

The CRITT TPR-DB 1.0: A Database for Empirical Human Translation Process Research

Michael Carl

Institute for International Business Communication
Copenhagen Business School,
2000 Frederiksberg, Denmark
mc.ibc@cbs.dk

Abstract

This paper introduces a publicly available database of recorded translation sessions for Translation Process Research (TPR). User activity data (UAD) of translators behavior was collected over the past 5 years in several translation studies with Translog¹, a data acquisition software which logs keystrokes and gaze data during text reception and production. The database compiles this data into a consistent format which can be processed by various visualization and analysis tools.

1 Introduction

Human translation process research (TPR) is a branch of descriptive translation studies (Holms, 1972) which analyzes the translation behavior of translators, such as types of units that translators focus on, conscious and unconscious translation processes, differences in expert and novice behavior, memory and search strategies to solve translation problems, etc. It seeks to identify the temporal (and/or contextual) structure of those activities and describes inter- and intra-personal variation. Various models have been developed that seek to explain translators' behavior in terms of controlled and uncontrolled workspaces (Göpferich, 2008), and monitor models (e.g. Tirkkonen-Condit, 2005) with trigger micro- and

macro-translation strategies. However, due to the lack of appropriate data and tools, only few attempts have been made to ground and quantify translation process models in empirical user activity data (UAD).

In order to close this gap, this paper introduces a database of translation process data which was collected over the past 5 years with Translog¹. More than 450 translation sessions were recorded in 10 translation studies and converted into a common format (Carl and Jacobsen, 2009). The database is now publicly available, together with a toolkit for analysis and visualization: as described in Carl and Jacobsen, (2009), the UAD consists of product and process components which are processed in different components in the CRITT TPR-DB². A) We used the NLTK (Bird, 2009)³ for automatically POS tagging and lemmatization. B) In addition, the product data can be converted into treex format and visualized/annotated in TrEd⁴. C) The CRITT TPR-DB provides several tools to manually check and amend the automatic annotations. D) The product and process data is integrated by mapping keystrokes and fixations on the produced TT tokens (Carl, 2012) and via the alignment on the corresponding ST equivalents. This allows us to extract various different types of product and process units from the UAD and to mutually correlate the product and the process data. Translation sessions can thus be visualized in

¹ The translog website is www.translog.dk. The most recent version of Translog-II can be obtained for free for academic purposes from the author.

² CRITT (www.cbs.dk/en/CRITT) is the "Center for Research and Innovation in Translation and Translation Technology" at Copenhagen Business School. We refer to the UAD database as CRITT TPR-DB.

³ NLTK is a Python platform to work with human language data: <http://nltk.org/>

⁴ TrEd is a programmable graphical editor and viewer for tree-like structures: <http://ufal.mff.cuni.cz/tred/>

the form of translation progression graphs (Carl and Jacobsen, 2009) or statistically analyzed e.g. with R⁵.

In this paper we give a short introduction to translation process research and the data that we obtain from Translog. We describe the structure of the CRITT TPR-DB and the origin/intention of the various studies it contains. We will then describe how the raw logging data is compiled into a database structure which allows for more detailed analysis and evaluation of the translation processes. While much of this compilation is fully automatized, the database design also contains a number of tools to manually adjust the annotations. Finally we give an overview of the Metadata that is stored with the CRITT TPR-DB.

2 Empirical TPR with Translog

While in the beginnings of TPR, user activity data (UAD) could only be elicited via traditional methods of introspection such as questionnaires, think-aloud experiments (TA) or retrospection (Krings, 1986; Lörscher, 1992; Tirkkonen-Condit & Jääskeläinen, 2000), computer-based analysis techniques have been applied in empirical translation studies for about 15 years.

Around the 1990s, most texts and most translations were typed on computer keyboards, and software was developed to log the writing process (all keystrokes, pauses and changes), for example ScriptLog (Holmqvist et al, 2002), Proxy (Pacte group), Translog (Jakobsen and Schou, 1999 and Inputlog (Leijten/Van Maes, 2006)). This can be regarded as the beginning of digital translation process research (DTPR). With these tools a complete log can be created of all the keystrokes made in producing a text, including typos, pauses, deletions, changes, mouse clicks, cursor movements. Several larger translation process projects were carried out with keystroke logging combined with retrospection and post-process dialogues.

Since 2006 CRITT⁶ has developed a data acquisition software, Translog (Jakobsen and

Schou, 1999, Carl 2012) with which translators' keystroke and gaze activities can be recorded⁷. This tool is now the most widely used tool of its kind (Jakobsen, 2006).

