Initial Report on User Interface Studies, Cognitive and User Modelling

Carl, Michael; Hill, Robin

D1.1: Initial report on user interface studies, cognitive and user modelling

Michael Carl and Robin Hood

Distribution: Public

CASMACAT
Cognitive Analysis and Statistical Methods
for Advanced Computer Aided Translation

ICT Project 287576 Deliverable D1.1

SEVENTH FRAMEWORK PROGRAMME

Project funded by the European Community
under the Seventh Framework Programme for
Research and Technological Development.

Information Society Technologies

Project ref no.	ICT-287576
Project acronym	CASMACAT
Project full title	Cognitive Analysis and Statistical Methods for Advanced Computer Aided Translation
Instrument	STREP
Thematic Priority	ICT-2011.4.2 Language Technologies
Start date / duration	01 November 2011 / 36 Months

Distribution	Public
Contractual date of delivery	October 31, 2012
Actual date of delivery	October 29, 2012
Date of last update	October 29, 2012
Deliverable number	D1.1
Deliverable title	Initial report on user interface studies, cognitive and user modelling
Type	Report
Status & version	Draft
Number of pages	51
Contributing WP(s)	WP7
WP / Task responsible	CBS, UEDIN
Other contributors	
Internal reviewer	
Author(s)	Michael Carl and Robin Hood
EC project officer	Kimmo Rossi
Keywords	

The partners in CASMACAT are:

University of Edinburgh (UEDIN)
Copenhagen Business School (CBS)
Universitat Politècnica de València (UPVLC)
Celer Soluciones (CS)

For copies of reports, updates on project activities and other CASMACAT related information, contact:

The CASMACAT Project Co-ordinator
Philipp Koehn, University of Edinburgh
10 Crichton Street, Edinburgh, EH8 9AB, United Kingdom
pkoehn@inf.ed.ac.uk
Phone +44 (131) 650-8287 - Fax +44 (131) 650-6626

Copies of reports and other material can also be accessed via the project's homepage:
http://www.casmacat.eu/

Executive Summary

This WP lays the empirical foundations for the development of the CASMACAT workbench. A series of experiments will establish basic facts about translator behaviour in computer-aided translation, focusing on the use of visualisation option and input modalities. Another series of studies will deal with individual differences in translation, in particular translator types and translation styles.

The initial report deals with translation types and styles, text types and reading model adapted for machine translated texts. It covers the first periode of Tasks 1.3, 1.4, and 1.5. The deliverable is structured into three sections which biefly summarize the work and an appendix which contains more detailed information about the produced material and a number of papers.

An experimental setup (see section 2.1) and a questionnaire (see section 1.1) was designed to obtain consistent data from various translators in different languages under similar conditions. Translation data was collected in several locations (section 2.2) and assembled into a TPR database, as described in section 1.2. Preliminary studies were conducted to investigate post-editing and translation styles (section 1.3). Translation data was also collected in the first CASMACAT field trial. The assessment is provided in Deliverable d6.1. Section 3 describes the first Edinburgh Eyetracking experiment while the Appendix contains furter material.

Contents

1 Translator Types and Translation Styles. (Task 1.3) Design and deploy questionnaire to determine translator types; analyze translator types and correlate data with inter- and intra-translator variance in translation styles.

1.1 Design of the questionnaire

Two questionnaires have been designed, to interrogate the translator before the experimental session takes place, and another one after the session is over. These questionnaires are reproduced in Appendix 4.2. The Meta data gathered from these questionnaires was added to the TPR-DB.

1.2 Release of the TPR-DB

Prior to investigating the CASMACAT translation experiments, a substantial amount of time has been spent on the conceptualization and implementation of a consistent database format for translation process data (TPR-DB). Legacy data from Translog-II experiments were converted into the database format and more than 240 translation sessions were recorded and added to the TPR-DB (see Task 1.4, below). A first version of this database was released as TPR-DB V1.0 in the context of the TPR summer school (August 13, 2012, see Deliverable d7.1. `http://www.cbs.dk/content/download/189944/2411764/file/Balling&Carl.pdf`). The TPR-DV V1.0 is described in an AMTA workshop paper (reproduced in Appendix 4.5) and a description to more recent additions of the feature extraction component are submittied as a Coling workshop paper, attached in Appendix 4.6. .

The TPR-DB V1.0 is publicly available and can be downloaded from: `https://dl.dropbox.com/u/7757461/TPR-DB.zip`

Within the CASMACAT field trial, more than 90 translation sessions were recorded with the CASMACAT prototype-I. This data has also been converted into the TPR-DB format and is available from: `https://dl.dropbox.com/u/7757461/Casmacat%20Field%20Trial%201.zip`

A further release of the TPR-DB V1.1 is planned for Winter 2012 in which the additional data will be publicly released, together wit a number of fixes. Additional data will (most likely) include:

- Process data of the first CASMACAT field trial
- Additional English → Spanish translations
- Additional English → German translations
- Additional Chinese → Portuguese translations
- Process data for English → Farsi (no gaze data, only keylogging)
- Authoring data, which allows to compare text production (journalistic Spanish text production) with translation
- Additional features, as described in the paper in Appendix 4.6

1.3 Preliminary qualitative investigation into post-editing and translation styles

An initial qualitative analysis of the data collected in T1.4 has been carried out for several language pairs. A number of pilot studies were conducted to explore the differences between translation and post-editing of texts through the analysis of user activity data. The following papers are published based on the TPR-DB data:

- English → German (Gutermuth, 2012), paper presented at the EyeTrackingBehaviour conference, Leuven, 2012 (see appendix 4.3)
- English → Spanish (Lao-Mesa, 2012), paper presented at the ETP workshop, Copenhagen 2012 (see appendix 4.4)
- English → Hindi (Jaiswal et at, 2012), paper submitted to the ETNLP workshop, Mumbai, 2012

2 Text Type. (Task 1.4) Conduct translation experiments to investigate the correlations between translation styles, different text types and preferred visualisation options in the CASMACAT editor.

2.1 Set up of experimental translation design

To compare from-scratch translation (T), post-editing (P) and monolingual post-editing (E), six texts were chosen to be translated by various student and experienced translators. Since the CASMACAT workbench is still at an experimental stage, Translog-II was chosen as data acquisition software. The six source texts were permuted in a systematic manner so as to make sure that each text was translated by every translator and every translator translated two different text in each translation mode. See Appendix 4.1 for the list of texts and tasks distribution. The same order and naming schema was kept identical for all translation experiments. The three texts consisted of 3 news texts and 3 sociological texts from an ececlopeda.

2.2 Collection of experimental data

More than 240 translation sessions have been conducted within the context of CASMACAT covering more than 80 hours for translation, post-editing and monolingual editing from English into several languages:

- 20 hours English → Spanish, these recordings were conducted at UAB Barcelona by Bartholome Mesa-Lao with translation students and professional translators
- 20 hours English → German, these recordings were conducted at the University of Mainz, mainly by Sikle Guthermuth with translation students and professional translators.
- 10 hours English → Chinese, these recordings were conducted at the University of Macao, by Marcia Schmaltz with translation students and professional translators.
- 30 hours English → Hindi, these recordings were conducted at the CDAC Noida, India, mainly by Nishtha Jaiswal and Michael Carl with translation students and professional translators.

Note that the Chinese and Hindi translations were not a committment of the CASMACAT task and were not payed by CASMACAT money. It is, however, highly interesting additional data which promotes the CASMACAT studies far beyond the European borders and allows us to compare translation processes into very different languages.

The product data gathered from these translation experiments was tokenized word-aligned (most of the word alignments are manually corrected/checked), PoS tagged, and some of it also parsed (see see Appendix 4.5). Further experiments English ↔ Danish, English ↔ Spanish, and English ↔ German are planned. To allow for cross-translator/language comparision, all these experiments use the same set of English source texts. Meta data has been collected according to questionnaire shon innAppendix 4.2 and added to the TPR-DB.

3 Cognitive Modeling. (Task 1.5) Build cognitive models that capture processing difficulty in translation (reading models for source and target text).

3.1 Edinburgh Eye-tracking Experiment 1

Objective: to explore human error-checking behaviour in a simulated post-editing environment.

Phase 1 starts with the easiest case and provides the baseline condition: monolinguals (native English speakers) reading MT output in an error-spotting task. Phase 2 will provide the main contrast by using participants who are bilinguals and professional translators. There is an established convention for investigations into bilingual (dis)advantages to contrast mono- and bilinguals on the same tasks (e.g. Sandoval, Gollan, Ferreira, and Salmon, 2010). Additionally, the experimental materials will include non-language specific errors, such as letter transposition or other typographic errors. The aim is to validate the experiment by replicating existing (monolingual) error-detection findings and task effects (e.g. Kaakinen and Hyona, 2010; Rayner, White, Johnson, and Liversedge, 2006) while also enabling comparisons with language-critical (i.e. translation) errors.

3.2 Technical details

A selection of target sentences were drawn from materials extracted from project-related corpora in order to ensure authentic stimuli, e.g. the Edinburgh submission for the German-to-English WTM12 shared task (part of the EuroMatrixPlus project) `http://matrix.statmt.org/matrix/output/1692?run_id=2517`. Each sentence contained a single error. These experimental items were combined with a set of filler materials extracted from native-English corpora (i.e. fluent, error-free sentences). Participants are therefore presented with a mixture of error-containing and error-free sentences in random order.

Eyetracking is an extremely useful technique for examining language processing, including recently the reading of machine translated text (e.g. Doherty, OBrien, and Carl, 2010). Adopting this paradigm, eye movements were recorded using a SR Research Eyelink 2K running in desktop mount mode and sampling at 1KHz in combination with a Samsung 22 monitor operating a refresh rate of 120Hz and 1680 x 1050 resolution.

3.3 Procedure

Participants are instructed to fully read each sentence that is presented to them and to then decide if it contains an error by clicking the left mouse button for yes or the right mouse button for no. Following a yes decision, the sentence is redisplayed on the screen and the task becomes to click on the first word of any error. A quarter of sentences (irrespective of whether they contain an error) are followed by a comprehension-testing question, ensuring task compliance. Dependent Variables pertaining to full sentence reading, re-reading of correctly identified problem sentences, error detection followed by error location, and comprehension are obtained.

