Measuring the Translation Process

Michael Carl[1] Isabel Lacruz[2] Masaru Yamada[3] Akiko Aizawa[4]
1 National Institute of Informatics and Copenhagen Business School
2 Kent State University
3 Kansai University
4 University of Tokyo and National Institute of Informatics

1 Alternatives to Rule-based Translation

By the mid-1980s it was clear that unrestricted high quality machine translation would not be achievable in the foreseeable future and alternative directions to the then dominant rule-based paradigm were proposed. The appropriate level of linguistic representation was difficult to determine, hard to compute; and translation relations were incomplete, error prone and time consuming to formalize within the current-state rule-based translation formalisms. At the same time, with the upcoming availability of Personal Computers (PCs), more translations were produced in electronic form. As translators produce translations daily, implicitly solving those translation problems that are so hard to formalize, Isabelle (1992) said that “Existing translations contain more solutions to more translation problems than any other existing resource.“ New horizons for using MT were thus sought, which led to a number of different paradigms, some of which are briefly described.

1.1. The Translators Amanuensis

Martin Kay (1980) suggested a Translators Amanuensis, a pragmatic, incremental approach to joining human and machine to address the problem of how machines should be used in language translation. Humans and machines would collaborate to produce not only a translation of the text but also a device whose contribution to the translation would be constantly enhanced. The system would thereby only accumulate those solutions that had been agreed upon between both humans and the machine, which would at each stage do only what was agreed upon and was known could be done reliably. Kay suggested three steps for the development of the Translators Amanuensis which would build on each other and incrementally become more complex

1. Text Editing (bilingual text editor, basic operations)
2. Translation Aids (dictionary access, special marker, morphological generation)
3. Machine Translation

Thirty years later, this vision has come true in the form of computer assisted translation (CAT) and Translation Memories, which are widely in used in the translation industry.

1.2. Statistical Machine Translation (SMT)

Brown et al. (1988) suggest a fully automatic translation approach, which would be based on the “view that every sentence in one language is a possible translation of any sentence in the other language. We assign to every pair of sentences (e,f) a probability P(e | f) ... the probability that a translator will produce e in the target language when presented with f in the source language.” Bilingual corpora would be used to train these probabilities and Bayes theorem would justify a segmentation into a translation model and a language model. By now SMT has become the main-stream MT paradigm and it has undergone a number of modifications and extensions, from word-based, over syntax-based to phrase-based SMT, with still a very active research community.

However, as Brown et al (1993) point out “[a]s a representation of the process by which a human being translates a passage from French to English, this [Bayes] equation is fanciful at best. One can hardly imagine someone rifling mentally through the list of all English passages computing the product of the a priori probability of the passage, P(e), and the conditional probability of the French passage given the English passage, P(f | e)”

1.3. Example-Based Machine Translation (EBMT)

Nagao (1984) suggests a cognitive approach to translation, which would mimic the human translation process. As a reaction to the then predominant rule-based translation paradigm, he states that “[m]an does not translate a simple sentence by doing deep linguistic analysis, rather, [...] first, by properly decomposing an input sentence into certain fragmental phrases ..., then by translating these phrases into other language phrases, and finally by properly composing these fragmental translations into one long sentence.” Based on this model, a large number of different computer systems have been developed.

As a simulation of the human translation process, and in line with many earlier models of the human translation process (Nida, 1964; Seleskovitch, 1975), these systems assume that translators proceed:

- Sequentially, in a sentence-by-sentence mode
- Stratificationally, in three steps by decomposing – translating – recomposing

However, none of the EBMT systems is based on actual empirical investigations of the human translation process. The sequential and stratificational nature of the human translation process is taken for granted. However, within the past 20 years, empirical translation process research has developed innovative methods and generated knowledge which suggest a more complex picture.

2 Empirical Translation Process Research

While earlier methods of translation process research relied on introspection and the evaluation of retrospective or think aloud protocols (TAP), more recent computational methods allow for collecting objective user activity data (UAD). Keystroke-logging and eye-gaze tracking make it possible to track translation behavior in detail and in a non-invasive manner, to play back translation sessions and to compare and quantify similarities and differences across different translators, different languages, different texts, etc. Jakobsen (2002) distinguishes three phases in the translation process: 1) initial orientation, in which the translator gets a gist of the source text, 2) translation drafting in which a raw version of the translation is produced and 3) translation revision, in which the draft is refined and finalized. By comparing the translations of 12 professional and 12 student translators, Carl et al. (2011) find that translation students and professional translators show different behaviour in all three phases. Student translators have a systematic initial orientation, they prefer large-context planning during drafting and often skip the revision phase. Professionals are more frequently head-starters, they show small-context planning during drafting and often have an extensive end-revision.

However, both groups show sequences of challenged translation production which requires a large amount of conscious effort and un-challenged translation production which is probably triggered by automatized routines.