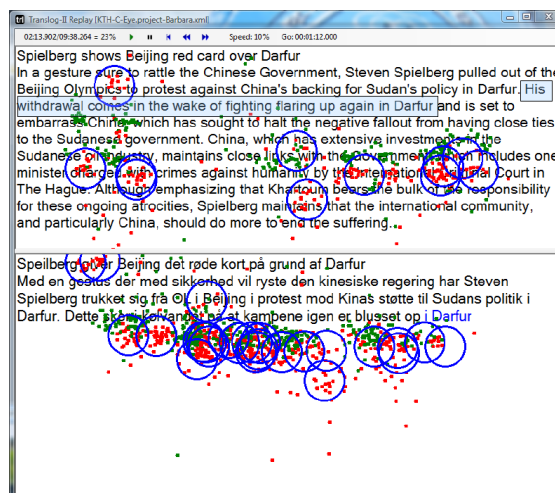


Figure 1: Screenshot of Translog-II replay: fixations in blue circles

As shown in figure 1, Translog separates the screen into two windows: the source text is shown in the upper window while subjects type a translation into the lower window. Figure 1 also shows the accumulations of gaze fixations (in blue) during the time span in which a translator reads the beginning of the source language sentence “China which has extensive investments in the Sudanese oil industry, maintains close” and begins producing (i.e. typing in) its translation.

Translog-II can be used to record reading and writing activities, as well as sessions of post-editing and revision. For post-editing (e.g. of MT output), the translation session can be prepared in such a way that the translation to be revised appears in the lower window of the screen while the upper window contains the original source text. Writing studies would be initiated by preparing Translog-II to show only the lower window, and reading experiments would plot only the upper window. In a similar way, a revision (or editing) scenario of a text without a source can be produced by plotting the lower (write enabled) window with

⁵ R is a free software environment for statistical computing and graphics. It can be downloaded from <http://www.r-project.org/>

⁶ CRITT aims at building up new knowledge of translation and communication processes and provide a basis for technological innovation in this field.

⁷ Translog-II has interfaces to Tobii eye-tracker; a connection to eye-link 1000 is currently being implemented.

a pre-defined text. Note that the screen can also be divided in a vertical manner.

3 Translation Process Database

CRITT has collected over the past 5 years a substantial amount of translation process data from numerous translation sessions. The analysis of this data has given rise to more grounded translation models and an extended understanding of the underlying human translation processes (Mees and Göpferich, 2009, Göpferich, Jakobsen, Mees, 2009; Göpferich, Alves, Mees, 2010).

As the collected UAD was recorded with various Translog versions producing different logging formats, the data has been converted into one

In each session, a translator had to translate (T), post-edit (P), Edit (E) or copy (C) a source text. In the case of post-editing, MT output was shown in the target window, and in the case of editing the MT output was shown without the source text (monolingual editing of MT output). A total of 19 different source texts were used in these studies, so that there are on average 24 translations per text. Table 1 shows the distribution of translations for each source text. While some texts (Text1, Text2, Text3 and Text8) have been translated more than 50 times into various languages and have been re-used in several translation studies, other texts are translated only few times. Text12, Text13, Text14 and Text15 are only used in one study and have been translated only by 2 and 3 translators

Table 1: Distribution of recordings per Study and ST in the CRITT TPR-DB V1.0: lines represent different Studies, rows different source texts

Study Text	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
ACS08				14	16	15	15													60
BD08								10												10
BML12	9	11	10					10										10	10	60
JLG10												2	3	2	3	5	5			20
KTHJ08	24	24	23																	71
LWB09									12	14	14									40
MS12	3	9	7					10										8	7	44
NJ12	15	19	14					17										18	17	100
SG12	6	5	5					6										5	5	32
TPR11	10		9																	19
Total translations	67	69	67	14	16	15	15	53	12	14	14	2	3	2	3	5	5	41	39	456

consistent data format (Carl and Jakobsen, 2009) and annotated with Metadata (Jensen and Carl, 2012). In addition, more than 230 translation sessions were recorded in the past year to complement the legacy TPR UAD with more target languages and with post-editing sessions. In its current version, the CRITT TPR-DB consists of 10 translation studies which amount to a total of 456 (translation) sessions, distributed as follows:

- T:** 257 Translation (from scratch)
- P:** 129 Post-editing
- E:** 40 Editing
- C:** 30 Text Copying

respectively. Each source text is between 100 and up to 236 words in length and designed in a way such that it fits on one Translog screen (to avoid scrolling). 13 of the 19 source texts are English, and two translation studies, JLG10 and LWB08, use respectively Portuguese and Danish source texts to be translated into English. Some of the source texts only differ in few words, as they seem to be slightly modified in some experiments. With respect to the target languages, the CRITT TPR-DB is more varied than with the source languages, with a total of 7 different target languages. The table 2 shows the distribution of translation, post-editing, editing and copying experiments together with the respective source

and target languages. Note that the source language is also given in the editing experiments (even though the text was not visible for the editor) and that copying experiments have identical source and target languages.