3.4 Impact

The experimental data will then feed into the development of the cognitive modelling of translators and research on bilingualism, providing an insight into how translators read and evaluate translated text. Subsequent experiments will then utilise eye-tracking to investigate post-edit checking when both source and target text is present simultaneously (the basis of the CASMACAT tool). This will provide further support for the classification of translator types and styles, as well as leading to optimised presentation for efficient translator behaviour. Careful manipulation of the quantity of linguistic material available at any moment may help reduce the cognitive costs associated with switching between languages and between comprehension and production processes, for example (e.g. Gollan and Ferreira, 2009). Research on bilingualism has typically focused on two key cognitive mechanisms that introduce differences between bilinguals and monolinguals: the reduced frequency of language-specific use (weaker links); and competition for selection within the language system in bilinguals (interference) (Mindt, et al., 2008). It may be possible to either exploit or minimise these differences where appropriate.

Humans are susceptible to non-statistical linguistic factors in their translation choice/decision, and so predicting preferences will have to take this into consideration. For instance there can be a tendency to prefer cognates (translations similar in meaning and form) rather than noncognates (translations similar in meaning only) (e.g. Ibez, Macizo, and Bajo, 2010).

3.5 References

- Doherty, Stephen, OBrien, Sharon, and Carl, Michael (2010). Eye tracking as an MT evaluation technique. Machine Translation, 24(1), 1-13. doi: 10.1007/s10590-010-9070-9.
- Gollan, T. H., and Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. Journal of Experimental Psychology: Learning Memory and Cognition, 35(3), 640-665. doi: 10.1037/a0014981.
- Ibez, A. J., Macizo, P., and Bajo, M. T. (2010). Language access and language selection in professional translators. Acta Psychologica, 135(2), 257-266. doi: 10.1016/j.actpsy.2010.07.009.
- Kaakinen, J. K., and Hyona, J. (2010). Task effects on eye movements during reading. [Article]. Journal of Experimental Psychology: Learning Memory and Cognition, 36(6), 1561-1566. doi: 10.1037/a0020693.
- Mindt, M. R., Arentoft, A., Germano, K. K., D'Aquila, E., Scheiner, D., Pizzirusso, M., et al. (2008). Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. Neuropsychology Review, 18(3), 255-268. doi: 10.1007/s11065-008-9069-7.

- Rayner, K., White, S. J., Johnson, R. L., and Liversedge, S. P. (2006). Raeding wrods with jubmled lettres - There is a cost. Psychological Science, 17(3), 192-193. doi: 10.1111/j.1467-9280.2006.01684.x
- Sandoval, T. C., Gollan, T. H., Ferreira, V. S., and Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. Bilingualism: Language and Cognition, 13(2), 231-252. doi: 10.1017/s1366728909990514.

4 Appendix

4.1 Experimental Design

The spreadsheet show the distribution of texts and tasks to successive translators (Participant01 ... Participant24). Each experiment consisted of six translation translation session in which one text was translated/edited. The first two texts were translated (T) followed by two texts to be post-editied (P) and two texts to be edited (E). This order was kept constant, but the actual texts were permuted according to the schema. Texts have between 100 and 200 words and fit on one screen. A translation session lasted approx 20. mins (sometimes for some translators also more than one hour).

	Date	Time	From Scratch		Post-editing		Editing	
Participant 01			T1	T2	P3	P4	E5	E6
Participant 02			T3	T4	P5	P6	E1	E2
Participant 03			T5	T6	P1	P2	E3	E4
Participant 04			T2	T1	P4	P3	E6	E5
Participant 05			T4	T3	P6	P5	E2	E1
Participant 06			T6	T5	P2	P1	E4	E3
Participant 07			T1	T3	P2	P4	E5	E6
Participant 08			T3	T5	P4	P6	E1	E2
Participant 09			T5	T1	P6	P2	E3	E4
Participant 10			T2	T4	P1	P3	E6	E5
Participant 11			T4	T6	P3	P5	E2	E1
Participant 12			T6	T2	P5	P1	E4	E3
Participant 13			T1	T3	P2	P5	E4	E6
Participant 14			T3	T5	P4	P1	E6	E2
Participant 15			T5	T1	P6	P3	E2	E4
Participant 16			T2	T4	P1	P6	E3	E5
Participant 17			T4	T6	P3	P2	E5	E1
Participant 18			T6	T2	P5	P4	E1	E3
Participant 19			T6	T3	P2	P5	E4	E1
Participant 20			T2	T5	P4	P1	E6	E3
Participant 21			T4	T1	P6	P3	E2	E5
Participant 22			T5	T4	P1	P6	E3	E2
Participant 23			T1	T6	P3	P2	E5	E4
Participant 24			T3	T2	P5	P4	E1	E6

4.2 Design of Questionnaires

TRANSLATING / POST-EDITING / EDITING OF MACHINE TRANSLATION

QUESTIONNAIRE 1

Name:…………………………………………………………………………………………………

Sex: ○ M ○ F

Wear Glasses: ○ Yes ○ No

Years of formal translator training: …................ Years

Years of translator experience: …................ Years

Languages L1 …......................... L2 …......................... L3 ….........................

How frequently do you use machine translation?

- ○ Every day
- ○ Every 2 - 3 weeks
- ○ Every month
- ○ Once or twice a year
- ○ Never

From your previous experience with machine translation outputs, how would you rate your level of satisfaction in relation to machine translation?

- ○ Highly satisfied
- ○ Somewhat satisfied
- ○ Neutral
- ○ Somewhat dissatisfied
- ○ Highly dissatisfied

Do you think that you will want to apply machine translation in your future translation tasks?

○ Yes ○ No ○ I'm not sure

<u>In general</u>, how feasible do you think it is to apply machine translation to professional translation services?

- ○ Very likely
- ○ Somewhat likely
- ○ Neutral
- ○ Somewhat unlikely
- ○ Very unlikely

Have you ever post-edited[1] machine translation?

○ Yes ○ No

[1] In this context, *post-editing* refers to "the process of improving a machine-generated translation with a minimum of manual labour by a human translator". A person who post-edits is called a *post-editor*.

TRANSLATING / POST-EDITING / EDITING OF MACHINE TRANSLATION

QUESTIONNAIRE 2

Name:……………………………………………………………………………………………………

How satisfied are you with the translation you have produced:

through post-editing	**through editing?**
○ Highly satisfied	○ Highly satisfied
○ Somewhat satisfied	○ Somewhat satisfied
○ Neutral	○ Neutral
○ Somewhat dissatisfied	○ Somewhat dissatisfied
○ Highly dissatisfied	○ Highly dissatisfied

Would you have preferred to work on your translation from scratch instead of post-editing machine translation?

○ Yes ○ No

Do you think that you will want to apply machine translation in your future translation tasks?

○ Yes, at some point ○ No, never! ○ I'm not sure yet

Based on the post-editing task you have performed, how much do you rate machine translation outputs on the following attributes?

	Well Below Average	Below Average	Average	Above Average	Well Above Average
Grammaticality	○	○	○	○	○
Style	○	○	○	○	○
Overall accuracy	○	○	○	○	○
Overall quality	○	○	○	○	○

Based on the post-editing task you have performed, which of these statements will you go for?

○ I had to post-edit ALL the outputs.
○ I had to post-edit about 75% of the outputs.
○ I had to post-edit 25 -50% outputs.
○ I only had to post-edit VERY FEW outputs.

Based on the post-editing task you have performed, how often would you have preferred to translate from scratch rather than post-editing machine translation?

○ Always.
○ In most of the cases (75% of the outputs or more).
○ In almost half to the cases (approx. 50%).
○ Only in very few cases (less than 25%).

4.3 Evaluation of English $\rightarrow$ German

The next pages contain a presentation given at the Tobii conference EyeTrackingBehavior, Leuven, 2012

Post-editing machine translation
–
a usability test for professional translation settings

Silke Gutermuth & Silvia Hansen-Schirra
University of Mainz
Germany

Post-editing?

- “term used for the correction of machine translation output by <u>human</u> linguists/editors” (Veale & Way 1997)
- “taking raw machine translated output and then editing it to produce a 'translation' which is suitable for the <u>needs of the client</u>” (one student explaining it to another)
- “is the process of improving a machine-generated translation with a <u>minimum</u> of manual labour”

  (TAUS Report 2010)

Degrees of Post-editing

- light or fast postediting
  - essential corrections only
  - time factor: quick

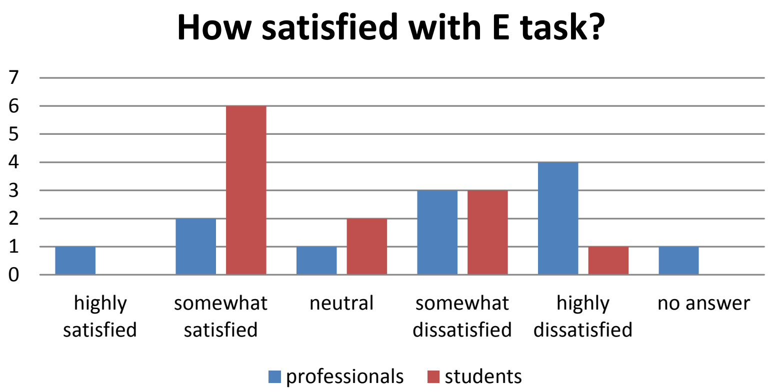
- full post-editing
  - more corrections -> higher quality
  - time factor: slow

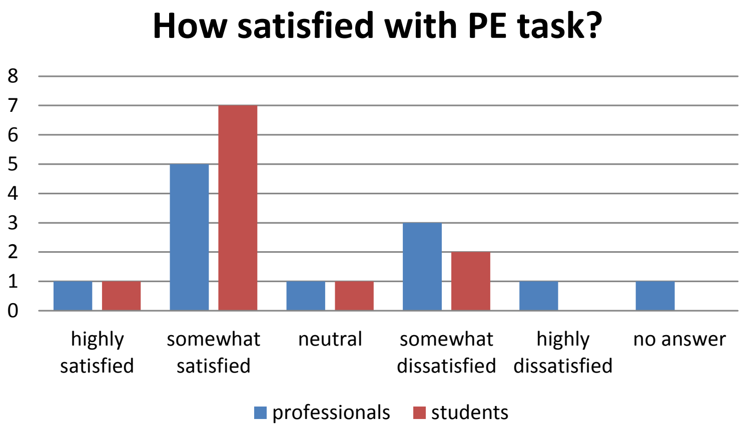
(O‘Brien 2009)

