Characterization of Human Translation for Different Levels of Expertise

Pascual Martínez-Gómez[1] **Akiko Aizawa**[1] **Michael Carl**[2]
[1]The University of Tokyo
[1]National Institute of Informatics
[2]Copenhagen Business School
{pascual,aizawa}@nii.ac.jp mc.isv@cbs.dk

1 Introduction

Despite of their lack of precision, machine translation is widely used to satisfy a wide array of communication needs. These needs are characterized by their spontaneity and low requirements on translation quality, such as judging the relevance of e-mails in our inbox or navigating websites in unfamiliar languages. However, there are other applications that require precise translations, such as patent review processes or translation of legal documents. In these cases, we need human intervention to meet the highest standards in translation quality that certify the semantic preservation of translated documents with respect to their original counterparts.

Human translation involves an important amount of effort, and Computer Assisted Translation (CAT) was developed to help humans in this laborious task. Modern CAT systems (Barrachina et al., 2009) receive sentences from a source language and suggest possible translations to supervising human translators. Then, translators may accept (part of) the translations, or amend them. It is thus reasonable that these systems are evaluated according to the effort they save to the supervising humans.

Translation effort is often measured as the ratio of keystrokes, mouse clicks (Barrachina et al., 2009) or e-pen strokes (Toselli et al., 2011) (with respect to the length of the sentence) that are necessary to transform automatic translation suggestions into the final translation that the supervising human wishes to achieve. For practical reasons, these measurements are estimated from human translation simulations. However, these metrics do not take into account other human activities that are involved during the translation process, such as understanding the source sentences, or checking the adequacy and fluency of the translated sentences. These activities are difficult to observe because there are no explicit actions to record from human translators. Moreover, translators with different levels of expertise may have different strategies in translation, but CAT systems assume homogeneous distributions of preferences and translation styles.

Our objective is to better understand how translators spend their time and what are the perception and motor differences between translators with varying levels of expertise. Thus, our hypothesis can be stated as:

Hypothesis: Translators have different perceptual and motor activities, depending on their level of expertise.

To test this hypothesis, we carry out a systematic and quantitative analysis of human translation activities during real translation tasks. We hope this analysis may inform us on how to develop more effective CAT tools that maximize the impact on translator effort reduction.

2 Methodology

Translators and translation sessions can be characterized from many perspectives and data granularities. In our analysis, we characterize them according to:

- the activities translators engage in, and
- the characteristic patterns of translator's eye-movements.

When characterizing translators according to their activities, our observations will consist of *sets* of eye movement and typing events. First, we will compute the proportion of time on:

- reading sentences from source language,
- reading sentences from target language,
- producing character insertions, and
- producing character deletions.

Then, we will investigate whether there are differences in these proportions between novice and expert translators. Within the community of translation process research, there is no agreement on the exact definition of what translation activities are. However, it is well known that translators may engage in concurrent activities that involve different subsets of the activities described above. The activity combinations that we consider in this second analysis are:

1. Source text reading.
2. Target text reading.
3. Source and target text reading.
4. Target text typing.
5. Target text typing and source text reading.
6. Target text typing and target text reading.
7. Target text typing, source text reading and target text reading.
8. Translator idle.

For our analysis, we will automatically segment eye movement and keystroke events from translation sessions into a sequence of these translation activities, investigate in what proportions these activities occur, and whether there are differences between novice and expert translators.

When characterizing translators according to their eye movements, we will focus on two types of events, namely *fixations* and *saccades*. Fixations are periods of time when eyes remain still at a fixed location, and are used mainly for lexical recognition. Saccades are rapid eye movements that are used to locate the gaze point at a different text area. Fixation durations and saccade lengths are known to reflect different aspects of attentional effort and cognitive processes (Rayner, 1998), and we believe they are different for translators with different levels of expertise. Following standard practice in psycholinguistics, we will test these differences using Analysis of Variances (ANOVAs).

3 Experiments

The data in our analysis consists of translation sessions in TPR-DB database (version 2)[1] (Carl, 2012a), that have been recorded using Translog-II (Carl, 2012b). Translation sessions in this dataset contain the temporal sequence of keystrokes and eye movements that translators produce when translating a text from a source language into a target language. There are several indicators of translators' expertise. We use the information on whether translators are certified or not, that is, whether they have received formal authorization to work as translators or interpreters. There were 204 sessions that had information about whether the translator was certified or not, where 99 sessions were produced by 47 non certified translators, and 105 sessions were produced by 47 certified translators.