Table 2: Distribution of recordings with respect to source and target language and type of session.

Source	Target	T	P	E	C	Total
en	da	111				111
en	hi	39	61			100
en	es	20	20	20		60
en	zh	15	19	10		44
en	de	12	19	10		41
da	en	40				40
en	en				30	30
en	pt	10	10			20
pt	en	10				10

With the exception of study JLG10 (20 translation sessions), all of the studies contain keystroke and gaze data. Gaze data was collected with Tobii eyetracker 1750 (BD08, ACS08, KTHJ09 and LWB09), Tobii T120 (TPR11, BML12, MS12, NJ12) and Tobii TX300 for SG12. The 10 studies were conducted for different reasons and with different research goals. While the collected data has been evaluated in numerous publications, the primary purpose of the studies were as follows:

ACS08: 30 translations (en->da) and 30 text copying sessions (en->en). The aim of this study was to explore the way in which translators process the meaning of non-literal expressions (Sjørup, 2011)

BD08: 10 translations (en->da), collected in the context of the Eye-to-IT project, to investigate production pauses (Dragsted, 2010)⁸.

KTHJ08: 72 translations (en->da) to investigate translators' allocation of cognitive resources (Jensen, 2011).

LWB09: 40 translations (da->en) to investigate the impact of syntactic processing in translation from L1 to L2 (Sjørup et al. 2009)

JLG10: 10 translations en->pt and 10 translations pt->en to investigate the impact of direct (L2-L1) and indirect (L1-L2) translations. (Gonçalves and Alves, 2012)

TPR11: 10 post-editing sessions en->pt and 9 post-editing sessions en->de collected in the context of the TPR summer school 2011.

The following four studies were conducted in the context of the CASMACAT⁹ project, with the aim to compare translation, post-editing and editing activities. A set of 6 English texts was translated and post-edited into Spanish, Chinese, Hindi and German.

BML12: 20 translation, 20 post-editing and 20 editing sessions, all en->es (Mesa-Lao, 2012)

MS12: 15 translation, 19 post-editing and 10 editing sessions, all en->zh (Schmalz, 2012)

NJ12: 39 translation and 61 post-editing sessions, all en->hi (Jaiswal et al. 2012)

SG12: 12 translation, 10 post-editing and 10 editing sessions, all en->de (Hansen and Gutermuth, forthcoming)

4 Database Compilation

The collected TPR UAD is processed and annotated to allow for more detailed analysis and evaluation of the translation processes. For each of the logging files a compilation process produces the following four types of resources (in several different files) which, in addition to the metadata, constitute the CRITT TPR-DB 1.0:

1. Logged UAD (output of Translog)
2. Aligned and annotated product data
3. Treex representations of the product data
4. Unit tables for (quantitative) analysis and visualization of translation progression graphs