Background

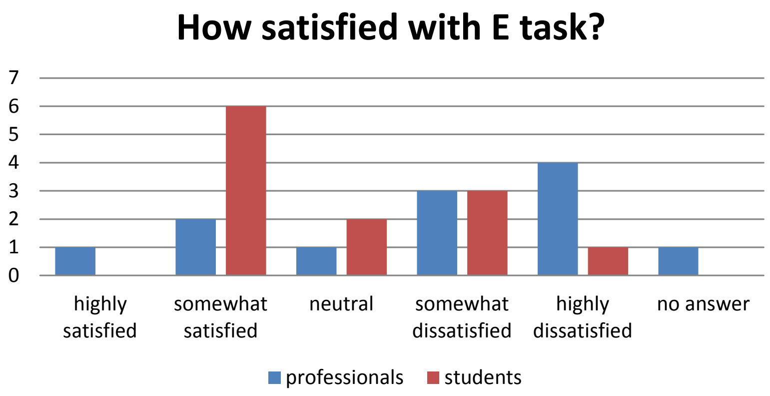
- Motivation: evaluation of machine translation (MT), post-editing of MT, eye-enhanced CAT workbenches (e.g. O'Brien 2011, Doherty et al. 2010, Carl & Jakobsen 2010, Hyrskykari 2006)
- Project: in cooperation with the project CASMACAT, Copenhagen Business School (http://www.cbs.dk/Forskning/Institutter-centre/Institutter/CRITT/Menu/Forskningsprojekter)
- Experiment:
  - English-German
  - translation vs. post-editing vs. editing
  - 6 source texts (ST) with different complexity levels (Hvelplund 2011)
  - 12 professional translators, 12 semi-professional translators
  - eye-tracking (Tobii TX 300), key-logging (Translog), retrospective interviews, questionnaires

