvement. *Pract. Period.Struct. Des. Constr.* 19 (pp. 63–67). [

Wallberg, I. (2017). ISO 18587- a standard for the post-editing process. In J.Porsiel (ed.), *Machine translation*. Berlin: BDÜ Fachverlag (pp. 148–155).

Jacobs, F.; Cain, W.; Lu, R.; Daugherty, A. Case Study: Teachingwith Industry (TWI) Using NewVideoconferencing Technology and Innovative Classroom Setups. *Educ. Sci.* 2022, 12, 128. <https://doi.org/10.3390/educsci12020128>

Websites

Joint Postgraduate Studies Programme (J.P.P.S.) in “Conference Interpreting and Translation”. (2019). Programme of Studies – Structure: Contemporary Translation Practices, <https://pp.enl.auth.gr/en/trans-1-520>.