International Journal of Applied Linguistics and Translation
2022; 8(4): 148-152
http://www.sciencepublishinggroup.com/j/ijalt
doi: 10.11648/j.ijalt.20220804.13
ISSN: 2472-1166 (Print); ISSN: 2472-1271 (Online)

Measuring Cognitive Effort with Translation Process Database

Wang Jiayi, Xiao Chenyixuan

College of Foreign Languages, Hunan Institute of Engineering, Xiangtan, China

Email address:
boriswjy@163.com (Wang Jiayi)

To cite this article:
Wang Jiayi, Xiao Chenyixuan. Measuring Cognitive Effort with Translation Process Database. *International Journal of Applied Linguistics and Translation*. Vol. 8, No. 4, 2022, pp. 148-152. doi: 10.11648/j.ijalt.20220804.13

Received: October 14, 2022; **Accepted**: November 7, 2022; **Published**: November 16, 2022

Abstract: The analysis and measurement of cognitive effort could be complicated when involved in translation production. And it therefore attracts researchers' great attention to the investigation of this topic. Different from traditional data collection methods, the Translation Process Research Database (TPR-DB) utilizes the large corpus to record the translation process, including translation process data (e.g. keystrokes, fixations, mouse movements) and translation product data (e.g. ST, TT and links between tokens in both texts) from more than ten language pairs and dozens of translation and associated studies. After reviewing the studies and some findings on measuring cognitive effort with the TPR-DB, the present study proposes that features of HTra, HCross, AUs and PWR in the TPR-DB tables are frequently used as indicators for the measurement of cognitive effort during translation and post-editing processes. The attempts to measure cognitive effort with the TPR-DB have not only yielded some interesting findings but also added fresh insights to facilitate understanding and examination of cognitive effort. The present study pointed out that the TPR-DB provides a new and effective method to measure cognitive effort. It will further support and promote the future research in this field.

Keywords: Cognitive Effort, TPR-DB, Measurement

1. Introduction

Translators' cognitive effort remains the research focus for its essential role in exploring how the black box of the translating brain works. Cognitive effort, according to Lacruz [11], refers to the mental effort involved in reading the texts, thinking about how to translate and how to correct mistranslations, selecting the desired product, and reflecting on the chosen solutions. One of the first authors who investigated this issue is Krings [9] who classified three distinctive but closely relevant kinds of cognitive effort in the post-editing process. The temporal effort refers to working time. The technical effort involves keyboarding activities, such as inserting, reordering and deleting. Cognitive effort exerts a further profound influence on the former two types of effort because it comprises the necessary "type and extent" for improving machine translation [9]. It was considered the most challenging to understand and is the most difficult to measure [9, 11]. Krings' pioneering work offers "a good framework" [11] to explore translators' efforts in the translation process and other related activities. The follow-up studies explored the translator's cognitive effort through various methods and yielded many convincing findings [4, 11-16]. The present study introduces studies on the cognitive effort based on Translation Process Database (TPR-DB).

2. Translation Process Database

In recent years, Carl [3] introduced the Translation Process Research Database (TPR-DB) to TPR. TPR-DB provides researchers with efficient, consistent, reliable, and large-scale statistics by establishing topic-related databases and setting features accordingly. TPR-DB mainly consists of user activity data (UAD), including translation process data (e.g. keystrokes, fixations, mouse movements) and translation product data (e.g. ST, TT and links between tokens in both texts). Currently, with the Creative Commons licence, users can gain access to both the raw and the processed data from the database, which includes ten language pairs, over 500 hours of text production, and more than 40 translation and associated studies (through Translog, Translog-II and the CASMACAT workbench). Figure 1 presents the architecture of the TPR-DB compilation process.

As can be seen from Figure 1, the TPR-DB compilation process is practiced through three stages: data collection, data annotation, and data integration and evaluation. The logged UAD data, which include product data and process data, are first collected via Translog-II. Then, they are annotated respectively (at the top and bottom process). The annotations of the product data "include tokenization, sentence and token alignment and (optionally) lemmatization, PoS tagging among others" [3]. For process data annotation, Translog-II allows users to perform manual gaze-to-word re-mapping. In the third step, keystroke-to-token and fixation-to-token mappings are computed to integrate the product data and process data. Finally, tables are generated to assist in further analysis and visualisation.