The interface of Translog-II consists of two main windows arranged vertically. The window on the top contains several sentences of a passage in a source language. The window on the bottom is initially empty, and translators need to type there the translation of the text appearing in the top window. Since Translog-II is capable of recording the (x, y) coordinates and the duration of every fixation, we are able to record perception activities on the source or target text.

3.1 Translation Activities

Regarding general perception activities, translators spent 14% of their time fixating on the source text, and between 17% to 37% of their time fixating on the target text. There were significant differences between certified and non-certified translators regarding to the proportion of time they spent fixating the target text ($\mathrm{F}(1, 202) = 84.52, \mathrm{MSE} = 2.05, p < .001$). However, there were no significant differences in the proportion of time translators spent fixating the source text between certified and non-certified translators (see Figure 1).

Regarding motor activities in the form of keystrokes, translators spent between 35% and 42% of their time inserting characters, and around 4% of their time deleting characters. There were no statistically significant differ-

[1] Available at http://bridge.cbs.dk/platform/?q=node/18

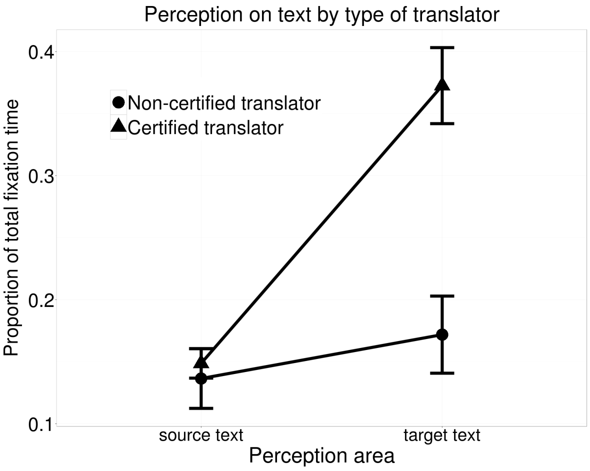

Figure 1: Proportion of translation session fixating source and target text. Certified translators spent larger proportion of time fixating the target text, compared to non-certified translators ($p < .001$).

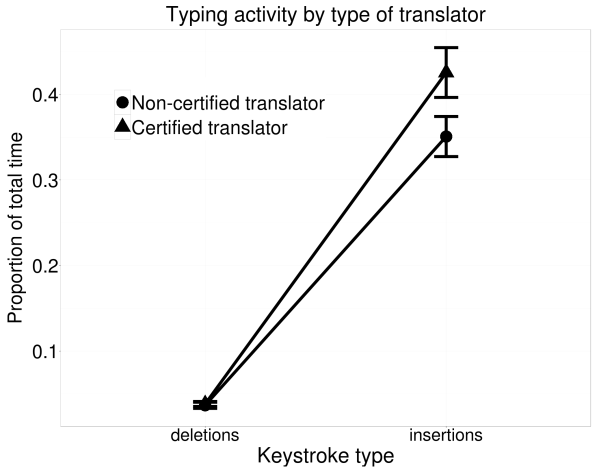

Figure 2: Proportion of translation session inserting and deleting characters. Certified translators spent larger proportions of time inserting characters, compared to non-certified translators ($p < .001$).

ences in the proportion of time that certified and non-certified translators spent on deleting characters, but there were significant differences in proportion of time inserting characters ($\mathrm{F}(1, 202) = 15.81, \mathrm{MSE} = 0.28, p < .001$). Figure 2 displays differences in proportion of time producing character insertions between certified and non-certified translators.

We also analyzed the proportion of time that translators spent on each of the 8 activities described in Section 2. Translators spent between 45% to 65% of their time on activity 7, which is a combination of concurrent activities "target text typing", "source text reading" and "target text typing". Certified translators spent significantly larger proportions of time on activity 7 than non-certified translators ($\mathrm{F}(1, 202) = 38.81, \mathrm{MSE} = 1.80, p < .001$). Translators spent around 14% of their time on activity 6, which is a combination of concurrent activities "target text typing" and "target text reading". There were no statistically significant differences between certified and non-certified translators regarding to this activity. Translators spent between 5% and 15% of their time in activity 8, which we labeled as "translator idle". This activity does not necessarily reflect that the translator was not doing anything, but rather that there were no fixation or keystroke events recorded by Translog-II during a certain period of time. Such phenomenon may occur when the translator is gazing away from the screen without producing keystrokes, sometimes when drafting mentally a translation or searching for appropriate terminology. We found that non-certified translators engage in activity 8 for significantly larger proportions of time than certified translators ($\mathrm{F}(1, 183) = 39.88, \mathrm{MSE} = 0.49, p < .001$).

