**Impact case study (REF3)**

REF2021

| **Institution:** University of Sheffield |
| --- |

| **Unit of Assessment:** B-11 Computer Science and Informatics |

| **Title of case study:** Enhancing quality and productivity and reducing costs of machine translations. |

| **Period when the underpinning research was undertaken:** 2012–2020 |

| **Details of staff conducting the underpinning research from the submitting unit:** |

| **Name(s):** | **Role(s) (e.g. job title):** | **Period(s) employed by submitting HEI:** |
| --- | --- | --- |
| Specia, L. | Professor of Language Engineering | 2012–present |
| Scarton, C. | Research Fellow | 2017–present |
| Blain, F. | Postdoctoral Researcher | 2015–present |
| Cohn, T. | Senior Lecturer | 2010–2014 |

| **Period when the claimed impact occurred:** 2015–2020 |

| **Is this case study continued from a case study submitted in 2014?** N |

**1. Summary of the impact** (indicative maximum 100 words)

Machine translation (MT) is inexpensive, fast, and accessible, but it lacks the reliability of human translators. Sheffield research on quality estimation (QE) in MT has enabled the identification of the likelihood of error, allowing MT to be used with greater confidence and underpinning impacts for multiple organisations. Microsoft introduced MT for text strings in program code in 37 languages, allowing more than 20% of translated software to be published without human intervention. Facebook improved its offering of approximately 6bn automated translations per day in >300 language pairs. In addition, the research shaped the latest development of the European Commission Directorate-General for Translation's (DGT's) MT system.

**2. Underpinning research** (indicative maximum 500 words)

In the context of the global spread of social media and a global population that is mostly monolingual, MT allows individuals to communicate across language barriers in a way that was impossible a generation ago. Professor Specia's research established the field of QE for MT, with the developed approaches summarised in **[R1]**, the only book on the topic, written by invitation for the main book series on natural language processing (NLP). Specia developed algorithms to score machine-translated documents for their predicted translation quality at the word, sentence, or document level. These methods can successfully identify text likely to be ungrammatical and poorly written and text that reads fluently but is likely to contain meaning errors.

In 2012, Specia and her collaborators established the Quality Estimation Shared Task (QuEst) **[R2]**, which provides automatic methods for estimating the quality of MT output at run time without reliance on reference translations. The task has run yearly since. In 2013, Specia published QuEst, the first sizeable research output in the area **[R3]** and the first framework released as a tool for MT QE. This framework provided the first language-independent feature set, which made QE applicable to any language pair and offered sentence-level quality prediction. In 2015, Specia and Dr Scarton expanded the framework to include word-level

prediction and document-level prediction by adding novel, appropriate features for these two levels, leading to the development of QuEst++ **[R4]**.

In 2018, Specia and Dr Blain developed and released deepQuest, a version of QuEst relying on deep learning, in response to new developments in MT software **[R5]**. Unlike previous approaches, deepQuest avoids the need for "feature engineering", thereby providing increased language independence as well as improved performance. Standard NLP pre-processing is extremely computationally intensive, but the lightweight deep learning approach in deepQuest, called the bidirectional recurrent neural network (biRNN) approach, does not require pre-processing or pre-training. In 2017, deepQuest was shown to perform on par with the state-of-the-art approach, Predictor-Estimator, with the advantage that it can be trained in a fraction of the time.

Finally, in 2020, Specia and Blain developed and released a new version of deepQuest **[R6]**; combining the biRNN approach with state-of-the-art, pre-trained representations. Their investigation of the new version showed that it allowed QE with unsupervised learning and provided state-of-the-art performance based on current benchmarks.

QuEst, deepQuest and QuEst++ are freely available on GitHub.

**3. References to the research** (indicative maximum of six references)

Sheffield staff and students in **bold**.

**R1.** **Specia, L.**, **Scarton, C.**, & **Paetzold, G. H.** (2018). Quality Estimation for Machine Translation. *Synthesis Lectures on Human Language* Technologies, *11*(1), 1–162. https://doi.org/10.2200/s00854ed1v01y201805hlt039. Cited by 31.