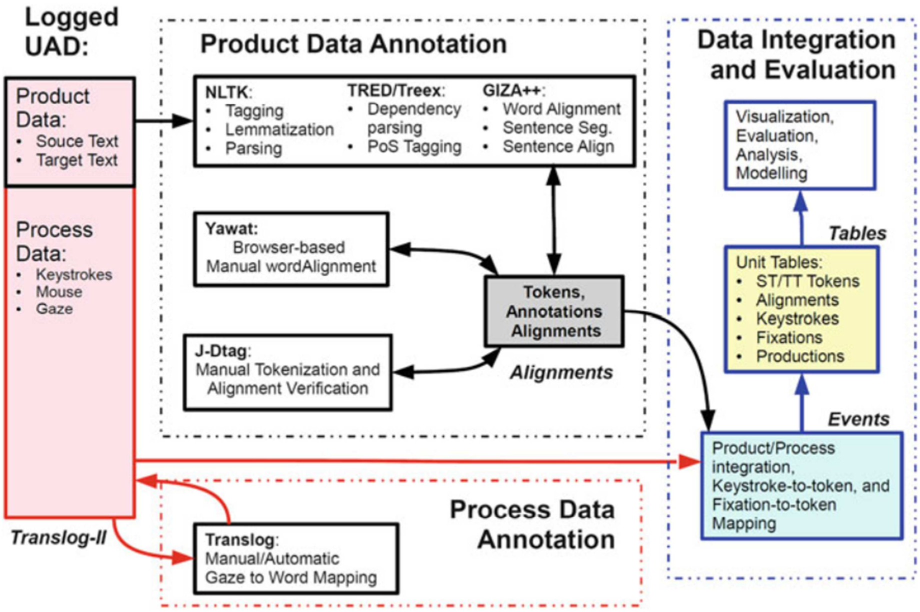

Figure 1. *Architecture of the TPR-DB compilation process (Carl et al. 2016: 16).*