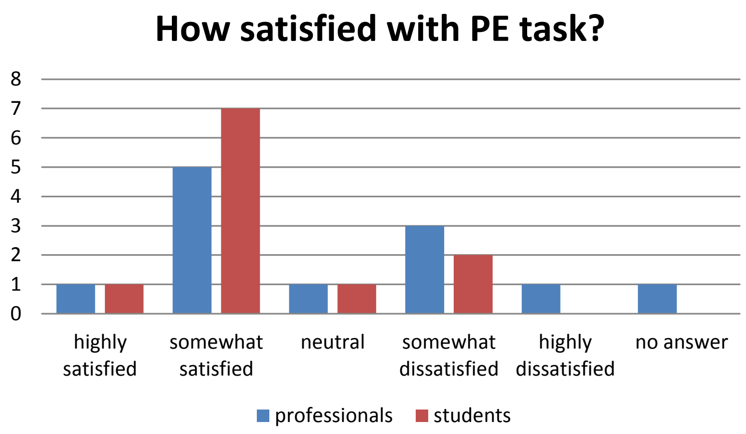
Translators‘ self-estimation

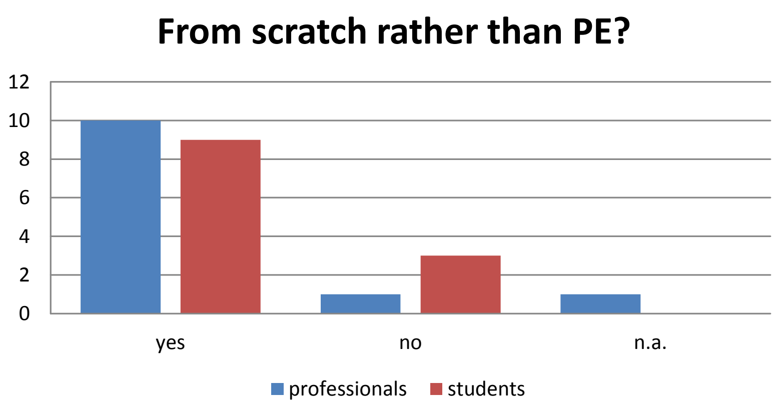




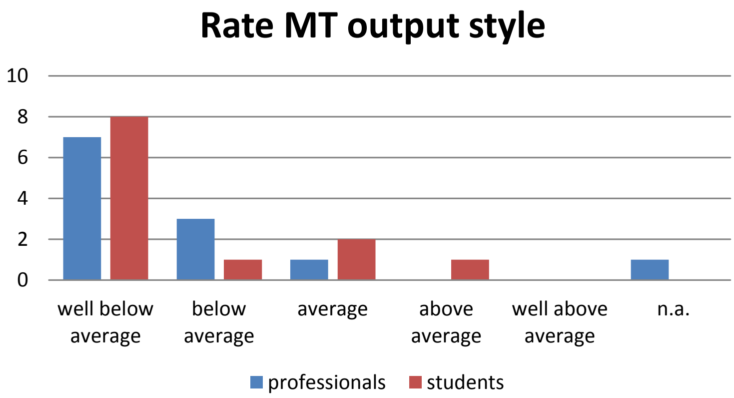
Translators‘ self-estimation

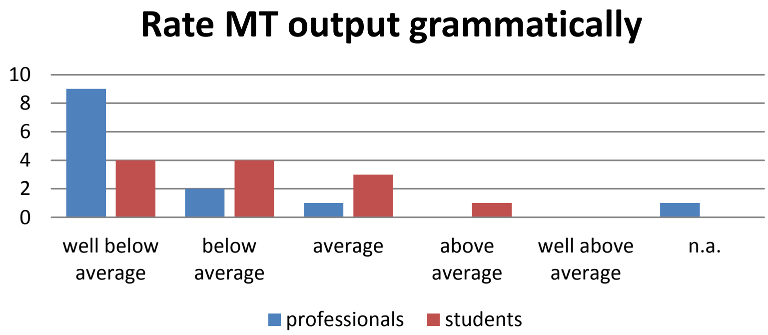




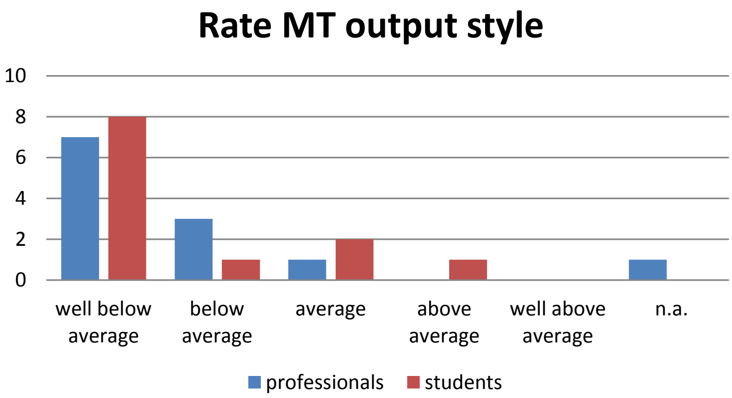


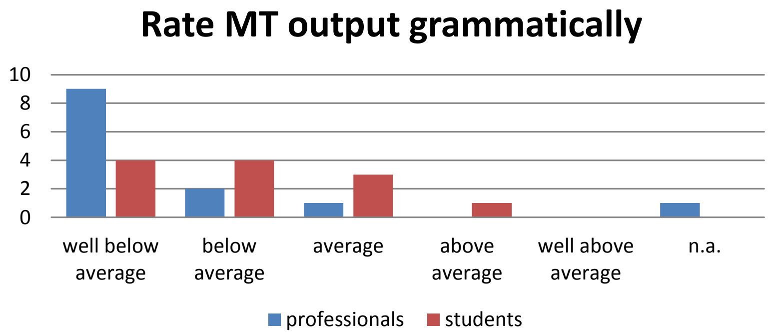
Translators‘ evaluation of MT quality

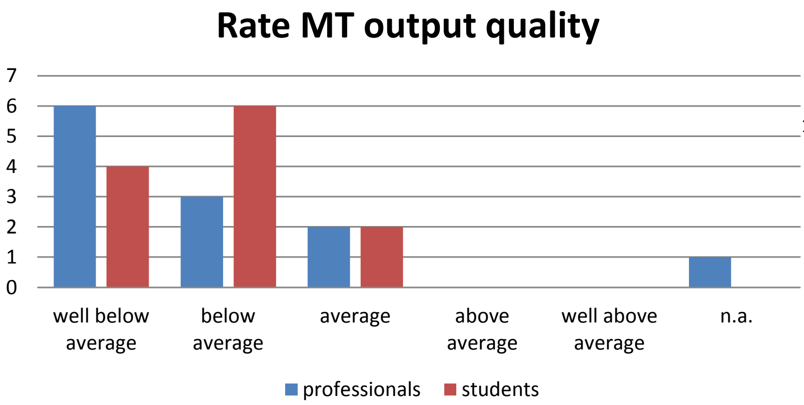


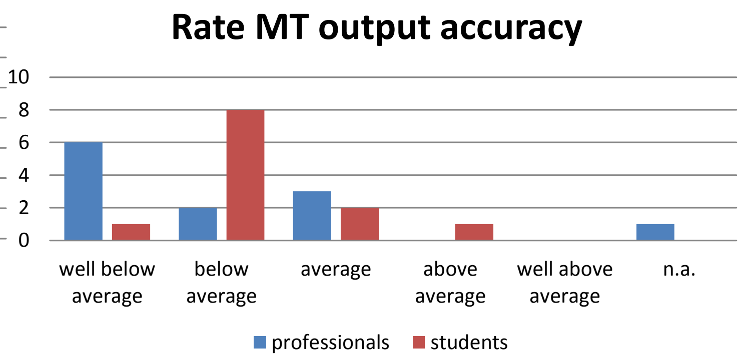


Translators‘ evaluation of MT quality

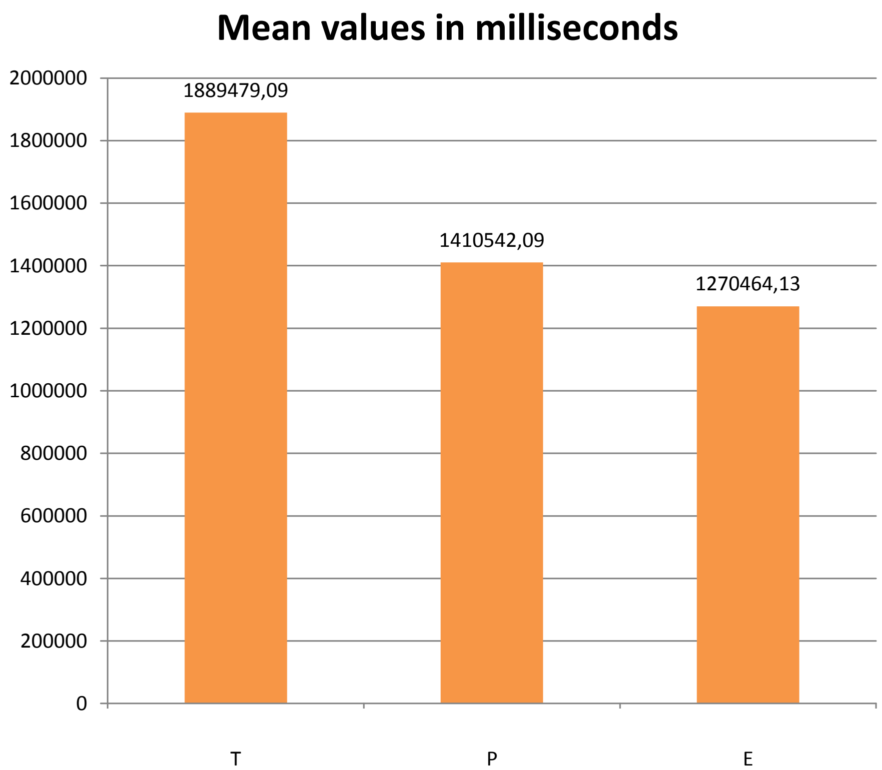




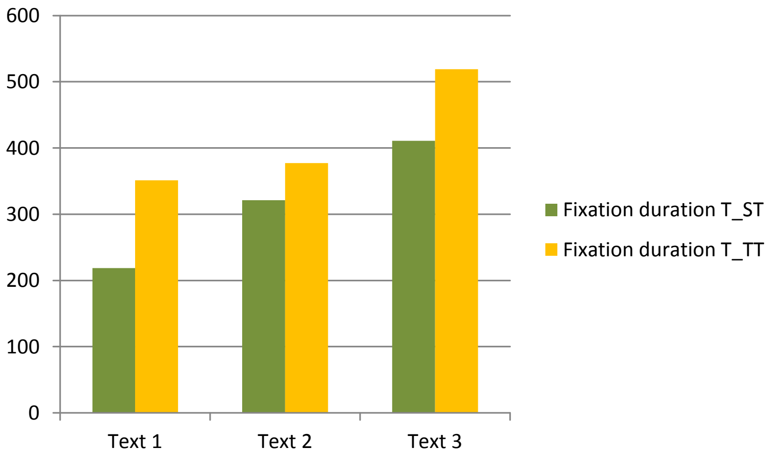


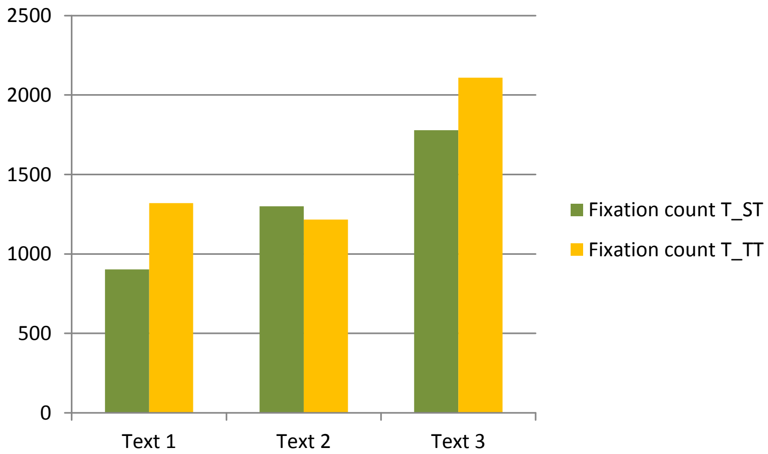


Processing time

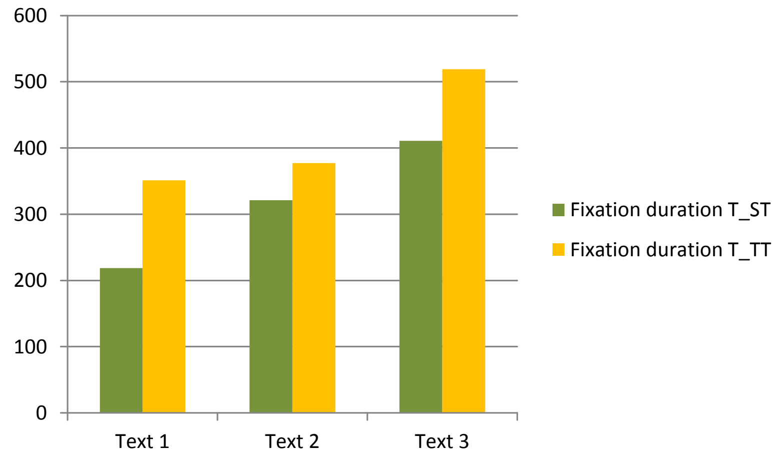


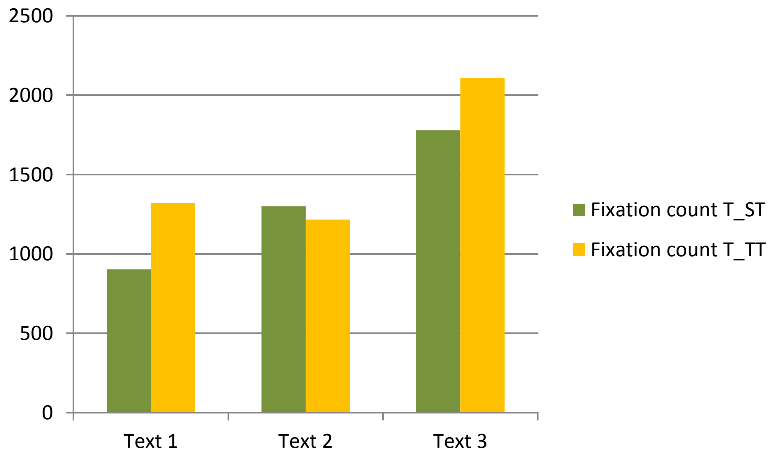
Processing of ST vs. TT Translation

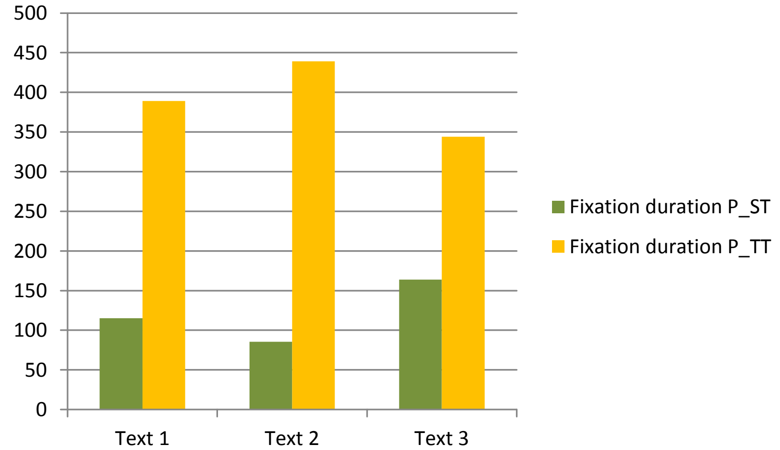


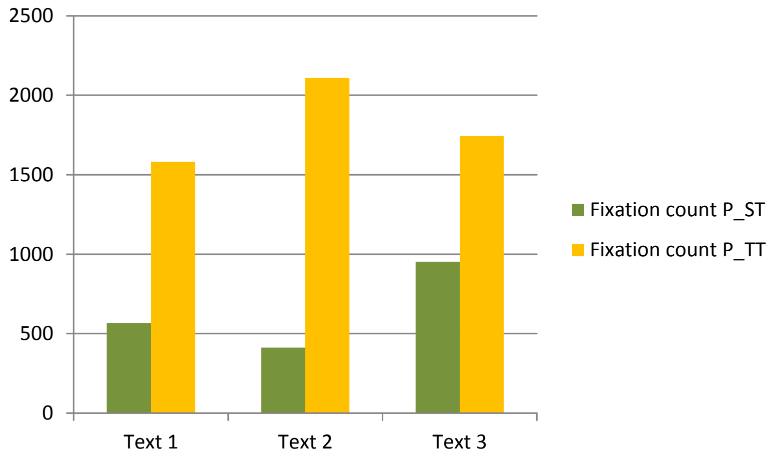


Processing of ST vs. TT
Translation vs. Post-editing

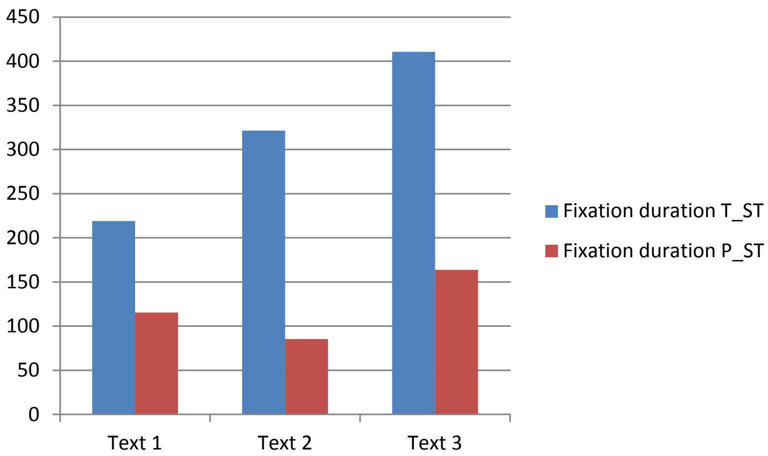


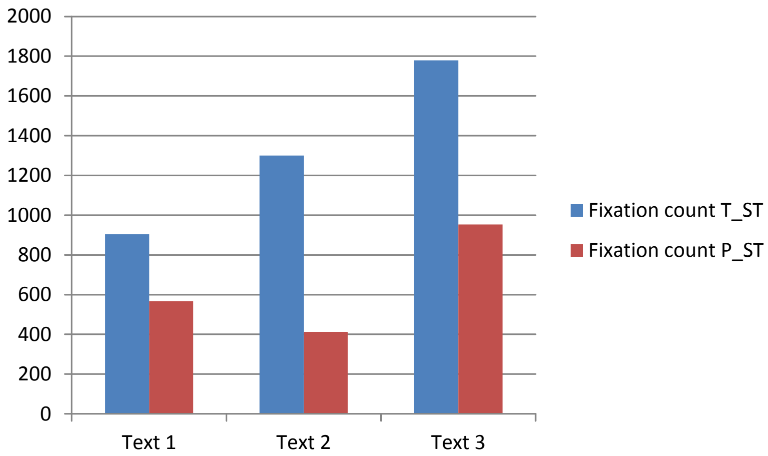




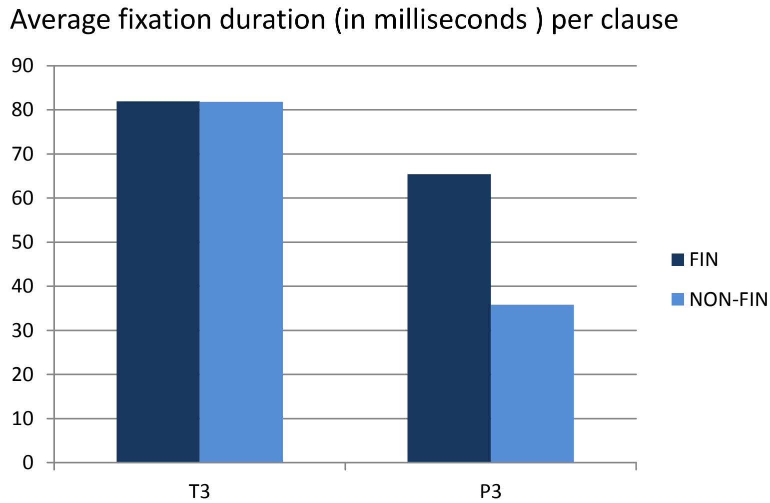


Processing of ST
Translation vs. Post-editing

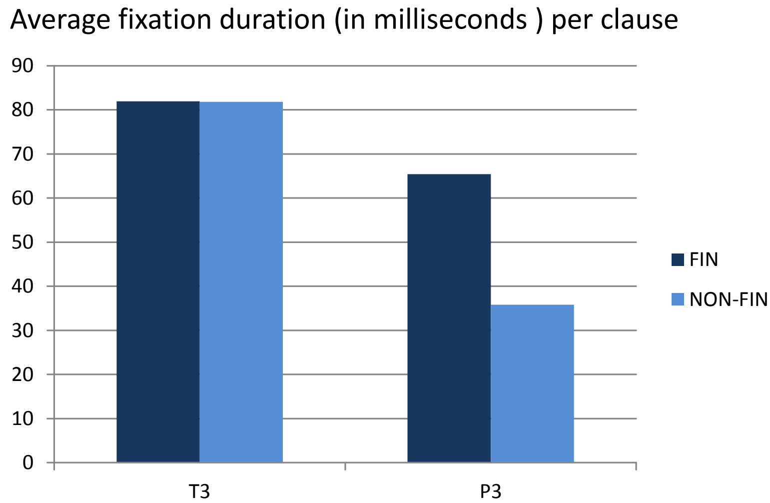




Fixation duration of clauses



Fixation duration of clauses



<u>Quality of MT for non-finite clauses</u>

ST: to end the suffering
ST: Although emphasizing that
ST: to protest against
ST: in the wake of fighting flaring up again in Dafur

TT-P: um das Leiden zu beenden
TT-P: Obwohl betont wird, dass
TT-P: um gegen … zu protestieren
TT-P: im Zuge des Kampfes gegen ein erneutes Aufflammen in Darfur

What‘s next?

- Analysis of other contrastive differences and gaps
- Analysis of ambiguities and processing problems
- Comparison of complexity levels
- Analysis of monitoring processes during TT production (with Translog)
- Comparison of professionals vs. semi-professionals
- Correlations between process data and the quality of the participants’ outputs
- Comparison with other translation pairs

Bibliography

- **Carl**, Michael and Jakobsen, Arnt Lykke (2010): Relating Production Units and Alignment Units in Translation Activity Data, In *Proceedings of International Workshop on Natural Language Processing and Cognitive Science (NLPCS)*, Madeira, Portugal.
- **Doherty**, Stephen, O'Brien, Sharon and Carl, Michael (2010): Eye tracking as an MT evaluation technique. *Machine Translation*, 24, 1, pp1-13.
- **Hvelplund**, Kristian Tangsgaard (2011): *Allocation of cognitive resources in translation an eye-tracking and key-logging study*. PhD thesis, Department of International Language Studies and Computational Linguistics, Copenhagen Business School
- **Hyrskykari** , Aulikki (2006): *Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid*. Dissertation, Tampere University Press.
- **O'Brien**, Sharon (2011): Towards Predicting Post-Editing Productivity. *Machine Translation*, 25, 3, pp197-215.
- Postediting in Practice. A TAUS Report, March 2010 p.6
- **Veale**, T. and Way, A. (1997). *Gaijin: A Bootstrapping Approach to Example-Based Machine Translation.* Recent Advances in Natural Language International Conference, 239-244.

Contacts

- Silvia Hansen-Schirra
  (hansenss@uni-mainz.de)
  http://www.staff.uni-mainz.de/hansenss/

- Silke Gutermuth
  (gutermsi@uni-mainz.de)

4.4 Evaluation of English → Spanish

This abstract was presented at the ETP workshop, Copenhagen, August 2012

Translating vs Post-Editing: A pilot study on eye movement behaviour across source texts

Bartolomé Mesa-Lao
Universitat Autònoma de Barcelona, Spain barto.mesa@uab.cat