**R2.** Callison-Burch, C., Koehn, P., Monz, C., Post, M., Soricut, R. & **Specia, L.** (2012). Findings of the 2012 Workshop on Statistical Machine Translation. *Seventh Workshop on Statistical Machine Translation*, 10–51, WMT, Montréal, Canada. http://eprints.whiterose.ac.uk/171411/. This was the first edition of the Quality Estimation Shared Task. Cited by 647.

**R3.** **Specia, L., Shah, K.,** de Souza, J.G.C. & **Cohn., T.** (2013). QuEst - A translation quality estimation framework. *51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 79–84, ACL, Sofia, Bulgaria. http://eprints.whiterose.ac.uk/171412/. Cited by 170.

**R4.** **Specia, L.**, **Paetzold, G.**, & **Scarton, C.** (2015). Multi-level Translation Quality Prediction with QuEst++. *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, 115–120. https://doi.org/10.3115/v1/p15-4020. Cited by 94.

**R5.** Ive, J., **Blain, F.** & **Specia, L.** (2018). deepQuest: A framework for neural-based quality estimation. *Proceedings of the 27th International Conference on Computational Linguistics*, 3146–3157, Association for Computational Linguistics, Santa Fe, New Mexico, USA. http://eprints.whiterose.ac.uk/171414/. Cited by 25.

**R6.** **Fomicheva, M.**, Sun, S., Yankovskaya, L., **Blain, F.**, Guzmán, F., Fishel, M., **Aletras, N.**, Chaudhary, V., & **Specia, L.** (2020). Unsupervised Quality Estimation for Neural Machine Translation. *Transactions of the Association for Computational Linguistics*, *8*, 539–555. https://doi.org/10.1162/tacl_a_00330. Cited by 14.

**4. Details of the impact** (indicative maximum 750 words)

The Sheffield research contributed to rapid improvements in the quality of MT services. QuEst++ gives a reliable indication of probable quality, enabling users to focus efforts more effectively (reducing time and costs) and increasing confidence and awareness of quality risks when using raw MT texts. These improvements have led many organisations to use MT as their primary translation method, with human translators performing editing.

**Microsoft E&D Global**

In May 2018, Microsoft E&D Global introduced MT with automated QE based on QuEst into software localisation (i.e., the process of adapting software for a specific region or language, by translating text and adding locale-specific information). Although Microsoft was previously using MT for ~75% of their content, its use for software proved more complex, as the risks were greater since functional bugs in a product negatively impact customer experience (and result in a loss of customers). By 2019, MT was enabled for 37 languages on 90% of the Office product line. QuEst has allowed Microsoft to publish >20% of machine-translated word count without review, with no measurable negative impact on customer satisfaction **[S1]**.

According to Microsoft E&D's Senior Program Manager, *"The key benefit of this process for Microsoft has been savings of 20% in the cost of human translation and post-editing, but it has also delivered time savings by removing a stage from the translation and publishing process. With over a million words now published using raw MT across 37 languages, these savings are significant. […] Without QuEst we would not have been able to develop and deploy a quality estimation solution fit for purpose, therefore QuEst was absolutely fundamental in achieving the volumes we have managed to reach to date. [...] QE provided us with confidence without which I don't think we would have had buy-in from product owners to start publishing raw MT. If they did approve it, we would have been severely restricted in the volumes we would have been able to publish". Summarising, the Senior Program Manager stated, "We can safely say that QE gave us a 20-30% relative gain, in terms of volume of raw MT that we could publish, compared with using other mechanisms"* **[S2]**.

**Facebook**

Specia collaborated with Facebook within the REF assessment period on research to inform Facebook's development of prediction models of translation quality. The Research Scientist Manager at Facebook AI explained, *"Facebook offers its users machine translation of posts into their chosen language. Some translations are shown automatically instead of the original foreign language post – and the decision on which translations are shown automatically is based on confidence level. If our system is confident that the translation is likely to be of good quality then it will be shown automatically – if not, then it will not. Professor Specia's original research and the software she has been pushing forward (QuEst++ and deepQuest) largely set the standard for the area and provided inspiration for the prediction model used to make these decisions"* **[S3]**.

*"Facebook has set up a research agenda around quality estimation based on her work and in collaboration with Professor Specia, creating new corpora and data sets to increase the coverage of the prediction model from higher research languages (e.g. German, Chinese) to lower research languages (e.g. Sinhalese, Nepali)"* **[S3]**.

