Literality and Cognitive Effort: Japanese and Spanish

Isabel Lacruz, Michael Carl, Masaru Yamada

Kent State University, Kent, Ohio, United States

CRITT, Copenhagen Business School, Copenhagen, Denmark

Kansai University, Osaka, Japan

ilacruz@kent.edu, mc.ibc@cbs.dk, yamada@apple-eye.com

Abstract

We introduce a notion of pause-word ratio computed using ranges of pause lengths rather than lower cutoffs for pause lengths. Standard pause-word ratios are indicators of cognitive effort during different translation modalities.The pause range version allows for the study of how different types of pauses relate to the extent of cognitive effort and where it occurs in the translation process. In this article we focus on short monitoring pauses and how they relate to the cognitive effort involved in translation and post-editing for language pairs that are different in terms of semantic and syntactic remoteness. We use data from the CRITT TPR database, comparing translation and post-editing from English to Japanese and from English to Spanish, and study the interaction of pause-word ratio for short pauses ranging between 300 and 500ms with syntactic remoteness, measured by the CrossS feature, semantic remoteness, measured by HTra, and syntactic and semantic remoteness, measured by Literality.

Keywords: pause-word ratio, Literality, cognitive effort

1. Introduction

The Multiling subset of the CRITT TPR-DB database (Carl et al., 2016a) provides a large corpus of translation process data that facilitates comparisons across different languages and different translation modalities. It assembles user activity data obtained from translation tasks into several languages using a common set of six short English source texts. In particular, keystroke and eye tracking data were recorded during from-scratch translation sessions and during post-editing of machine translations.

In this paper, we focus on translation and post-editing data from the BML12 study for English-to-Spanish (Mesa-Lao, 2014) and the ENJA15 study for English to Japanese (Carl et al., 2016b). By introducing refinements of the pause-word ratio measure of cognitive effort (Lacruz and Shreve, 2014) given by different ranges of pause lengths, we identify different patterns of cognitive effort for the two language pairs. These point to possible differences in the translation process when languages are more or less remote from each other that merit systematic investigation. These differences are potentially of interest to researchers in Natural Language Processing and Text-to-Speech synthesis.

In terms of language structure, Spanish is much closer to English than Japanese is to English. It is therefore to be expected that translation related tasks will be more effortful for English>Japanese than for English>Spanish. In addition, typed production of Japanese using an input method editor (IME) is more complex, and so more effortful, than typed production of Spanish. The expected extra effort involved in English>Japanese translation tasks as compared to English>Spanish translation tasks has been confirmed, for example in Carl et al. (2016b) and Schaeffer et al. (2016).

Linguistic complexity, as opposed to typing complexity, is a factor that contributes to increased cognitive effort expended on translation tasks (Dragsted, 2011). One type of complexity arises from translation entropy, which is computed by HTra in the CRITT TPR database. Translation entropy of a source text word is derived from the number of different translation choices made by different translators. For example, English *seat* could be rendered in several ways in both Spanish and Japanese. Examples include *asiento* or *silla* in Spanish, and 倚子 or 安楽椅 in Japanese. The actual value of HTra for a source text word is determined by the relative frequencies of different target language translation choices in the data set. HTra is 0 for a word for which only a single translation is provided, but increases with the number of translation choices provided. Details can be found in Carl et al. (2016a). Carl et al. (2016b) provide evidence that HTra is higher when the target language is more remote from the English source language. In particular, HTra is about double for English>Japanese than for English> Spanish when the same source texts are used.

