Trados-to-Translog-II: Adding Gaze and Qualitivity data to the CRITT TPR-DB

Masaru Yamada, Takanori Mizowaki
Rikkyo University, Tokyo, Japan
masaru.yamada@rikkyo.ac.jp;
taka.m0118@icloud.com

Longhui Zou, Michael Carl
Kent State University, Ohio, USA
{lzou4.mcarl6}@kent.edu

Abstract

The CRITT (Center for Research and Innovation in Translation and Translation Technology) provides a Translation Process Research Database (TPR-DB) and a rich set of summary tables and tools that help to investigate translator behavior. In this paper, we describe a new tool in the TPR-DB that converts Trados Studio keylogging data (Qualitivity) into Translog-II format and adds the converted data to the CRITT TPR-DB. The tool is also able to synchronize with the output of various eye-trackers. We describe the components of the new TPR-DB tool and highlight some of the features that it produces in the TPR-DB tables.

1 Introduction

Much of the translation process research (TPR) has been conducted with Translog-II, which is an editor that allows to record keystrokes and eye tracking data during translation sessions (Carl, 2012). The collected data can then be uploaded to the Translation Process Research Database (CRITT TPR-DB) which provides numerous tools for data analysis and data visualization (Carl et al., 2016). However, Translog-II does not offer professional editing possibilities nor is it a translation environment which professional translators normally use. To alleviate these shortcomings, to emulate translators' real-world working conditions in experimental conditions and thus increase the ecological validity of TPR, we have implemented a new Trados interface that uses Qualitivity [1] to log keystroke and that converts the output into a CRITT TPR-DB format. In addition, eye-tracking data (currently: Tobii, Eyelink, GazePoint) can be collected and synchronized with the Qualitivity keystroke data while uploading the data to the CRITT TPR-DB.

2 Uploading Trados data to TPR-DB

Trados Studio[2], a commercial CAT software, has a non-invasive free-of-charge plugin called Qualitivity that collects typing activities from the translator. Qualitivity captures modifications in the editor (usually induced through keystroke, but also through automated processes) and assigns each event a timestamp as well as the segment number in which the modification occurs. This data can be converted into a Translog-compatible XML format and integrated into the CRITT TPR-DB via the newly added uploading option in the CRITT TPR-DB management tool, as shown in the following Figure 1.

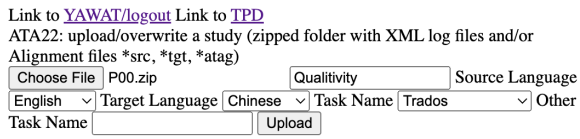

Figure 1: Upload CRITT TPR-DB management tool

3 Synchronizing Eye-tracking data

In order to investigate translator's gazing behavior during the translation processes, we integrated an add-on to the Trados-Translog-II conversion tool that merges eye tracking data with the Qualitivity keystroke data in a seamless way. This allows us to exploit user activity data collected during translation sessions in Trados as a combination of eye movement and keyboard logging. However, unlike Translog-II, Trados (or Qualitivity) does not offer the possibility to connect directly to external eye trackers. Qualitivity also does not record where on the screen (X/Y positions) the edited word or segment occurs. Due to the

[1] https://community.rws.com/product-groups/trados-portfolio/rws-appstore/w/wiki/2251/qualitivity
[2] https://www.trados.com/products/trados-studio/

different ways in which the gaze data is recorded in Translog-II and within the Trados setting, some different processing strategies are required.

Qualitivity and the eye-tracker software (Tobii, Eyelink, GazePoint) are independent programs, each of which is equipped with an independent keylogger. That is, every keystroke pressed by the translator in Trados is logged twice, once by Qualitivity (as a text modification), and once by the eye tracker software, each with independent timers and thus potentially different timestamps. The mapping of gaze data and textual data works in three steps 1) Comparing sequences of keystrokes recorded with Qualitivity and the eye tracker allows us finding an offset between the timestamps, and successively synchronize gaze data with the Qualitivity keystrokes. 2) While the eye tracker provides us with a stream of X/Y coordinates that reveal where on the screen the translator is looking at a certain point in time, Qualitivity tells us for each keystroke which segment was edited at a given point in time. The synchronized, combined information allows us, then, to relate the course of gaze events (X/Y coordinates) with (sequences of) keystrokes (or text modifications) that occur in a certain Trados segment[3] 3) Processing the data within the CRITT TPR-DB implies a) mapping each keystroke on the target word that it produces (Carl, 2012).and b) aligning the source text and the target texts on a word-level. This makes it possible to map the synchronized gaze data on the emerging target text via the alignment relations on source segments, just as with Translog-II data,

4 Gaze-path features

However, in contrast to Translog-II - which provides the possibility for gaze–to–word mapping at runtime - we do not know which word(s) the translator is looking at when collecting data with Trados. In the Trados setting we only know the X/Y coordinates of the gaze path, whether the gaze occurred on the ST or the TT windows, and which segment the translator was working on. This information is, however, highly informative and provides indicators of translation effort, as encoded, for instance, in the relative and total distances between fixations, the number of regressions, the gaze movements in the X and Y direction of the screen, parallel (concurrent) reading and typing behavior, etc. Within the TPR-DB, we compute this gaze path summary information for various process and product units. For instance, gaze path information can be computed on the segment level (SG) for production units (PU), alignment groups (AG), or also for each word (ST), which provides novel ways to assess translation effort and effects on various levels of granularity.

5 Conclusion

The new Trados-TPR-DB interface provides the possibility to record translation behavior in an ecologically realistic translation environment. We are now able to investigate patterns of reading and typing activities in a widely and professionally used CAT tool, and thus to achieve a better understanding of factors that impact professional translation activity. The collected data can be uploaded and processed in the CRITT TPR-DB, which offers a wide variety of analysis and visualization tools (Carl et al., 2016) as well as detailed information about translation effort and translation effects. We are able to expand the features with customized algorithms and open-source NLP packages to conduct further analysis regarding the translation effect, such as automatic translation assessment tools (e.g., BLEU and COMET), and linguistic complexity metrics (e.g., LingX) (Zou et al., 2022). Several ongoing studies are already using this new tool and more results are likely to be available soon.

References

Carl, Michael, Moritz Schaeffer, and Srinivas Bangalore. 2016. The CRITT translation process research database. In *New directions in empirical translation process research* (pp. 13-54). Springer, Cham.

Carl, Michael. 2012. Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research. *LREC*,12: 4108-4112.

Zou, Longhui, Michael Carl, Mehdi Mirzapour, Hélène Jacquenet, and Lucas N.Vieira. 2022. AI-Based Syntactic Complexity Metrics and Sight Interpreting Performance. In: Kim JH et al. (eds) *Intelligent Human Computer Interaction. IHCI 2021. Lecture Notes in Computer Science* (vol. 13184, pp. 534-547). Springer, Cham.

[3] We assume that gazing patterns preceding a (sequence of) keystrokes are indicative of the exerted effort related to their production.