New technologies are creating new translation workflows as well as new professional profiles. Post-editing is gradually becoming one of the most requested services in localisation as opposed to full human translation. Major language service providers now pre-translate source texts using existing translation memories and then automatically translate the remaining text using a machine-translation engine. This hybrid pre-translated text is then given to human translators to post-edit. Following guidelines the post-editors correct the output from machine translation to produce a target text with different levels of quality. The main purpose of this pilot study is to explore the differences between translation and post-editing of texts through the analysis of user activity data. A group of ten professional translators translated and post-edited four different texts from English into Spanish while their eye movements were being tracked. Each participant translated two texts from scratch and post-edited two further texts using a first machine translation draft. Our aim and interest when comparing these two different modalities was ultimately to study the effects on eye movements when reading the same text for two different purposes, i.e. translation vs. post-editing. Research was devised so as to find out to what extent reading a source text while translating results in different degrees of visual attention in comparison with the attention devoted to it by the translator while post-editing a machine-generated translation of the same text. Four different measures were registered during the translation process in order to make comparisons between reading for translation and reading for post-editing: 1) task time, 2) fixation frequency, 3) total gaze time duration, and 4) transitions across source and target areas on the monitor screen. If differences were found between reading for translation and reading for post-editing, we would certainly have empirical data to start thinking about what the actual role played by the source text is in post-editing. Similarly, we could evaluate how much attention it deserves when designing computer-aided translation interfaces which integrate post-editing tasks as part of their translation workflow. Preliminary results show significant differences in the way translators approach the source text when it comes to translating or post-editing it.

4.5 The CRITT TPR-DB V1.0

Reproduction of AMTA workshop paper 2012

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 Translog [1], 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 Translog[1]. 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-DB[2]. 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 TrEd[4]. 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 R[5].

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 CRITT[6] has developed a data acquisition software, Translog (Jakobsen and Schou, 1999, Carl 2012) with which translators' keystroke and gaze activities can be recorded[7]. This tool is now the most widely used tool of its kind (Jakobsen, 2006).

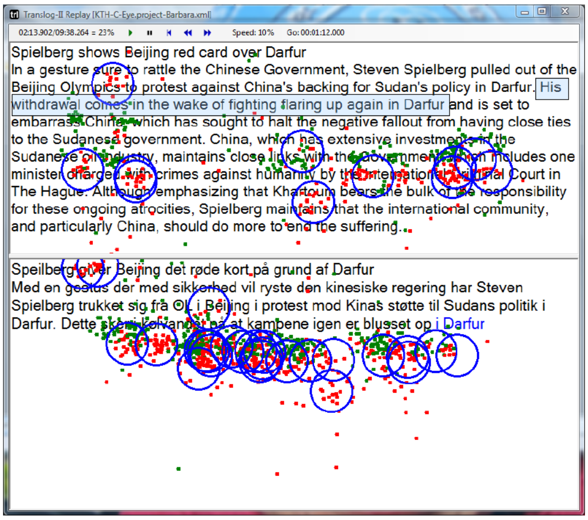


Figure 1: Screenshot of Translog-II replay: fixations in blue circles

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 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

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 respectively.

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

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 or the TPR summer school 2011.