Another type of linguistic complexity arises when word alignments differ in the source and target languages. For example, adjective alignments are different in English, where the default is adjective – noun, and in Spanish, where the default is noun – adjective. As another example, verb alignments with the subject and object are different in English (subject – verb – object) and Japanese (subject – object – verb). The CrossS feature in the CRITT TPR database gives information about alignment differences between the source and target texts. Working left to right, as each new word is encountered in the source text CrossS counts how many skips must be made in the target text (positive for skips right, negative for skips left) to reach the position of the corresponding target text word. When words align perfectly in source and target, the CrossS value will be 1. The absolute value |CrossS| will tend to be higher when the syntactic structures of the source and target languages require more re-ordering during translation. A more complete discussion can be found in Carl et al (2106a). Carl et al. (2016b) provide evidence that |CrossS| values are higher when the target language is more syntactically remote from the source language. In particular, |CrossS| values are almost double for English>Japanese than for English>Spanish when the same source texts are used.

HTra is a measure of semantic remoteness and |CrossS| is a measure of syntactic remoteness between source and target language segments. High values of either one are likely to be associated with less literal translations, since

the source text will need to be extensively reframed to be easily comprehensible. Carl et al. (2016a) introduced the Literality[1] feature into the CRITT TPR database as a way to simultaneously measure semantic and syntactic remoteness. At the word level,

$$Literality = HTra * |CrossS|;$$

at the segment level, Literality is computed as the average Literality of all words in the segment. Segment Literality will be lower when there is low variability from one translator to another and when translations of a source text segment differ little from straight word-by-word renderings. It might therefore be expected that translations of segments with higher levels of Literality will be achieved with higher expenditure of cognitive effort. We investigate this line of reasoning using the BML12 and ENJA15 subsets of the CRITT TPR database. When cognitive effort is measured by pause-word ratio (Lacruz and Shreve, 2014), the arguments outlined above suggest that high segment Literality will be associated with high cognitive effort in from-scratch translation and in post-editing for both English>Spanish and English>Japanese.

2. Pause-Word Ratio for Pause Ranges

Pauses in language production have long been associated with cognitive effort. (e.g., Schilperoord, 1996; Dragsted, 2004; Kumpulainen, 2015). Lacruz et al. (2012) and Lacruz and Shreve (2014) introduced pause metrics that have been associated with cognitive effort in a variety of translation tasks. The simplest is the Pause-Word Ratio (PWR), the ratio of the number of pauses in a segment to the number of words in the segment.

Correlations with other measures of cognitive effort mostly have not been sensitive to the precise pause threshold time (but see Schaeffer et al. (2016)), and commonly used thresholds have ranged from 300ms to 5000ms. However, Lacruz and Shreve (2014) observed a variety of pause patterns during post-editing and noted that clusters of short pauses sometimes accompanied particularly effortful edits. They suggested that these short pauses might be associated with monitoring during the post-editing process. By its nature, monitoring occurs relatively late in the cognitive process, so studying short pauses could give insight into later stages of the post-editing or translation processes.

Previous studies have focused on pauses whose lengths were above a certain threshold, specifically above 300ms, above 500ms, above 1000ms, above 2000ms, and above 5000ms. However, in order to isolate the effects of short monitoring pauses in the translation process and to understand better their influence on cognitive effort, we examine here segment level pause-word ratios in the BML12 and ENJA15 studies for pauses whose lengths fall into different time ranges, specifically 300-500ms, 500-1000ms, 1000-2000ms, 2000-5000ms, and at least 5000ms. Loosely speaking, these time ranges correspond to short, medium, and long pauses. Monitoring pauses are likely to be short, in the 300-500ms or 500-1000ms ranges.

[1] Literality is a counter-intuitive term, since highly literal translations can have a low Literality score.

Other studies have focused on the time course of the translation process (cf. Lacruz, 2017). Several have investigated the early stages (cf. Schaeffer et al., 2016). To date, however, little research has been carried out to focus specifically on later stages of the translation process.

There were noticeable differences in the distribution of these range PWRs between the BML12 and ENJA15 data, both for post-editing and for from-scratch translation.