3.2 Eye movement characteristics

We analyzed differences in fixation durations and saccade lengths between different translation activities and types of translators. Fixations had an average duration of 210 milliseconds and a median of 133 milliseconds. There were no significant differences between certified and non-certified translators, and there were small but statistically significant differences across translation activities ($\mathrm{F}(7, 78) = 6.61, \mathrm{MSE} = 6.5e7, p < .001$). Under the assumption that fixation duration reflects attentional effort, this suggests that different translation activities have different demands of attentional effort from translators. Regarding the length of saccades, certified translators had significantly longer saccades than non-certified translators ($\mathrm{F}(1, 78) = 11.79, \mathrm{MSE} = 7.8e6, p < .001$), which may evidence higher document navigational skills or a higher degree of parallel processing by certified translators.

4 Discussion and Future Work

We have observed that translators spent from 39% to 46% of their time in typing activities, whereas they spent between 31% to 51% of their time in perception activities (reading the source or target text). Post-edition and interactive machine translation systems strive to reduce translator effort as measured by the keystroke ratio. As we observed, typing (keystroke) activities only correspond to roughly half the time of a human translation process. Under the assumption that current CAT tools only minimize typing effort, the upper bound in translation time reduction would be 50%. However, it remains unclear whether CAT tools are also capable of speeding the mental formulation of translations by humans, since phrase suggestions may also reduce workload of translators.

We have also observed that certified translators spent considerably larger proportions of time reading the target text, and larger proportions of time on inserting characters. One explanation is that certified translators impose on themselves higher standards of fluency and adequacy, which leads them to spend larger proportions of time on reviewing or revising the resulting translations. An alternative explanation is that certified translators have lower typing skills and spend larger proportions of time monitoring their own typing activities to detect typos.

The most common activity that translators engaged in was activity 7, which corresponds to the concurrent combination of "source text reading", "target text reading" and "target text typing". However, moving the attentional focus from the source to the target text (and vice versa) may lead to occasional corrective saccades (Prablanc et al., 1978), which are eye-movements to correct the location of the landing fixation towards the intended fixation location. Carefully designed user interfaces that minimize the distance between the source and the target passage may contribute to the reduction of corrective saccades and their associated cognitive disruption. A possible interface design could consist of an initial display containing the full source passage, where sentences can be read in context. Then, the view mode could be changed to satisfy the necessities associated to the drafting phase, where source sentences are displayed above or below the text box for their translation.

5 Conclusions

We first analyzed the proportions of time in translation sessions that humans dedicate to source text reading, target text reading and target text typing. We found that translators engage 14% of their time in source text reading, between 17% to 37% in target text reading, between 35% to 42% inserting characters and 4% deleting characters. Certified translators spent significantly larger proportions of time in target text reading and target text typing than non-certified translators. The most common translation activity was the concurrent combination of "source text reading", "target text reading" and "target text typing", which occurred around 45% of the time for non-certified translators and 65% of the time for certified translators. There were no significant differences in fixation durations between certified and non-certified translators, but certified translators showed longer saccade lengths than non-certified translators.

References

S. Barrachina, O. Bender, F. Casacuberta, J. Civera, E. Cubel, S. Khadivi, A. Lagarda, H. Ney, J. Tomás, E. Vidal, and J.M. Vilar. 2009. Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1):3–28.

M. Carl, 2012a. *The CRITT TPR-DB 1.0*, pages 9–18. Association for Machine Translation in the Americas (AMTA).

M. Carl. 2012b. Translog-II: a program for recording user activity data for empirical reading and writing research. In *LREC*, pages 4108–4112.

C Prablanc, D Masse, and JF Echallier. 1978. Error-correcting mechanisms in large saccades. *Vision research*, 18(5):557–560.

K. Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422, November.

A. H. Toselli, E. Vidal, and F. Casacuberta. 2011. *Multimodal interactive pattern recognition and applications*. Springer.