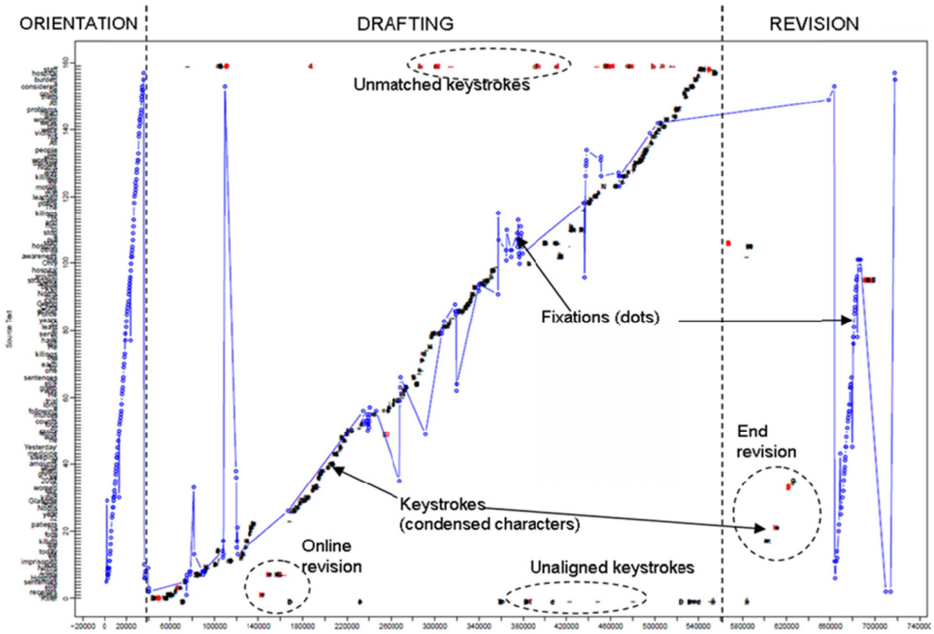

Figure 2. *Translation progression graph (Dragsted & Carl, 2013: 141) [5].*

TPR-DB is used to visualise how translations emerge in time. When compiling TPR-DB, the translation process data and translation product data were integrated and then presented in different tables. Information from those tables can be analysed, evaluated and visualised. The translation progression graph (TPG) is a new tool advocated by Carl to visualise the translation process. It plots the processing information from data each time and maps the progress of translation. The TPG reveals the occurrence of “pauses and deletions” and the time distribution of “keystrokes and gaze activities” [2]. Figure 2 is an example of a TPG. The vertical axis on the left shows the words of the ST (from 1 to 100); the horizontal axis lists the time nodes when translations of the ST were produced. The fixations and keyboard activities of the translator are plotted as well. Blue, black and red dots refer to the fixations, insertions and deletions respectively.

TPR-DB can also be used to investigate activity units (AUs). AUs, defined by a pause threshold (e.g. 1 ms), are units of the consecutive or concurrent gazing and keyboarding activities involved in translation. Start time (Time), duration (Dur), and segment (Seg) are the indicators to evaluate an AU [2]. Each row in the AU tables contains information regarding one AU, which can be one of six different types: reading the ST (type 1), reading the TT (type 2), translation typing (type 4), writing while reading the ST (type 5), writing while reading the TT (type 6) and no activity recorded for more than 2.5 seconds (type 8). Jensen [8] compared student and experienced translators’ performance in allocating cognitive resources. It was found that both groups allocated more cognitive resources to TT processing than ST processing. Student translators spent more time on ST processing than professionals. Professionals allocated more time to parallel ST/TT processing (AUs of type 5 and type 6) than student translators. There was no text difficulty effect on translators’ management of cognitive resources. Hvelplund [6] investigated the processing flow between two successive AUs. He designed two transition matrixes to compare professionals and student translators’ transition activity. He observed that in 65.5% of cases, professionals shifted from ST reading (type 1) to typing-related activities (AUs of type 4, 5 or 6), whereas student translators did this in only 52.2% of cases. TT reading (type 2) succeeded ST reading (type 1) more often for student translators than for professionals, which suggests that student translators needed to confirm their meaning hypotheses more often than professionals. Another similar study [14] investigated successive AUs by comparing the transitions during translation and post-editing processes. The authors found that post-editors make more transitions between ST reading (type 1) and TT reading (type 2) than translators. In post-editing mode, the possibility that a post-editor will switch to TT reading from ST reading state is 81%. Once the state is TT reading, the probability that a post-editor will shift to ST reading is 56%. In translation state, however, the ST-to-TT and TT-to-ST transition rates are 52% and 42%, respectively. Another interesting finding is that the highest probability that a post-editor will start a writing activity (types 4, 5 and 6) occurs when transitioning from TT reading (41%) rather than ST reading (16%). In translation mode, the probability is more balanced: a 54% chance of transition from TT to writing activity and 42% from ST to writing activity.

Therefore, TPR-DB facilitates the proposal and proof of hypotheses centering on the translation process of different language combinations and modalities. It also provides researchers with reliable and consistent resources to research the cognitive effort in translation and post-editing, reading and copying across diverse languages and individuals.