2.1 Translation from Scratch

For the BML12 from-scratch translation data, there is a consistent pattern where the PWR values for a segment show a steady downward trend as the pauses become longer. However, the ENJA15 data consistently shows a different pattern, where the PWR values for a segment rise to a maximum for the 500-1000ms range and then follow a steady downward trend as the pauses become longer. In almost all ranges, the PWR values for English>Japanese translation are higher than those for English>Spanish. This is consistent with previous findings that translation is more effortful when the languages are more remote (Carl et al., 2016b; Schaeffer et al., 2016). For a sample segment, these patterns are illustrated in Figure 1.

It is notable that the shorter pauses, which we have argued are likely to be monitoring pauses, contribute significantly to the pause-word ratio for both language pairs. In other words, it appears that a considerable amount of the cognitive effort expended in translating from scratch is devoted to monitoring what is being produced.
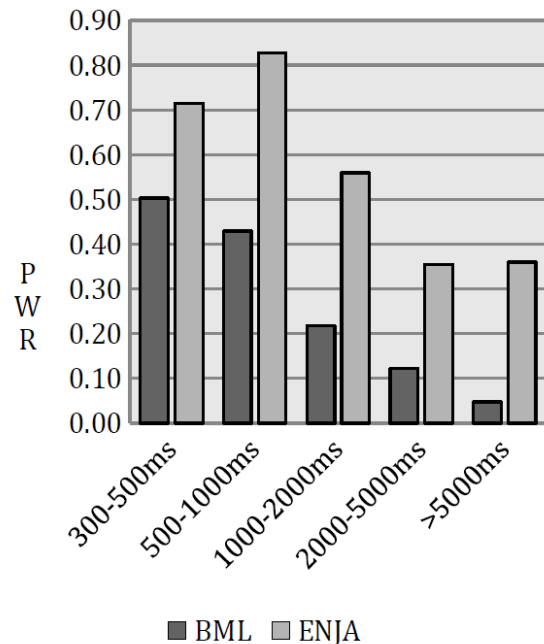

Figure 1: Translation PWR comparisons

2.2 Post-editing

The PWR patterns for post-editing share characteristics of those for from-scratch translation, except that in both language pairs PWR values for the 1000-2000ms pause range tend to be lower than for the longest pauses. Across almost all pause ranges in both language pairs, PWR values for post-editing are lower than for translation from scratch. This is consistent with previous findings that

post-editing is less cognitively effortful than translation from scratch (e.g., Green et al., 2013). The proportional discrepancy between English>Japanese and English>Spanish is more apparent for post-editing than for from-scratch translation, which suggests that the post-editing effort advantage might be less for English>Japanese. See Figure 2 for a representative visualization of the post-editing PWR patterns.
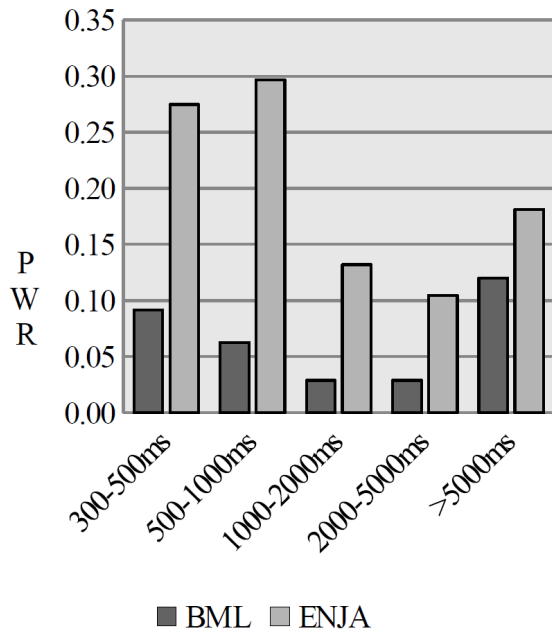

Figure 2: Post-editing PWR comparisions

As with translation from scratch, it is notable that the shorter pauses, presumed to be monitoring pauses, contribute significantly to the pause-word ratio for both language pairs. In other words, it appears that a considerable amount of the cognitive effort expended in post-editing is devoted to monitoring what is being produced. Also notable is the clear up-tick in very long pauses in post-editing as compared with translation from scratch. This is possibly due to the need for post-editors to completely rework areas that are unintelligible in the MT text.