The following four studies were conducted in the context of the CASMACAT[9] 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 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 visiualization of translation progression graphs

[8] http://cogs.nbu.bg/eye-to-it/

[9] http://www.casmacat.eu/

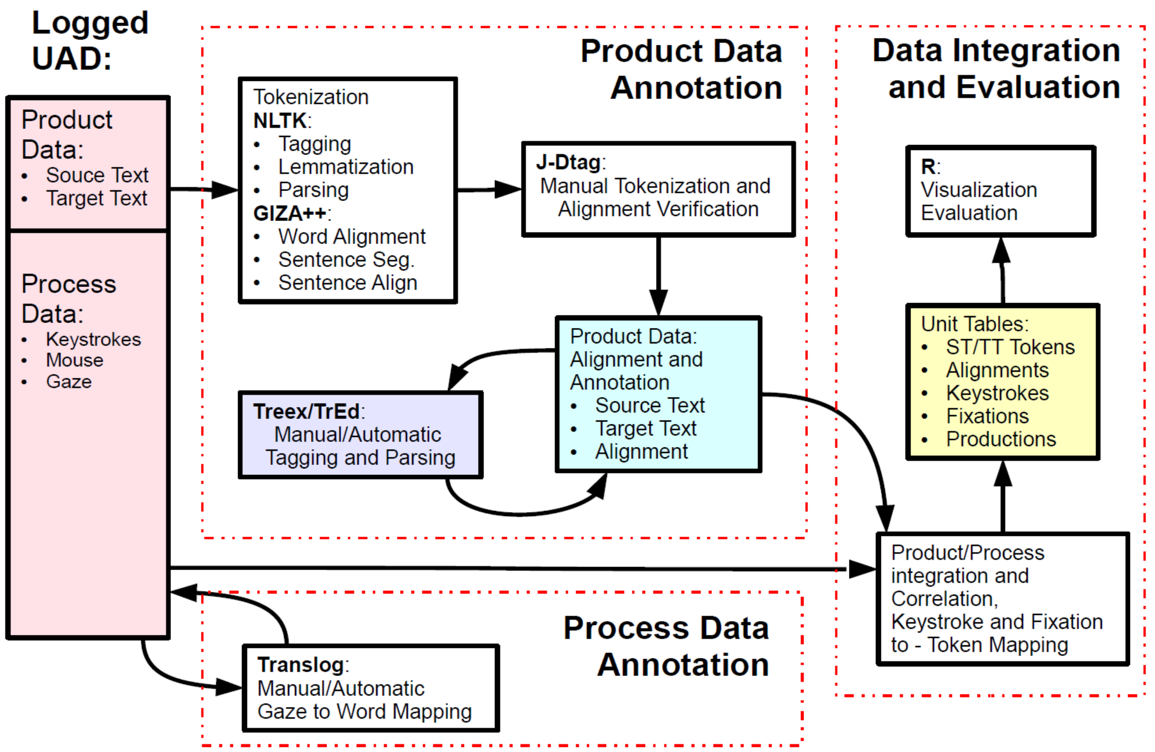


Figure 2: Diagram for the compilation of CRITT TPR-DB V1.0: from the logged UAD is semi-automatically generated 1. aligned and annotated product data, 2. treex representations and 3.unit tables.

Note that the CRITT TPR-DB follows a consistent naming strategy for the folders and files. To annonymise 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 Treex, 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 tools[10], 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.

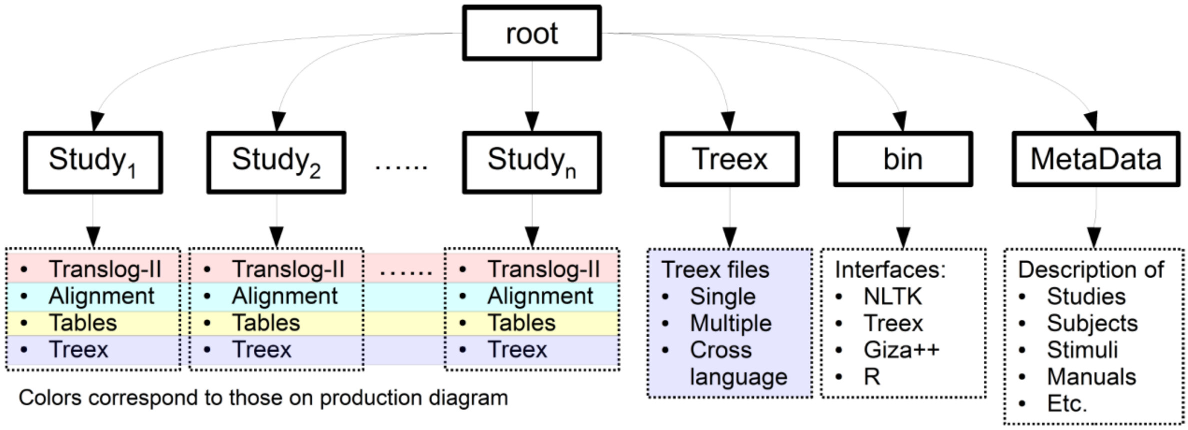


Figure 3: Representation of the CRITT TPR-DB V1.0: the initial Translog-II logging data is enriched with alignments and annotations, as well as with MetaData. Further studies and recordings can be added and processed by a set of programs and scripts in the bin folder.

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 schema[11]. 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:

- 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/

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 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).

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	udviklingl[l]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_

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, an 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 line[12], as shown in figure 4.

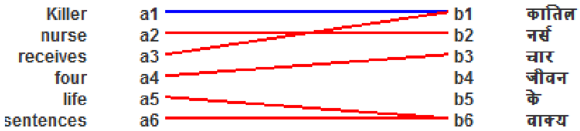


Figure 4: example of alignment visualization in Jdtag

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 Treex[13] programming interface. Figure 5 shows an example of the GUI.

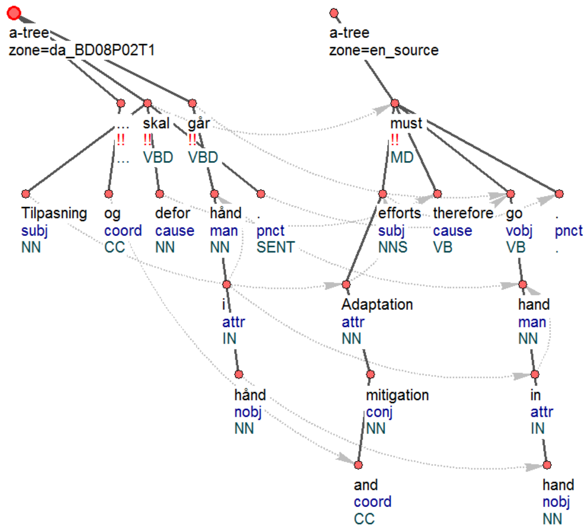


Figure 5: Example of dependency tree alignment and annotation in TrEd

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.

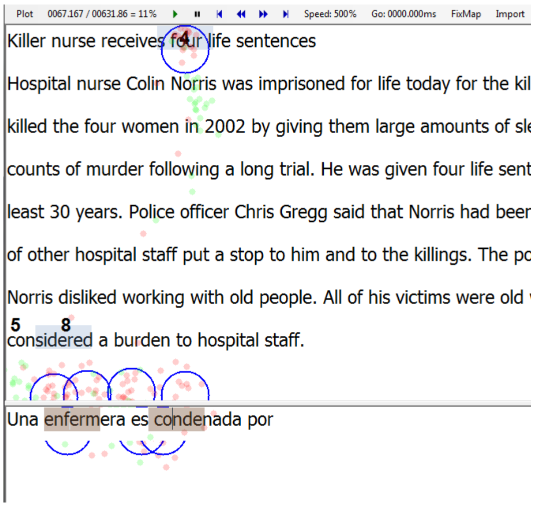


Figure 6: manual fixation correction in Translog-II: erroneous gaze-to-word mapping caused by gaze drift of can be manually.

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.
- 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.

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.

- 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. Kopenhagen: 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.

4.6 Tables and Features in the CRITT TPR-DB

The following paper is submitted to COLING workshop 2012 “Eye Tracking and Natural Language Processing”

Gaze and Keystroke Data in the TPR-DB: Feature Representation of Base and Complex Units

Michael Carl
Institute for International Business Communication,
Copenhagen Business School, Denmark
mc.ibc@cbs.dk

ABSTRACT

The CRITT Translation Process Research Database V1.0 (TPR-DB) was released in August 2012. It contains more than 450 text production sessions, including translation, post-editing, editing and text copying. For each session, keylogging and for most of them also eye-tracking data was recorded. The data was compiled into a consistent format for analysis and visualization of the product and process data.

This paper describes the feature tables of the CRITT TPR--DB. The TPR--DB currently distinguishes between seven different types of units: base units are keystrokes (insertions and deletions) and fixations on the source or target text. From these base units are derived production units and fixation units which represent sequences of coherent reading and writing. Three text-based units are derived from the final translation product: source and target tokens and alignment units. For each of these seven units, features are generated which describe their textual (product) and temporal (process) properties. For analysis and visualization purposes it thus becomes possible to link those dimensions. The paper describes the features, visualization and analysis of the some feature combinations.

1 Introduction

The Center for Research and Innovation in Translation and Translation Technology (CRITT) aims at building new knowledge of translation and communication processes to provide a basis for technological innovation. In particular, for more than 10 years CRITT has been involved in Translation Process Research (TPR) and developed a data acquisition and visualization toolkit (Jakobsen, 1999, Carl, 2012). Experiments in reading and writing have been collected over the past 5 years and much of the data has recently been compiled and released in a CRITT TPR-DB1.0. This database contains recorded logging data, as well as derived and annotated information. Seven kinds of simple and compound process- and product units are identified which are suited for process research and user modeling. The database provides tables for these seven kinds of units which are characterized by a number of features:

1. Keystrokes: basic text modification operations (insertions or deletions), together with time of stroke, and the word in the final text to which the keystroke contributes.
2. Fixations: basic gaze data of text fixations on the source or target text, defined by the starting time, end time and duration of fixation, as well as character offset and word index of fixated symbol in the source or target window.
3. Production units: coherent sequence of typing (cf. Carl and Kay, 2011), defined by starting time, end time and duration, percentage of parallel reading activity during unit production, duration of production pause before typing onset, as well as number of insertion, deletions.

4. Fixation units: coherent sequences of reading activity, including two or more subsequent fixations, characterized by starting time, end time and duration, as well as scan path indexes to the fixated words.
5. Source tokens: as produced by a tokenizer, together with TT correspondence, number, and time of keystrokes (insertions and deletions) to produce the translation, micro unit information.
6. Target tokens: as produced by a tokenizer, together with ST correspondence, number, and time of keystrokes (insertions and deletions) to produce the token, micro unit information, amount of parallel reading activity during .
7. Alignment units: transitive closure of ST-TT token correspondences, together with the number of keystrokes (insertions and deletions) needed to produce the translation, micro unit information, amount of parallel reading activity during AU production, etc.