⁸ <http://cogs.nbu.bg/eye-to-it/>

⁹ <http://www.casmacat.eu/>

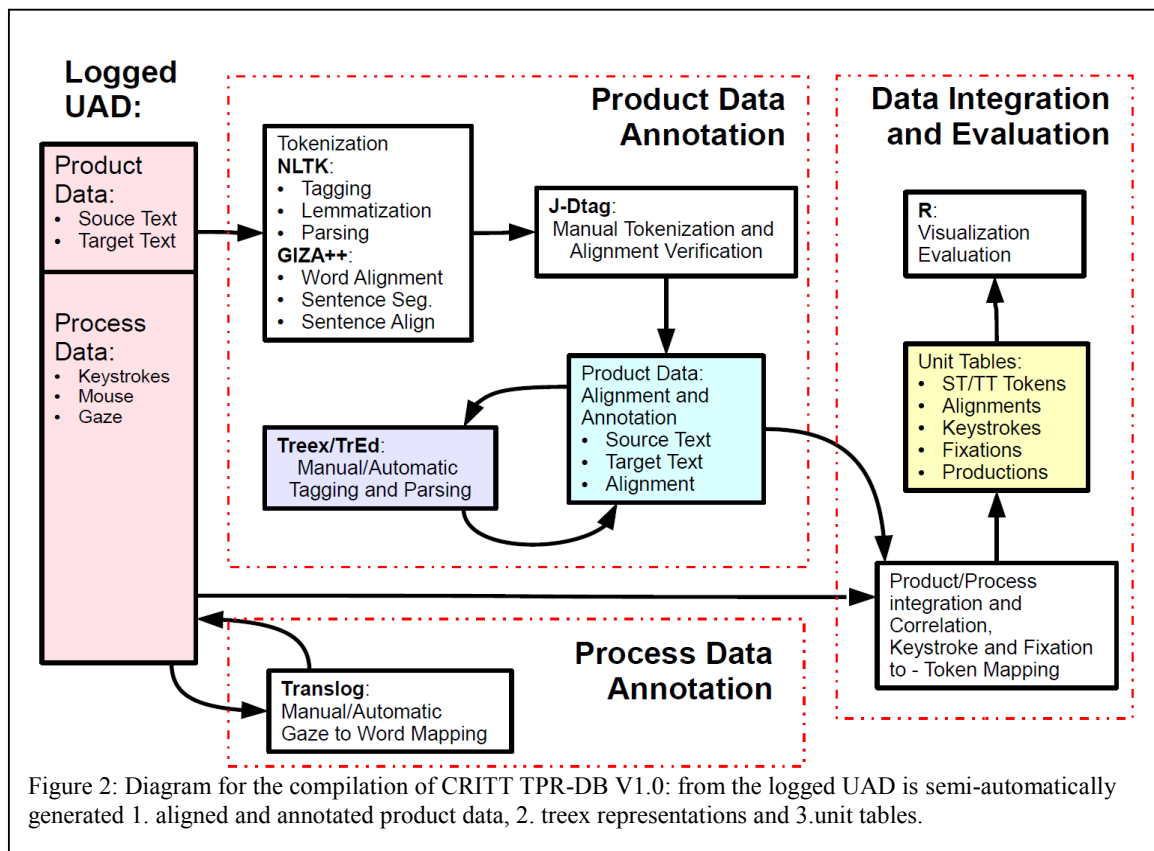


Figure 2: Diagram for the compilation of CRITT TPR-DB V1.0: from the logged UAD is semi-automatically generated 1. aligned and annotated product data, 2. treex representations and 3. unit tables.

Note that the CRITT TPR-DB follows a consistent naming strategy for the folders and files. To anonymise the recordings, filenames consist of a naming strategy which enumerated the participant, the task (translation, post-editing, etc.) and the text. Thus, a recording with the file root *P02_T1* e.g. in BD08 would refer to the recording of participant no. 2 (*P02*) for a translation task of text 1 (*T1*) in that particular study. This file root is kept consistent for all derived and annotated information for this recording. The concatenation of the study name and the file root – e.g. *BD08P01T1* – thus gives a unique identifier for a recording.

Figure 2 plots the processing steps in which the CRITT TPR-DB 1.0 is generated while Figure 3 shows the structure of the database. Besides the studies folders, the database also contains a Treeex, a MetaData, and a bin folder.

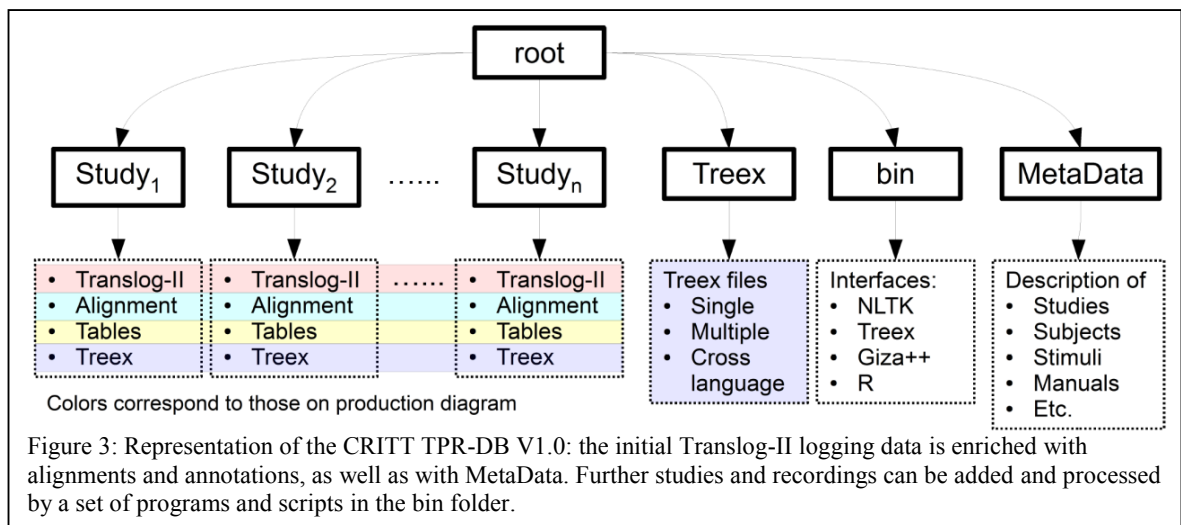
Following the description in Carl and Jakobsen (2009), a distinction is made between product data and process data. Figure 2 shows that both types of data are, to a certain extent, processed

independently and then integrated for the production of unit tables. This information is stored under the Study folder in separate subfolders. The product data (i.e. the final source and target texts) are extracted from the Translog-II logging protocol and linguistically processed in the following steps:

1. Tokenization
2. Sentence segmentation
3. Sentence alignment
4. Word alignment
5. POS tagging and Lemmatization
6. Dependency annotation

Tokenization and sentence segmentation is processed based on our own tools¹⁰, while sentence and word alignment was pre-processed with Giza++ and manually checked and corrected for all of the 456 translation sessions. POS tagging and lemmatization alignment was achieved with the tree tagger for German, English, Danish. We plan

¹⁰ Chinese Tokenization was manually corrected based on a tool provided by Derek Fai Wong, University of Macao.



to manually annotate dependency relations for all source files, as well as for all the sessions in the target files of BD08 study, using the DTAG annotation schema¹¹. The TPR-DB product data is also represented in the Treex format to be visualized in TrEd and to manually correct the linguistic annotation. The Treex folder contains two types of treex representations:

The annotated product data is integrated with the process data by mapping keystrokes and fixations - which occur during the text production - on the source and target language tokens that are being typed or gazed at. The underlying algorithms are described in (Carl and Jakobsen, 2009) and an updated version is available in (Carl, 2012). The integration of the product and process data allows

Table 3: example of alignment units (AU) table showing source and target unit with, the typed string, length of the typed sequence (insertions, deletions), as well as starting time and pre-unit production pause.

AUtarget	AUsource	Len	Ins	Del	Time1	Pause1	Typed
Selvom	Although	7	7	0	1267	12395	Selvom_
udviklingslande_forståeligt	developing_countries	34	31	3	7414	3029	udvikling[]slande_forståelig...
er_nok	are_understandably	7	7	0	688	142	nok_er_
tilbageholdende_med	reluctant	32	26	6	17525	841	tilbageholdende[_edned]dend...
at	to	65	34	31	61505	89	at_gå_på_kompromis_med[de...
ødelægge	compromise	9	9	0	2156	5767	ødelægge_
deres	their	6	6	0	847	120	deres_
chancer_at	chances	11	11	0	1026	237	chancer_at_
for_opnå	of	9	9	0	343	128	for_opnå_

- For each recording a separate treex file is generated, containing only the source text and one translation
- For every source text one treex file is generated, containing all translations for this text.

There are thus 456 treex files of the former and 19 treex files of the latter type.

¹¹ <http://code.google.com/p/copenhagen-dependency-treebank/>

us to generate various unit tables which can then be analyzed and visualized, for instance with R. Currently, the following seven unit tables are produced, each line describes:

Source tokens: enumeration of ST token

Target tokens: enumeration of TT token together with ST correspondence, number, time and value of production keystrokes (number of insertions and deletions).

Keystrokes: text modification (insertions or deletions), together with time of stroke, and the word in the final text to which the keystroke contributes.

Fixations: starting time, end time and duration of fixation, as well as character offset and word id of fixated symbol in the source or target window.

Production units: starting time, end time and duration of coherent sequence of typing (cf. Carl and Kay, 2011), percentage of parallel reading activity during unit production, duration of production pause before typing onset, as well as number of insertion, deletions.

Fixation units: starting time, end time and duration of coherent sequence of reading activities as defined in (Carl and Kay, 2011), as well as ids of fixated words.

Alignment units: source and target correspondences of AU, number of production keystrokes (insertions and deletions) duration of production and revision time, amount of parallel reading activity during AU production.

Each of the units is characterized by a number of features with a consistent naming strategy, so as to easily map contents of different tables. Table 3 in an example of alignment units table: each line describes an AU with a number of features. The data can be statistically evaluated (e.g. with R, for which various scripts exist) for quantitative analysis of translation processes. Given the richness of the CRITT TPR-DB and the structured representation of the data, a large number of additional features may be generated with little effort. Future evaluation of the data will generate needs for additional features which can be easily integrated in the existing framework.

5 Manual Correction

Manual correction and verification of the automated annotation processes are important at all levels of representation. The CRITT TPR-DB compilation process anticipates several steps to manually interfere and checking mechanism are put in place to ensure that the data remains consistent. Currently there are three programs

Jdtag: is a java implementation of a simplified version for bilingual alignment which is compatible with the dtag tool (Kromann, 2003). It allows to visualize word alignments and to modify alignment information in a command line¹², as shown in figure 4.