3. Literality, HTra, and CrossS

We investigate more closely how the remoteness of one language from another could influence the cognitive effort involved in monitoring during from-scratch translation and post-editing. For this we focus on how Literality, HTra, and CrossS levels of source text segments correlate with PWR computed from the shortest pause range, 300-500ms., which we assume to measure monitoring effort.

3.1 Translation from Scratch

For translation from scratch, significant correlations ($p < .05$) are marked with a star in Table 1. As expected, Literality, which is a measure of semantic and syntactic remoteness is strongly and significantly correlated with monitoring effort for both English>Spanish and English > Japanese. This effect does not carry over to HTra, which we take as a measure of conceptual remoteness. For each language pair, there is no signifcant correlation between HTra and monitoring effort. The lack of correlation might

be attributable to the fact that semantic difficulties will have been largely resolved before translation production begins. The lack of distinction between language pairs breaks down for CrossS, which is a measure of structural remotenesss. For English>Japanese, there is strong and significant correlation between |CrossS| and monitoring effort. However, there is only a moderate, non-significant correlation for English>Spanish. This distinction between the language pairs is a natural one. Since Japanese is structurally remote from English, it seems likely that there would need to considerable on-line monitoring of language production during translation in order to be sure of maintaining structural integrity. On the other hand, monitoring would likely be less intense for English> Spanish, since greater structural closeness will promote more fluency in the translation process.

	Literality	HTra	|CrossS|
BML	.93*	.03	.51
ENJA	.86*	.21	.82*

Table 1: Pearson correlations with 300-500ms PWR. Translation from scratch

3.2 Post-editing

For post-editing, significant correlations ($p < .05$) are marked with a star in Table 2.

	Literality	HTra	|CrossS|
BML	-.93*	-.86*	-.91*
ENJA	-.05	.91*	.16

Table 2: Pearson correlations with 300-500ms PWRs Post-editing

The post-editing correlations are markedly different from the from-scratch translation correlations. For English> Spanish, all the remoteness indicators correlate significantly, but strongly negatively, with monitoring effort. At first sight this might seem strange. However, MT quality from English to Spanish is good, so the machine will successfully resolve most of the difficult semantic and structural problems. These resolutions, which will likely be highly salient to the post-editor and which can usually be accepted quickly, will likely reduce the need to expend much monitoring effort on segments that would have required much more effort to translate from scratch. On the other hand, machine translation solutions for segments where there are few semantic or structural problems to resolve may be much less salient to the post-editor. Out of conscientiousness, the post-editor may search for hidden difficulties, leading to unnecessarily high monitoring effort. This type of paradox has previously been observed in a translation decision task carried out by Lacruz (2017).

The very different correlations for English>Japanese, where the only strong and significant positive correlation is between semantic remoteness and monitoring effort might be attributable to lack of confidence in the less reliable machine translation programs for English>

Japanese leading to multiple monitoring comparisons between the source text and the machine translation. Structural proposals are more transparent than semantic proposals by the machine, so will require less effort to accept or reject, and errors will generally be easier to correct than semantic errors.

4. Conclusion

We introduced the idea of grouping production pauses of similar lengths in order to make a preliminary investigation of potential differences between the translation and post-editing processes when the source and target languages are closely related (English and Spanish) and when they are semantically and structurally remote (English and Japanese). The first main preliminary finding was a difference in the distribution of pauses of different lengths for Spanish and Japanese translations or post-editings of the same English source texts. We argued that this may have been mostly caused by differences at the level of short, monitoring pauses. The second main preliminary finding was that structural and semantic differences between languages have a differential effect on monitoring effort during from-scratch translation and post-editing. This was particularly striking for post-editing.