The paper describes the units and the features[1] that are extracted from from logged and annotated data. Section 2 describes the two basic keystroke and fixation units. Section 3 illustrates examples of the derived production and fixation units. A special property of those units is parallel and alternating reading and typing behavior which indicates workload of the translator. The idea and the way to assess this property is described in section 4. Section 5 looks into characteristics of units that can be automatically derived from the final translation product: source tokens, target tokens and alignment units. Section 6 exemplifies how the translation construction of these production units can be decomposed into several micro units.

2 Keystrokes and Fixations

KEYid	Time	Type	Cursor	Char	STid	TTid
0	92016	ins	0	E	2	1
1	92172	ins	1	l	2	1
2	92313	ins	2	_	2	1
3	92375	ins	3	e	2	2
4	92563	ins	4	n	2	2
5	92828	ins	5	f	2	2
6	92938	ins	6	e	2	2
7	93047	ins	7	r	2	2
8	93266	ins	8	e	2	2
9	93610	del	8	e	2	2
10	93797	ins	8	m	2	2
11	93875	ins	9	e	2	2
12	93938	ins	10	r	2	2

Table 1: Keystroke information

The Keystrokes table encodes single events in time with no duration, all other tables encode textual/temporal units which strech over parts of one or more words and which have at leat one starting time and a duration, as described below.
The first column in each table is an identifier of the event or unit (KEYid, FIXid, FUid, STid, TTid, PUid, AUid). Successive columns encode various features which characterize the event or unit.

Keystroke data is stored in a file with the extension *.kd. As shown in Table 1, keystrokes have a Time at which they were produced, a Type, indicating whether it was an insertion or deletion, a position in the text (a Cursor offset) at which the text was modified, the actual character (Char) which was inserted or deleted, as well as

[1] Some of the features are only available in the CRITT TPR-DB V1..1

the target text token (TTid) to which the keystroke has contributed and the source text token (STid) of which the TTid is the translation. Note that the TTid refers to the token in the final text. Fixation data is stored in a file with the extension *.fd. During a fixation, the gaze is maintained on a single location. Reading involves fixating on a successive locations across a text, but neither is the eye perfectly steady during fixations, nor do the eyes move smoothly over a text. There are many methods to compute fixations. In Translog-II we currently use a density-driven fixation computation algorithm, which clusters gaze samples within a distance of 60 pixels into a single fixation, if the duration is longer than 40ms. The center of the fixation is then mapped on the closest character using build-in functions.

FIXid	Time	Win	Dur	Cursor	STid	TTid
251	93921	2	250	7	2	2
252	94171	2	150	9	2	2
253	94374	1	183	65	10	13
254	94546	1	267	25	4	5
255	94937	1	100	26	4	5
256	95077	1	184	25	4	5
257	95671	2	400	15	1	3
258	96062	1	316	791	152	170
259	96374	2	200	13	1	3
260	98765	1	217	25	4	5
261	98984	1	283	36	6	6
262	99265	1	217	24	4	5
263	99499	1	100	17	3	4
264	99624	1	116	17	3	4
265	99812	1	982	26	4	5
266	101562	1	1199	32	6	6
267	103812	2	299	32	4	5
268	105780	1	200	38	6	6
269	105999	1	117	425	82	86
270	108062	1	133	185	33	42
271	108359	1	100	54	8	8+9
272	108452	1	333	179	31	39
273	108796	1	133	295	54	62
274	109077	1	200	51	8	8+9
275	109452	1	117	58	9	12
276	110234	1	167	59	9	12

Table 2: Fixation information

The table in Table 2 indicates the beginning of a fixation (Time) and its duration (Dur). The fixation table shows in which window (Win) a fixation was detected, 1 for source text window and 2 for the target text window and the Cursor offset of the closest character at which the center of the fixation was detected. While the cursor offset refers to the text as it emerges, the STid and TTid refer to the source and target text tokens of the final text. Thus at a certain time during text production cursor position 5 of the TT may for instance contain an “a” which is part the word “asesino”. The fixation will be assigned TT4 if “asesino” turns out to be the 4th word in the final translation, irrespectively of where in the text this word occurred when it was fixated. In this way we can count the number of fixations on one word, even if the word changes its locations in the text during the editing process. Note, however, that the precision of this information has to be handled with care, since 1. movements of text fragments, particularly deletions, can be traced only very imprecisely, and 2. fixations and their mapping on the symbols may be quite noisy, due to different reasons of fixation drift.

3 Production and fixation units

Production units (PUs) are sequences of coherent typing activity (cf. Carl and Kay, 2011) which are stored in a file with the *.pu extension. A production unit boundary is defined as a delay of 1000ms or more without keyboard activity. It is assumed that coherent typing is interrupted beyond this delay of time, with a likely shift of attention towards another text segment. As a coherent temporal/textual segment PUs have a temporal beginning (Time) and a duration (Dur),

PUid	Time	Dur	Pause	Paral	Ins	Del	STid	TTid	Edit
0	92016	7250	1140	37.85	34	7	1+2+3	1+2+3+4	El_enfere[e]mero_asesiono_ re[er_ono]no_recibe
1	100406	1313	1875	29.55	8	0	3+4	4+5	_cuatro_
2	103594	4187	13735	0	23	3	4+5	5+7	sentencias_de_vida.__[__.]__

Table 3: Production units

and as they cover one or more insertion or deletion keystrokes (Edit operations) which contribute to build up one or more target text tokens (TTid). In the example in Table 3, the sequence:

El_enfere[e]mero_asesiono_re[er_ono]no_recibe

was typed within 7250ms, starting at time 92016 with no inter-key delay of more than 1000ms. A delay (Pause) of 1140ms follows this typing sequence before the next PU starts at Time 100406ms. The table 3 also indicates the number of insertions and deletions of the PUs. PU_0 contains 34 insertions and 7 deletions. The latter are within square brackets and must be read in the reverse direction. Thus, the substring "[er_ono]" is actually the deletion "ono_re" which reflects the correction of:

asesiono_re --> asesino_recibe

Note that PU_1 "_cuatro_" accounts for two target words (TT_{4+5}), as the blank, represented by an underscore "_" already counts as part of the next word.
Similar to PUs, Fixation Units (FUs) indicate sequences of coherent reading behavior and are stored under the file extension *.fu. Based on experiments in (Carl and Kay, 2011) we define a boundary between two successive FUs if a gazing pause is longer than 400ms. That is, if the stream of gaze samples indicates the gaze directs away from the screen for more than 400ms,

FUid	Time	Dur	Pause	Paral	Path
11	93921	1340	410	100	2:2+2:2+1:10+1:4+1:4+1:4+
12	95671	903	2191	100	2:3+1:152+2:3+
13	98765	2029	768	43.81	1:4+1:6+1:4+1:3+1:3+1:4+
14	108062	1507	665	0	1:33+1:8+1:31+1:54+1:8+1:9+

Table 4: Four fixation units

thus interrupting coherent reading activity, we assume a boundary of a fixation unit and the beginning of the next fixation. This may happen, for instance, when the gaze is shifts away from the screen to the keyboard, or to some other places.

Table 4 shows four FUs (FU_{11} to FU_{14}). As with the PUs, the Time indicates the beginning of the FU while the duration (Dur) indicates its length. The fixation path is a sequence of fixations on the source window (1) or the target window (2) and the word ID looked at. The path consists of one or more fixations indicated by Window:WordID where successive fixations are separated by a "+". The first FU in Table 4 (FU_{11}) shows a sequence of six fixations, first on the second word in the target window "enfermo" (2:2), followed by a number of fixations on fourth source word "four" (1:4). On the way from the target text word "enfermo" to the source text word "four", a fixation on word 10 "Colin" was recorded, which is just one line below the "four". Figure 1 shows the a screenshot of the Translog-II replay at time 98573, just before the start of the third FU. FU_{12} comprises of three fixations (marked by a blue circle), two of which are on word 3 "asesino" in the target text, while one fixation is at the end of the source text on word 152. While this accounts for the measured gaze data, it is more likely that a slight drift causes the second fixation is mapped into the ST window, while the translator was actually looking at the ST word.

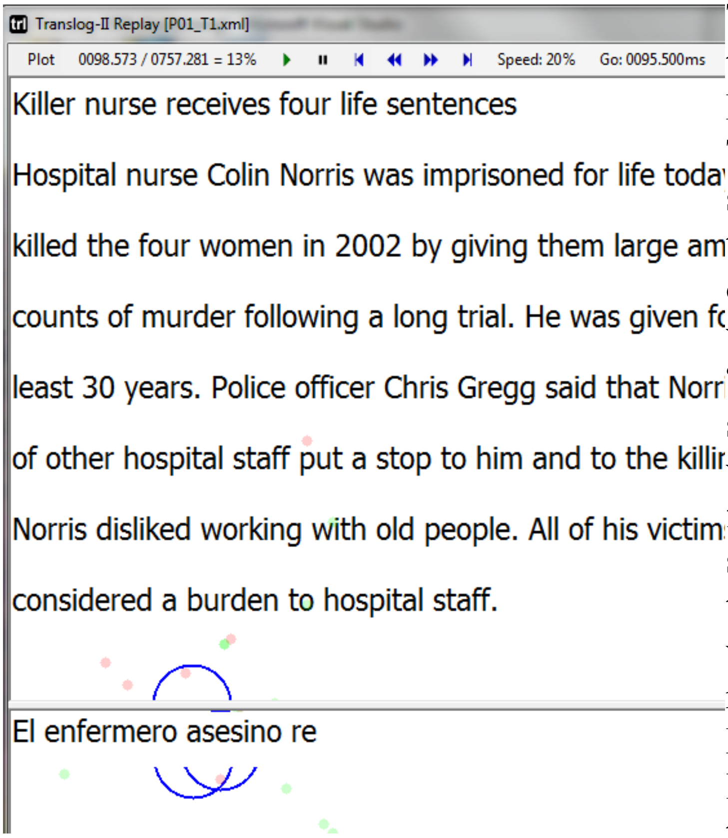


Figure 1: Screen shot of replay situation FU_{12}

The third fixation unit in Table 4, FU_{13} is plotted in Figure 2 and represents a reading sequence of the title (Killer nurse receives four life sentences). It shows how the eyes go back and forth between word 6 ("sentences"), 4 ("four") and 3 ("receives"). As it is not particularly difficult to understand the meaning of the sequence of words, the long reading time of more than 2 seconds (2029ms) suggests that a process of pre-translation takes place during ST reading, in which the translator reflects on how the translation should be rendered.