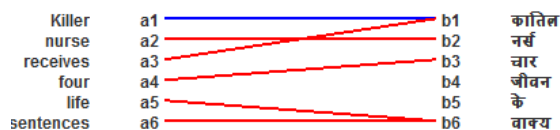


Figure 4: example of alignment visualization in Jdtag

Treex and TrEd: are free software distributed under GPL. TrEd is a fully customizable and programmable graphical editor and viewer for tree-like structures which runs on windows and Unix platforms. The conversion makes use of the Treex¹³ programming interface. Figure 5 shows an example of the GUI.

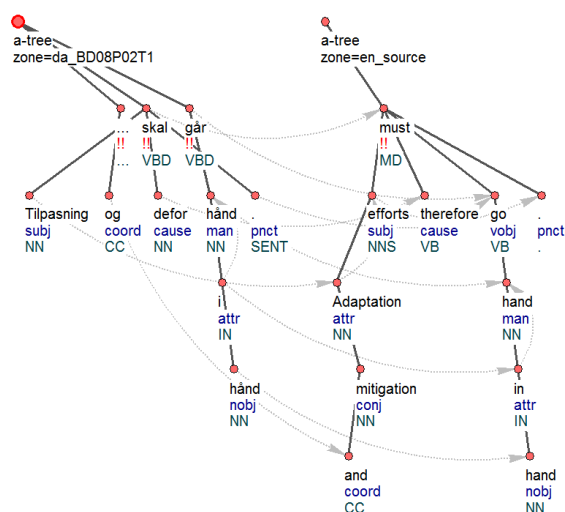


Figure 5: Example of dependency tree alignment and annotation in TrEd

Translog-II: While there are a number of tools and approaches to manually inspect, annotate and amend the product data (such as dtag, Jdtag and TrEd) there are only very few tools for annotating process data, such as the LITTERAE search tool (Alves & Vale 2011). Manual correction of process data includes amendment of logging errors, and the adjustment of gaze-to-word mapping. Due to free head movement and other sources of noise, calibration of gaze data gets often imprecise, so that the captured fixations often cannot be simply mapped to the closest underlying symbols. Despite a font size of 17pt, which was usually chosen in the translation studies, we frequently observe fixation drift to the next line. As shown in Figure 6, we implemented an additional replay mode (FixMap) in the Translog-II program which allows to manually re-

¹² Jdtag was implemented by Ragnar Bonk. It is free software that can be downloaded upon request.

¹³ <http://search.cpan.org/~tkr/Treex-Doc-0.08324/lib/Treex/Tutorial/Install.pod>

assign fixation mappings during the replay of translation sessions, and to store the amended file under a different name.

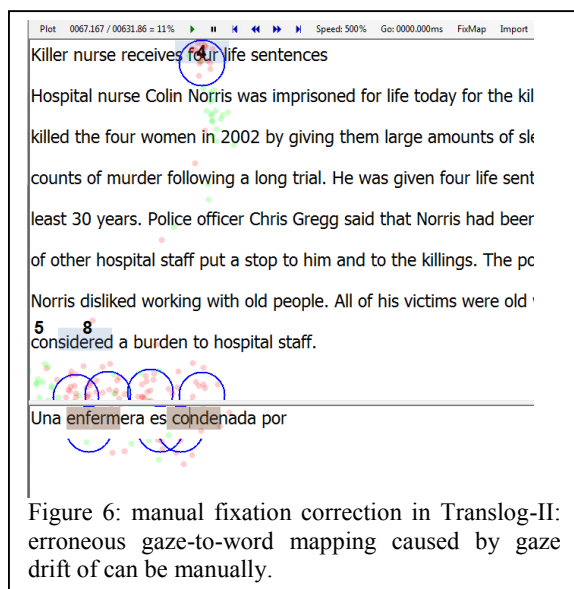


Figure 6: manual fixation correction in Translog-II: erroneous gaze-to-word mapping caused by gaze drift of can be manually.

6 Meta Data

The MetaData folder (see Figure 1) contains very detailed meta data information, as proposed in (Jensen and Carl, 2012). It consists of four csv files:

1. **Study MetaData:** enumerates the studies in the database, describes the purpose of the study, including a bibliography. It contains five categories of information:
 - ExperimentID is a unique identifier which is represented as a derived element in Stimulus metadata and Recordings metadata.
 - Abstract contains an abstract of the main study for which the process data have been collected.
 - Keywords lists the keywords of the experiment.
 - MainLiterature contains a reference to the main study for which data have been collected.
 - SecondaryLiterature contains references to other studies than the main study that have analysed data from the experiment.
2. **Stimulus MetaData:** describes the static properties of the source texts used in the study, their length, domain, etc. It contains the following categories of information:
 - StimulusID is a unique identifier which is represented as a derived element in Recordings metadata.
3. **Recordings MetaData:** provides background for the recordings, such as which texts were used, which hard and software configuration, source and target languages, and date of the recording etc.
 - EyeTrackerType specifies the eye tracking equipment that was used to collect the eye-tracking data.
 - RecordingSoftware specifies the eye tracking recording software that was used to collect the eye-tracking data.
 - EyeTrackerSoftwareVersion specifies the software version of the eye-tracking recording software.
 - Keylogger specifies the keylogging software that was used to collect the typing data.
 - KeyloggingSoftwareVersion specifies the software version of the keylogging software.
 - ExperimentalLocation specifies where the recording was carried out.
 - TargetLanguage specifies the language into which the source text was translated, copied, post-edited, etc
4. **Participants MetaData:** contains information about the participants from whom process data have been collected. It contains the following information:
 - ExperimentID is a derived identifier from Study metadata which links the participant explicitly to an experiment.
 - ExperimentParticipantID is a unique identifier which is represented as a derived element in Recordings metadata.
 - Sex of the participant.
 - YearOfBirth of the participant.
 - Programme that the participant was enrolled into.
 - Student at the time of recording (yes/no).
 - DegreeStartedYear specifies the year in which the participant was enrolled into a university programme.
 - DegreeFinishedYear specifies the last year of the participant's university programme enrolment.
 - YearsTraining specifies the number of years the participant received translation specific instruction.

- SourceLanguage states the language of the source text.
- LengthWords states the number of words of the source text.
- LengthCharacters states the number of characters of the source text.
- Text contains the source text in its entirety.

- CertifiedTranslator specifies whether or not the participant has received formal authorisation to work as a translator and/or interpreter.
- ExperienceYears specifies the number of years the participant has worked as a professional translator.
- L1 of the participant.
- L2 of the participant.
- L3 of the participant.
- OpticalAids specifies whether or not the participant uses optical aids such as glasses or contact lenses.
- LeftEye specifies the dioptre for the left eye.
- RightEye specifies the dioptre for the right eye.
- EyeColour of the participant.

Note that not all information is provided for all studies/participants/recordings. In fact it is difficult to gather all the data for experiments which have been conducted 5 years ago. While the naming convention in the Metadata is consistent with the study and recording name in as described in section 4, there is, as of now, no appropriate query tool available.

7 Conclusion

The paper describes the first public release of the CRITT TPR-DB. More than 450 translation sessions were recorded (more than 400 with gaze data) linguistically annotated and stored in a consistent data format. The database contains translations mainly from English into very different languages, such as Spanish, Hindi, Chinese and German, produced by novice and experienced translators. It contains from scratch translations, mono- and bilingual post-edited MT output (google and AnglaBharati (Sinha, 2005)) as well as text copying, with very detailed key logging and gaze data information. Some of the data also has detailed metadata information about the Stimulus, Recording and Participant. It is thus possible to compare translation behavior of the same participant across different studies and tasks (translation, post-editing, etc.) as well as compare translation strategies of different translators when translating the same text into different languages.

In future releases of the database we will add more experiments, complete the annotation (e.g. by adding more dependency annotations), but also add more tools to query the database and extract more features for the unit tables. Particular focus will also be given to the gaze data and gaze-to-word mapping strategies, as this seems to be the most noisy and least understood part in the database. Given the increased interest in post-editing, we

hope that the CRITT TPR-DB will attract researchers to analyze and compare translation and post-editing processes to better understand and model these different activities, and to finally develop tools that better support translators in their work.

Acknowledgments

This collection would not have been possible without the year-long work of the experimenters as well as the many study participants, translators, translation students and bilingual editors and annotators. In particular thanks to Zdenek Zabokrtsky and Jan Stephanek for helping with TrEd and Treex, to Abhijit Mishra who implemented the NLTK interface during a summer internship and Kristian TH Jensen for his ideas on the MeteData.

References

- Alves, F. & Vale, D. C. 2011. *On drafting and revision in translation: a corpus linguistics oriented analysis of translation process data*, Translation: Computation, Corpora, Cognition, Vol 1, No 1 <http://www.t-c3.org/index.php/t-c3/article/view/3>
- Steven Bird, Ewan Klein, and Edward Loper . Natural Language Processing with Python --- Analyzing Text with the Natural Language Toolkit
- Carl, Michael (2012). Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research, In Proceedings of the Eight International Conference on Language Resources and Evaluation, European Language Resources Association (ELRA)
- Carl, M. and Jakobsen A.L. (2009). Towards statistical modelling of translators' activity data. *International Journal of Speech Technology*, 12(4). <http://www.springerlink.com/content/3745875x22883306/>.
- Carl, M., and Kay, M. (2011). Gazing and Typing Activities during Translation: A comparative Study of Translation Units of Professional and Student Translators, *META* 56:4
- Dragsted, B (2010). Co-ordination of reading and writing processes in translation. Contribution to Translation and Cognition, Shreve, G. & Angelone, E. (eds.). John Benjamins in 2010.
- Gonçalves, José Luiz V. R. and Alves, Fabio, Investigating the conceptual-procedural distinction in the translation process: a relevance-theoretic analysis of micro and macro translation units, in press