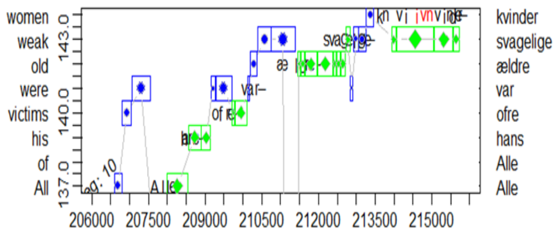

Figure 1: Example of an undisturbed translation progression

2.1. Automatized Translation Production

Figure 1 shows a visualization of an undisturbed translation progression. The English source sentence “All of his victims were old weak women“ (left) is translated into Danish “Alle[i] hans ofre var aeldre svagelige kvinder” (right). Translation activities are shown in the graph on a timescale, from 206,000 to 215,000 ms. The overall translation takes approximately 9 seconds. The figure shows keystrokes (insertions and deletions) reading patterns on the source text (blue rectangles) and monitoring of text production (green rectangles) in time. There are relatively few fixations on the source text words (blue boxes with dots), immediate translation typing activity and close monitoring of the target text production (green boxes with diamonds). It is likely that the production of this is highly automatized and translation activities occur early.

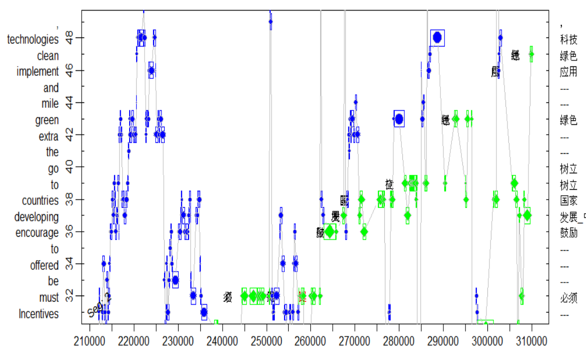

Figure 2: Progression graph with complex monitoring patterns

[i] “Alle” occurs twice in Fig. 1 as it is a translation of “All of”

2.2. Conscious Translation Effort

Figure 2 shows an excerpt from a translation session English → Chinese, with much more complex patterns of monitoring behavior, repeated regressions, re-reading, backtracking, deletions, revisions, etc. The production of this translation segment of 17 words took approximately 100 seconds, which is almost 5 times longer per word than the Danish translation in Figure 1. The ST segment “the extra green mile” was read at least 7 times, four times during an orientation phase between seconds 210 and 240 and then again three times during translation drafting. While it is likely that a large amount of conscious processing has taken place during the translation production of this segment, there are currently no fine-grained measures available to capture and describe the hidden translation processes appropriately and it is unclear how this observed behavior relates to translation strategies as isolated and described by TAP researchers.

3 Measuring the Translation Process

Keystrokes, mouse clicks, gaze data and other user activities constitute the directly observable translation behavior which can be recorded and quantified, and which reflect traces of the hidden translation processes. However, it is neither obvious how keystrokes and gaze data should be organized so as to constitute some sort of “observable states” nor is it uncontroversial what are the hidden states that are presumed to give rise to the recorded observations.

Within the CRITT TPR-DB (Carl et al. 2016) translators’ behavioral activity data is post-processed and a number of measures are implemented that render the observations quantifiable. As shown in Figures 1 and 2, a translation session consists of translation events (keystrokes and gaze fixations) which have, on the one hand, temporal aspect (horizontal axis) and, on the other hand, textual aspect as they contribute to the production of the target text (vertical axis). An appropriate analysis of the translation process should thus consider both the textual and the temporal dimensions. Current approaches accumulate UAD either in a window on the vertical (textual) or a window on the horizontal (temporal) axis.

	Reading measures	**Writing measures**	**R&W measures**
earlier processes	First-Fixation Duration	Single Keystrokes	Eye-key span
↓	First Pass Reading time	Production Time	
later processes	Regression Path Duration	Short Distance Revisions	Parallel R&W activity
	Total Reading Time	Revisions, Inefficiency	

Table 1: Textual Translation Process Measures

3.1. Textual Translation Process Measures

There is a large repository of linguistic concepts to describe textual segments, such as words, phrases or sentences, and to annotate the relationships between them. Textual process measures accumulate behavioral data that contribute to the perception and/or production of these segments which are also referred to as Areas of Interest (AOI) in the eye-tracking literature. As shown in Figures 1 and 2, a distinction can be made between behavioral measures which indicate early and late(r) processes, according to when, how often or how long an activity was observed on the items in the AOI.

Table 1 summarizes some of these measures. It distinguishes between reading measures, which capture gaze activities, writing measures, which describe typing processes, and R&W measures, which capture the relationship between reading and writing behavior. The first pass reading time, for instance, is the sum of fixation durations on a word (or another predefined text segment) starting from the first fixation before the eyes leave the AOI again. The Word Production Time is the total time needed to type a word (i.e. a translation), including all its possible revisions.