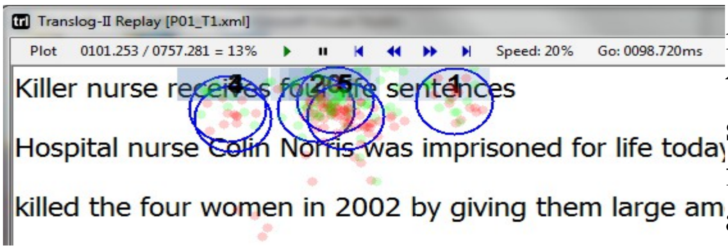


Figure 2: Screen shot of replay situation FU_{13}

Note that the sum of all FU durations may be longer than the sum of all fixation durations, since FUs include inter fixation delays shorter than 400ms which may not be part of any fixation.

4 Parallel and alternating reading and writing

Figure 3 illustrates the overlap of reading and writing activity. It puts into relation the source text (vertical axis) and the translation time (horizontal axis). Insertions are represented in black letters, deletions are red. The progression graph in Figure 3 plots the keystroke data of Table 1, the fixation data from Table 2, as well as the three production units of Table 4 and four fixation units from Table 3. The first part in Figure 3 (approx. Time 92000ms to 94000ms) reproduces the production of words 1 and 2 ("El enfermero") as plotted in Table 1. The linked blue x-es represent the fixations (Table 2). The red horizontally striped boxes indicate PUs while the green boxes represent FUs.

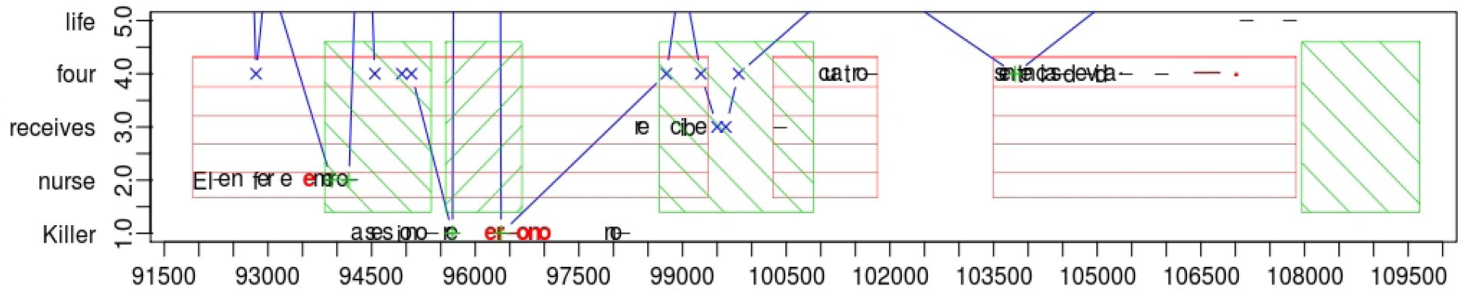


Figure 3: The progression graph shows information from Tables 1 to 4

Reading and writing activity can go on concurrently in parallel. For instance, the FU_{11} between Time 93921 -- 95260 and FU_{12} between 95671 -- 96574 take place while the translator performs a coherent typing activity at the same time generating PU_0. While FU_{11} and FU_{12} overlap 100% with PU_0, FU_{13} between Time 93921 -- 95260 only partially overlaps with two adjacent PU_0 and PU_1. While there is 43.81% overlap with production activity of FU_{13}, FU_{14} has no overlap at all. Progression graphs, as in Figure 1 may thus illustrate in a graphical manner the relation between reading and writing activities.

5 Tokens and Alignment Units

Besides fixation and production units, there are three more units represented in tables: Source Token (*.st), Target Token (*tt) and Alignment Units (AUs). Source and target tokens correspond to sequences of characters, usually separated by a blank, while AUs refer to m-to-n source-to-target token correspondences. The tables provide similar kind of information for these three different kinds of units. These tables contain various information concerning the source/target correspondances, who and how the translation was produced, and information concerning the session. Table 5 shows three English --> Spanish AUs: the column AUtarget contains the TL string, while AUsource has the corresponding SL string. The column "Study" gives the name of the study, "Person" indicates the study unique identification of the translator, the "Text" column indicates which text was translated, and "Task" gives the kind of text production (T: translation, P: post-editing, E: editing).

AUid	AUtarget	AUsource	SL	TL	Study	Person	Text	Task
44	de	of	en	es	BML12	P01	1	T
45	tranquilizantes	sleeping_medicine	en	es	BML12	P01	1	T

Table 5: Alignment unit

Table 6, 7 and 8 are continuations of the AU information. Table 6 gives session information, Table 7 (macro unit) production information and Table 8 decomposes the macro unit in Table 7 into various micro units.

AUid	Session	Draft	Revise
44	757281	92016	290391
45	757281	92016	290391

Table 6: Session information

In Table 6, the column "Session" indicates the duration to the translation/post-editing/editing session, "Draft" shows the lapse of time before the first keystroke was typed, i.e. the end of the orientation phase and beginning of the drafting phase, while "Revise" indicates the time when the drafting phase ended and the revision phase started. This is defined as the end of the first micro unit in which the last token of the text was translated (cf Jakobsen, 2002).

While Table 5 indicates for AU_{44} and AU_{45} that the final translation was "de" and "tranquilizantes" respectively, table 7 shows in the "Edit" column that first "de medicinas para dormir" was typed and later "medicinas para dormir" was again deleted. The table shows the overall number of keystrokes produced: there were 24 insertions, of which 21 characters (the string in square brackets) were later deleted. Even though "medicinas para dormir" and "tranquilizantes" are paraphrases, the former is part of AU44, since deletions are attributed to the preceding word. The time needed to type the translation is given by the duration feature (Dur).

AUid	Len	Ins	Del	Dur	Ratio	GazeT	GazeS	Edit
44	45	24	21	11407	22.5	23100	1245	de_medicinas_para_dormir[rimrod_arap_sanicidem]
45	15	15	0	1610	1	638	412	tranquilizantes

Table 7: AU production information

The editing effort (Ratio) is the ratio of the number of produced characters divided by the length of the final translation. This is equivalent to the number of insertions (Ins) and deletions (Del) devided by their difference: Ratio=Ins+Del/Ins-Del where Ins ≥ Del ≥ 0. Thus, for AU44 the length of the insertion and deletion keystrokes string amounts to 45 which, divided by the length of the final word "of", results in an editing effort of 22.5, while the length of keystroke string to produce "tranquilizantes" in AU45 amounts to the length of the final translation, and thus the editing effort is 1.
GazeT and GazeS indicate the total amount of gaze time on the source unit and the target unit respectively. In contrast to the "Paral" feature in Tables 3 and 8 this is not necessarily during translation production.

6 Micro units

Source and Target tokens, as well as AUs may be characterized by the number and type of micro units by which the translations are constructed. Alves and Vale (2012) refers to recurring editing activities of the same word translations as micro units. For them, "a micro TU is defined as the flow of continuous TT production ... separated by pauses during the translation process". A macro unit, then is a collection of micro units "that comprises all the interim text productions that correspond to the translator's focus on the same ST segment". The TPR-DB computes a micro unit as a coherent typing activity which contributes to the translation of the source or target token, or a AU. While there can be, in principle, any number of micro units (a translator can revise a piece of text very often), only information of the first two micro units is explicitly listed. Tables 8 shows the micro unit information for AU44 and AU45, while their macro unit information is given in table 7. The micro unit is characterised by the actual typing activity (Edit), the starting Time and duration (Dur) of the typing activity, the pause preceding that typing activity, and the amount of parallel reading and writing activity (Paral). Table 8 decomposes the production activity in Table 7 into two micro units: at Time 225703 the translator first types "de medicinas para dormir" in AU44. During a revision more than 4 minutes later, at time 569781 in micro unit2, the string "medicinas para dormir" is deleted and replaced by "tranquilizantes" at Time 570250 which is part of AU45, micro unit1. The duration of those activities is indicated, together with the pause

AUid	Edit1	Time1	Dur1	Pause1	Paral1
44	de_medicinas_para_dormir	225703	11110	187	8.55
45	tranquilizantes	570250	1610	172	60.62
	Edit2	Time2	Dur2	Pause2	Paral2
44	[rimrod_arap_sanicidem]	569781	297	22937	100
45	---	0	0	0	0

Table 8: Micro unit1 and micro unit2

following it and the parallel activity as described in section 4. Given the information in Table 6, we know that revision phase started in this translation session at time 290391, we see that micro unit 1 in AU_{44} takes place during translation drafting, while micro unit2 of AU_{44} and AU_{45} micro unit 1 are both revision events.

7 Conclusion

The paper describes several units and their feature characteristics in the CRITT TPR-DB. We hope that this can be a solid basis for future translation process research.

References

Fabio Alves, Daniel Couto Vale, 2012, *On drafting and revision in translation: a corpus linguistics oriented analysis of translation process data*. Translation: Corpora, Computation, Cognition. Special Issue on the Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation. Volume 2, Number 1. July 2012.www.t-c3.org

Jakobsen, Arnt Lykke (2002) "Translation drafting by professional translators and by translation students." *Empirical Translation Studies: Process and Product*. Copenhagen Studies in Language 27, 191-204.

Jakobsen, A. L. 1999. Logging target text production with Translog. In Hansen, G. (ed.), *Probing the process in translation: methods and results*, *Copenhagen Studies in Language*, volume 24. Copenhagen: Samfundslitteratur. Pages 9–20.

Carl, Michael and Kay, Martin, 2012, *Gazing and Typing Activities during Translation : A Comparative Study of Translation Units of Professional and Student Translators*. I: Meta, Vol. 56, Nr. 4, 2012, s. 952-975.

Michael Carl, 2012, Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research, Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC12),