- Göpferich, S., Jakobsen, A. L., and Mees, I. M., editors (2009). *Behind the Mind: Methods, Models and Results in Translation Process Research*. Copenhagen: Samfundslitteratur, Copenhagen
- Göpferich, Susanne (2008): *Translationsprozessforschung: Stand - Methoden - Perspektiven*. *Translationswissenschaft* 4. Tübingen: Narr
- Göpferich, Susanne/Alves, Fabio/Mees, Inger M., Hrsg. (2010): *New Approaches in Translation Process Research*. *Copenhagen Studies in Language* 39. Copenhagen: Samfundslitteratur.
- Hansen and Gutermuth (2013), forthcoming
- Holmes, James S. (1972/1988). *The Name and Nature of Translation Studies*. In Holmes, *Translated! Papers on Literary Translation and Translation Studies*, Amsterdam: Rodopi, pp. 67–80.
- Jaiswal, Sinha and Shukla (2012), *Analysing User Activity Data of English-Hindi Translation*, forthcoming
- Jakobsen, A.L. und Schou, L. (1999). "Translog documentation." *Copenhagen Studies in Language* 24: 149–184.
- Jakobsen, A.L. (2006). "Research methods in translation – Translog." In Sullivan, K. P. H. and Lindgren, E., (eds) *Computer keystroke logging and writing: Methods and applications*, volume 18. Oxford: Elsevier, 95–105
- Jensen, HKT and Carl, Michael. *User Activity Metadata for Reading, Writing and Translation Research*. In: *Proceedings of The Eighth International Conference on Language Resources and Evaluation. LREC 2012: Workshop: Describing LRs with Metadata: Towards Flexibility and Interoperability in the Documentation of LR*. ed. / Victoria Arranz; Daan Broeder; Bertrand Gaiffe; Maria Gavrilidou ; Monica Monachini ; Thorsten Trippel. Paris : ELRA, 2012. p. 55-59.
- Jensen, Hvelplund, K. T. (2011), *Allocation of cognitive resources in translation: an eye-tracking and key-logging study*. Ph.d.-afhandling. Copenhagen Business School. Copenhagen.
- Krings, H.P. (1986). *Was in den Köpfen von Übersetzern vorgeht*. Tübingen: Narr.
- Kromann, M. T. (2003). *The Danish Dependency Treebank and the DTAG treebank tool*. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT 2003)*, 14-15, November, Växjö. 217–220.
- Van Maes, L. and Leijten, M. (2006). "Logging writing processes with Inputlog." In Van Maes, L., Leijten, M. and Neuwirth, D. (eds). *Writing and Digital Media*. Oxford: Elsevier, 158-165.
- Lörscher, W. (1992). "Investigating the Translation Process." *Meta*, XXXVII (3):426–439.
- Mesa-Lao, Barto (2013), *Analysing User Activity Data of English-Spanish Translation*, *International Workshop on Expertise in Translation and Post-editing Research and Application*, Copenhagen
- Mees, I. M., Alves, F. and Göpferich, S., (eds) (2009). *Methodology, Technology and Innovation in Translation Process Research*, volume 38. Copenhagen: Samfundslitteratur.
- Sinha R.M.K., (2005), *Integrating CAT and MT in AnglaBharti-II Architecture*, *EAMT 2005*, Budapest, Hungary.
- Sjørup, Annette C. *Cognitive Effort in Metaphor Translation : An Eye-Tracking Study.. I: Cognitive Explorations of Translation*. red. / Sharon O'Brien. London : Continuum International Publishing Group Ltd, 2011. s. 197-214 (*Continuum Studies in Translation*).
- Sjørup, A.C., Jensen, K.T.H., & Balling, L.W. (2009). *Syntactic processing in translation from L1 to L2*. *Eye-to-IT conference on translation processes 28-29 April 2009*, Copenhagen, Denmark.
- Schmalz (2012). *English-Chinese Manual and Post-Ed Translation Process: a Comparative Pilot Study*, *International Workshop on Expertise in Translation and Post-editing Research and Application*, Copenhagen
- Tirkkonen-Condit, Sonja (2005) *The Monitor Model Revisited: Evidence from Process Research*, *META*, Volume 50, numéro 2, avril 2005, p. 405-414
- Tirkkonen-Condit, S. & Jääskeläinen, R. (eds) (2000). *Tapping and Mapping the Process of Translation and Interpreting*. Amsterdam & Philadelphia: John Benjamins.