3. Cognitive Effort and TPR-DB

One application of TPR-DB lies in investigating word translation entropy and syntactic entropy. After the product and process data are integrated, a number of different tables can be generated using the TPR-DB management tool, which offers many features for describing and modelling behaviour during translation. HTra and Cross are two of these features, as introduced by Carl [3]. HTra (word translation entropy), referring to the number of different corresponding translations for each ST word, is introduced by Carl et al. Carl [3] to investigate the semantic similarity between ST and TT. The metric is on the basis of the features (number and distribution) of available translations for a source word in the given context in TPR-DB. Another metric, Cross (syntactic entropy), records the order differences of the source words from ST and the translations from TT. When their orders are similar, the Cross value will be 1. On the contrary, if the first source word is relative to the fifth word in translation, the value will be 5. A positive relationship can be found between the absolute Cross value and the syntactic differences between ST and TT. Schaeffer et al. [15] investigated the effect of HTra and Cross on the measurement of eye movements in the translation process. The study examined 42,211 items (ST tokens) in the nine studies in TPR-DB. They found that the number of alternative translations for a single source word (HTra) and cross-linguistic syntactic re-ordering (Cross) had a significant positive effect on first fixation duration and total reading time. This finding suggests that in the process of translation, researchers can utilize data to evaluate the “relative word order and semantic overlap between lexical items of two different languages” [15], and observe their effect on translators’ eye movements. It is therefore likely that “the effect of CrossS and HTra on first fixation durations represents early, automatic cognitive alignment, which is less effortful in the case of ST items for which the overlap between ST and TT representations in terms of syntax and lexico-semantics, respectively, is greater (low HTra and Cross values)” [16]. Bangalore et al. [1] discussed the correlation between syntactic variation and priming in translation by using translation data from the TPR-DB. The authors stated that the priming effect should be considered in the translation process at both the lexical and syntactic levels. The results showed that syntactic entropy and restructuring effort (CrossS) had significant positive effects on the total ST reading time. When the language pairs shared similar syntactic structures, the response time in translation was shorter because of the priming effect activated by the co-activated networks.

In Jensen’s [7] eye-tracking experiment, two indicators were used to investigate whether the reordering of L1 syntax may result in increased processing effort. Total gaze time results confirmed that translators took significantly longer time dealing with segments when they needed to change the word order. However, the word order almost showed no effect on pupil dilation. The author argued that “this difference between the two eye-tracking indicators could suggest that they measure different processes, that pupil dilation is delayed or that gaze time does not equal processing effort" [7].

Higher pause-to-word ratio (PWR) values are related to more cognitive effort [10, 15]. Based on the pausing and typing structure, Schaeffer et al. [15] adopted Activity Units in the measurement of cognitive effort in TPR-DB. The data used in this study were extracted from 6 different studies (KTHJ08, BML12, SG12, MS12, ENJA15 and NJ12) in the TPR-DB. Followed Carl et al. [3], the authors in this study also identified 6 types of Activity Units according to reading and typing activities during from-scratch translation and post editing. The 6 types of Activity Units were Type 1 (ST reading), type 2 (TT reading), type 4 (translation typing), type 5 (ST reading and typing), type 6 (TT reading and typing) and type 8 ((no gaze or typing activity recorded for more than 5 seconds). After the investigation to the correlation of Translation Difficulty Index (TDI) and pause-to-word ratio (PWR), the study pointed out that TDI had a significant positive effect on the PWR with a pause threshold of 5000ms. The finding indicated that both TDI and PWR were suitable predictors to measure effort in the translation process. Because of the relatively lower PWR scores during the post-editing process than in from-scratch translation, the study assumed that translators had a less cognitive load in the post-editing mode. The result also showed the translation from the European languages to English was less effortful when compared with the translation from the Asian languages (Chinese, Japanese, Hindi) to English. It concluded that more remote language pairs required more cognitive effort in translation.

Lacruz et al. [12] focused on literality and cognitive effort. The study used data from the BML12 study for English-to-Spanish and the ENJA15 study for English to Japanese in TPR-DB to examine the linguistic complexity and semantic and syntactic remoteness across different language pairs and investigate their influence on the cognitive effort in translation and post-editing processes. HTra, CrossS and PWR were used as indicators to measure cognitive effort. According to the results, HTra and monitoring efforts has no strong connection for the two language pairs. While for CrossS, the results showed a distinction between the language pairs. To be specific, the translation process from English to Spanish showed no correlation between CrossS and monitoring efforts. While in the translation from English to Japanese, considering the structural differences between the two languages, the strong and close connection between CrossS and monitoring effort was noticed. It seemed reasonable to assume that the online monitoring of language production was necessary to assure structural integrity in the translation process. The study concluded that more cognitive effort was needed to translate more remote language pairs. This finding is consistent with the previous research conducted by Carl [3, 15] Schaeffer [15]. Lacruz et al. [12] also found that (1) in comparison with the post-editing process, from-scratch translation involves more cognitive effort, as indicated by PWR values, and (2) monitoring translation production needs great cognitive effort in translation from scratch or post-editing.