3.2. Temporal Production Measures

There are comparatively few approaches to fragmenting the translation process data. Available methods fragment the behavioral data based on pauses in the flow of keystrokes and/or fixation location on the source or target text. There is no consensus on an appropriate threshold for the duration of the pause which can last for anything between 300 ms and 5 seconds. A Production Unit (PU) in the TPR-DB, for example, is a coherent sequence of keystrokes where the lapse of time between successive keystrokes is below a given threshold, e.g., 1 sec. A PU can thus and contain a single or a large number of keystrokes irrespective of the number of words produced. Activity Units fragment the process data into sequences of source text reading, target text reading, typing or concurrent activity. It is sometimes assumed that longer pauses between successive keystrokes signal higher cognitive effort. Immonen (2006) finds that in translation, pause length is higher at word and clause boundaries. O'Brien (2006) suggests that analyzing pauses in post-editing is a useful tool to measure cognitive effort in post-editing. Lacruz et al. (2014) introduced the pause-word ratio as a more refined metric to measure cognitive effort in post-editing.

3.3. Textual-temporal Translation Process Measures

To date, no measures have been developed which would combine both the textual and the temporal aspects of translation production. For instance, the keystroke pausing segment of 25 seconds in Figure 2 between (roughly) time stamps 21000 and 235000 consists only of source text fixations, in which parts of the source words are repeatedly read. Available textual translation process measures (e.g., total reading time) would map accumulated fixation durations on the source or target language words (i.e. horizontally in Figure 2), irrespective of when they occur in time. Temporal production measures would qualify this segment as a source text reading segment, regardless of the order in which the words are fixated. However, in order to better understand the behavioral patterns of different translation strategies, to empirically assess their similarities and differences, and to obtain a better understanding of the human translation process, we need to develop and evaluate textual-temporal translation process measures and combine these with the available ones.

4 Conclusion

Pauses in the translation production process (i.e. gaps in typing activities) have been associated with cognitive meta-activity and pause analysis has been proposed as a method to detect the amount of 'cognitive effort' in translation. However, it is unclear what exactly the cognitive processes are that take place during keystroke pauses and it is an unsolved problem to determine what it is exactly that makes pauses more or less effortful. Recorded gaze data which fills the typing pauses could make it possible to "identify the specific motivation of a particular pause" (Kumpulainen, 2015: p 47). However, only a very limited number of tools have been developed and very few measures exist to date that are suited to analyze, identify and classify gaze patterns and to relate them to translation problems and translation strategies. Temporal-textual translation process measures are required to obtain a comprehensive picture of the human translation process and to eventually arrive at better anticipation of and assistance for human behavior during computer interaction in translation.

References

Brown, P. F., Pietra, S. A. Della, Pietra, V. J. Della, & Mercer, R. L. (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics, 19(2), 263–312.

Brown, P., Cocke, J., Pietra, S. D., Jelinek, F., Mercer, R., & Roossin, P . (1988). A statistical approach to language translation. COLING-88

Carl, Michael and Dragsted, Barbara and Lykke Jakobsen, Arnt. 2011 A Taxonomy of Human Translation Styles. In: Translation Journal, Vol. 16, Nr. 2, 2011

Carl, Michael; Moritz Schaeffer; Srinivas Bangalore. 2016. The CRITT Translation Process Research Database. In: New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB. ed. Michael Carl; Srinivas Bangalore; Moritz Schaeffer. Cham : Springer, p. 13-54

Immonen, Sini 2006: Translation as a Writing Process: Pauses in Translation versus Monolingual Text Production. In Target 18 (2), 313-335

Isabelle, P. (1992). Bi-Textual Aids for Translators. In Proc. of the Annual Conference of the UW Center for the New OED and Text Research (pp. 76–89).

Jakobsen, Arnt Lykke. 2002. "Translation Drafting by Professional Translators and by Translation Students." In Empirical Translation Studies. Process and product. Edited by Gyde Hansen, 191-204. Copenhagen: Samfundslitteratur

Minna Kumpulainen. 2015. On the operationalisation of 'pauses' in translation process research. Translation & Interpreting Vol 7 No 1 (2015) http://trans-int.org/index.php/transint/article/download/367/183

Lacruz, Isabel and Shreve, Gregory M. 2014. Pauses and Cognitive Effort in Post-Editing. In Post-editing of Machine Translation: Processes and Applications. Edited by Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia, Cambridge Scholar Publications

Nagao, M. (1984). A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle. In Proc. Of the International NATO Symposium on Artificial and Human Intelligence (pp. 173–180). New York, NY, USA: Elsevier North-Holland, Inc.

Nida, E. A. (1964). Toward a science of translating: with special reference to principles and procedures involved in Bible translating. Leiden: E.J. Brill.

Seleskovitch, D. (1975). Langage, langues et mémoire : étude de la prise de notes en interpretation consécutive. Paris: Minard Lettres Modernes.