*"Facebook's goal is to bring down language barriers and allow people to connect regardless of language. This is an important part of our mission, and we take it seriously. Professor Specia's research has helped us to achieve our mission by enabling the technology that ensures only the highest quality and most accurate translations are shown to our members. Facebook shows more than 6 billion translations daily in more than 300 direct language pairs, and each translation is accompanied by a quality estimation prediction"* **[S3]**.


**European Commission DGT**

The research on biRNNs played a key role in guiding the introduction of automatic QE into eTranslation, the MT system of the European Commission DGT. The DGT aimed to improve the usefulness of their services through the QE technology and methods developed in Specia's research, as the previous quality control system was overly simplistic. eTranslation is both a supporting tool for DGT translators (who translate over 2 million pages of official documents into 26 languages every year) and a main translation route for other EU institutions and institutions of its member states (courts, parliaments, etc.). eTranslation has been used to translate up to 1 million pages of text in a single day, and usage continues to grow **[S4]**.

The components of the QE module were successfully integrated, and the module became fully operational and interoperable with the eTranslation system. The Sheffield technology estimates the translation quality from two methods: MT and "translation memory" (the default method used by the DGT, which employs a database of previous translations). The likely superior version is then recommended. For other users, each MT includes a quality/accuracy "health warning" based on a range of factors. These warnings allow greater confidence in the translated texts and highlight the areas of greatest risk.

The Project Manager for Machine Translation at the DGT stated, *"Without the work of the Sheffield Research Group we would not have been able to contemplate integrating quality estimation functionality into eTranslation at this point in time. Specia worked closely to advise and guide our internal IT and translation experts, including delivering a workshop to our staff entirely devoted to quality estimation. Her influence on our direction of travel and our approach to the development of our software has been fundamental"* **[S5]**.

A pilot of the module is currently underway.


**Localisation and translation technology specialists**

Specia's work is the cornerstone for QE in MT across the industry. A host of localisation and translation technology specialists worldwide, such as Belgian consulting and systems integration company CrossLang **[S6]** and ModelFront (formerly SignalN) **[S7]**, have been guided by Specia's research or used her products to develop and benchmark their own product offering. The CEO of ModelFront stated, "*Specia's work has defined the category of Quality Estimation of machine translation, and all companies working in this area are indebted to her for her early vision and leading the research community for the last decade. As a dedicated provider of Quality Estimation technology, ModelFront is grateful to her for proving this technology both scientifically and in the eyes of the industry*" **[S7]**.

**5. Sources to corroborate the impact** (indicative maximum of 10 references)

**S1.** Paper presented by Microsoft E+D Global at the 2019 Dublin Machine Translation Summit. Confirms the use of DeepQuest in their Automated Localisation Workflow and citing

Specia's papers (pages 159 - 166). (Accessed 16th June 2020).
https://www.aclweb.org/anthology/W19-67.pdf#page=177

**S2.** Confidential testimonial letter from Senior Program Manager (Applied Science) on Microsoft's Global AI Experiences Team (2020). Corroborates the role of Sheffield's research in Microsoft's Quality Estimation tool and the improvement seen through its use.

**S3.** Confidential testimonial statement from Research Scientist Manager at Facebook AI (2020). Corroborates that a) Sheffield's technology has set the benchmark in this area, b) Facebook set up its own research programme based on Sheffield's research, c) Facebook worked with Sheffield to further develop the model, and provides d) quantitative data on the number of translations per day.

**S4.** European Commission Directorate-General for Translation 2019 Annual Activity Report. Reports facts and figures related to translation services (page 8 &9). (Accessed 16th June 2020). https://ec.europa.eu/info/sites/info/files/dgt_aar_2019_en.pdf

**S5.** Confidential testimonial letter from the Project Manager for Machine Translation at the European Commission Directorate General of Translation (2019). Corroborates the critical role of Sheffield's research in guiding and defining their forward strategy in this area.

**S6.** Confidential testimonial letter from the Founder and Director of CrossLang (2020). Corroborates the use of Sheffield's technology by CrossLang.

**S7.** Confidential testimonial letter from the technical co-founder and CEO of ModelFront (2020). Corroborates the influence of Sheffield's technology on ModelFront and the sector.