This preliminary work suggests that it may be interesting to carry out a more in-depth study, possibly using different language pairs, and also raises the possibility that it may be informative to examine different types of behavioral metrics obtained in various translation modalities using a similar time-range approach. For example, it may be instructive to investigate different ranges of gaze times in eye tracking experiments in order to identify the locus of cognitive effort associated with different parts of the translation process.

5. Bibliographical References

Carl, M., Schaeffer, M., and Bangalore, S. (2016a). The CRITT translation process research database. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research*. New York, NY : Springer, pp. 13-54.

Carl, M., Aizawa, A., and Yamada, M. (2016b). English-to Japanese translation vs. Dictation vs. Post-editing : comparing translation modes in a multilingual setting. In Nicoletta Calzolari (Conference Chair), et al., editors, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pages 4024-4031, Portorož, Slovenia, May 23-28. European Language Resource Association (ELRA).

Carl, M., Lacruz, I., Yamada, M., & Aizawa, A. (2016c). Measuring the translation process. Proceedings from *22nd Annual Meeting of the Society for Natural Language Processing (NLP 2016), Tohoku University, Sendai, Japan, March 7–10,* (pp. 957–960). Japan: Association for Natural Language Processing.

Dragsted, B. (2004). *Segmentation in translation and translation memory systems: An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process*. Copenhagen: Samfundslitteratur.

Dragsted, B. (2012). Indicators of difficulty in translation - Correlating product and process data. *Across Languages and Cultures* 13(1): pp. 81–98.

Green, S., Heer, J., and Manning, C.D. (2013). The efficacy of human post-editing for language translation. In Wendy E. Mackey (Conference Chair) et al., editors, Proceedings of the SIGCHI conference on human factors in computing systems (SIGCHI 13), pages 439-448, Paris, France, April 27 – May 2. Association for Computing Machinery (ACM).

Kumpulainen, M. (2015). On the operationalisation of 'pauses' in translation process research. *Translation & Interpreting*, 7(1): 47-58.

Lacruz, I., Shreve, G. M., and Angelone, E. (2012). Average pause ratio as an indicator of cognitive effort in post-editing: A case study. In S. O'Brien, M. Simard, & L. Specia (Eds.), Proceedings of AMTA 2012, 10th Conference of the Association for Machine Translation of the Americas, Workshop on Post-Editing Technology and Practice (WPTP 2012), pages 1-10, San Diego, CA, October 28–November 1.

Lacruz, I., and Shreve, G. (2014) Pauses and cognitive effort in post-editing. In S. O'Brien, L. Balling, M. Carl, M. Simard & L. Specia (Eds.), *Post-editing: Processes, technology and applications*. Newcastle upon Tyne (UK): Cambridge Scholars Publishing, pp. 246 – 272.

Lacruz, I. (2017). Cognitive effort in translation, editing and post-editing. In J. Schwieter & A. Ferreira (Eds.), *Handbook of translation and cognition*. Blackwell Handbooks in Linguistics. Malden, MA: John Wiley & Sons, pp. 386-401.

Lacruz, I. (2018). An experimental investigation of stages of processing in post-editing. In I. Lacruz & R. Jääskeläinen (Eds.), *Innovation and expansion in translation process research*. Volume XVIII of the American Translators Association Scholarly Monograph series. Amsterdam/Philadelphia: John Benjamins, pp. 217-240.

Mesa-Lao, B. (2014). Gaze Behaviour on Source Texts: An Exploratory Study Comparing Translation and Post-editing. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard, & L. Specia (Eds.) *Post-editing: Processes, technology and applications*. Newcastle upon Tyne (UK): Cambridge Scholars Publishing, pp. 219-245.

Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., and Carl, M. (2016). Word translation entropy : Evidence of early target language activation during reading for translation. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research*. New York, NY : Springer, pp. 183-210.

Schaeffer, M., Carl, M., Lacruz, I., & Aizawa, A. (2016). Measuring cognitive translation effort with activity units. *Baltic Journal of Modern Computing,* 4(2) : pp. 331–345.

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production* (Vol. 6). Amsterdam: Rodopi.