4. Conclusion

The compilation of the TPR-DB provides researchers with accessible, efficient and reliable translation process data to carry out relevant empirical research. The TPR-DB can be used to investigate activity units (AUs), visualise how translations emerge in time and draw comparisons across different individuals, language combinations and translation modalities. Features of HTra, HCross, AUs and PWR in the TPR-DB tables are frequently used as indicators for the measurement of cognitive effort during translation and post-editing processes. Studies show that the attempts to measure cognitive effort with the TPR-DB have not only yielded some interesting findings but also added fresh insights to facilitate understanding and examination of cognitive effort. It will further support and promote the future research of cognitive effort by providing new research methods in this field.

Acknowledgements

This study was supported by Hunan Social Science Fund of China (Grant No. 19ZDB008; 19YBA113) and Scientific Research Project of Hunan Provincial Department of Education (Grant No. 19A105).

References

[1] Bangalore, S., Behrens, B., Carl, M., Ghankot, M., Heilmann, A., Nitzke, J., & Sturm, A. (2016). Syntactic variance and priming effects in translation. In In M. Carl, S. Bangalore & M. Schaeffer (Eds.), New directions in empirical translation process research (pp. 211-238). Springer, Cham.

[2] Carl, M., & Jakobsen, A. L. (2009). Towards statistical modelling of translators' activity data. International Journal of Speech Technology, 12 (4), 125-138.

[3] Carl, M., Bangalore, S., & Schaeffer, M. (Eds.). (2016). New directions in empirical translation process research. London: Springer.

[4] Carl, M. (Ed.). (2021). Explorations in empirical translation process research. Heidelberg: Springer.

[5] Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-tracking and key-logging data. Journal of Writing Research, 5 (1), 133–158.

[6] Hvelplund, K. T. (2016). Cognitive efficiency in translation. Reembedding translation process research, 128, 149-170.

[7] Jensen, K. T. H., Sjørup, A. C., & Balling, L. W. (2009). Effects of L1 syntax on L2 translation. In I. M. Mees, F. Alves, & S. Göpferich (Eds.), Methodology, Technology and Innovation in Translation Process Research (pp. 319–336). Copenhagen: Samfundslitteratur.

[8] Jensen, K. T. H. (2011). Distribution of attention between source text and target text during translation. In S. O'Brien (Ed.), Continuum Studies in Translation: Cognitive Explorations of Translation (pp. 215–236). London & New York, NY: Continuum.

[9] Krings, H. P. (2001). Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes. Kent, Ohio: Kent state university Press.

[10] Lacruz, I., & Shreve, G. M. (2014). Pauses and cognitive effort in post-editing. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard, & L. Specia (Eds.), Post-Editing of Machine Translation: Processes and Applications (pp. 244-272). Cambridge: Cambridge Scholars Publishing.

[11] Lacruz, I. (2017). Cognitive effort in translation, editing, and post-editing. In J. W. Schwieter & A. Ferreira (Eds.), The handbook of translation and cognition (pp. 386-401). Hoboken: Wiley-Blackwell.

[12] Lacruz I, Carl M, Yamada M. (2018). Literality and cognitive effort: Japanese and Spanish. In A, Calzolari (eds.), Proceedings of the Eleventh International Conference on Language Resources and Evaluation (pp. 3818-3821). Paris: European Language Resources Association.

[13] Lacruz, I., Ogawa, H., Yoshida, R., Yamada, M., & Ruiz Martinez, D. (2021). Using a product metric to identify differential cognitive effort in translation from Japanese to English and Spanish. In M, Carl (eds.), Explorations in Empirical Translation Process Research (pp. 295-314). Springer, Cham.

[14] O'Brien, S. (2022). How to deal with errors in machine translation: Postediting. In D, Kenny (eds), Machine translation for everyone: Empowering users in the age of artificial intelligence, (pp. 105-120). Berlin: Language Science Press.

[15] Schaeffer, M., Carl, M., Lacruz, I., & Aizawa, A. (2016). Measuring Cognitive Translation Effort with Activity Units. Baltic Journal of Modern Computing, 4 (2), 331–345.

[16] Schaeffer, M., & Carl, M. (2017). Language processing and translation. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), Empirical modelling of translation and interpreting, (pp. 117–154). Berlin: Language Science Press.