D1.4: Experimental data gathered in Tasks 1.1, 1.2, 1.4 and 1.5

Michael Carl, Bartolomé Mesa-Lao, Moritz Schaeffer, Mercedes García-Martínez

Distribution: Public

CasMaCat
Cognitive Analysis and Statistical Methods
for Advanced Computer Aided Translation

ICT Project 287576 Deliverable D1.4

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 287576.

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, 2014
Actual date of delivery	November 3, 2014
Date of last update	November 3, 2014
Deliverable number	D1.4
Deliverable title	Experimental data gathered in Tasks 1.1, 1.2, 1.4 and 1.5
Type	Report
Status & version	Draft
Number of pages	9
Contributing WP(s)	WP1
WP / Task responsible	CBS, UEDIN
Other contributors	
Internal reviewer	Jesús González Rubio
Author(s)	Michael Carl, Bartolomé Mesa-Lao, Moritz Schaeffer, Mercedes García-Martínez
EC project officer	Aleksandra Wesolowska
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 deliverable describes the experimental data gathered in Tasks 1.1, 1.2, 1.4 and 1.5, it is related to deliverable D6.5.

Numerous translation and post-editing experiments have been conducted during the CasMaCat project and many of them have been assembled in a Translation Process Database (TPR-DB) which is hosted at the CRITT[1]. The current TPR-DB version 2.0 is an extension of the TPR-DB version 1.0 which was described in deliverable D1.1, Appendix 4.5.

This deliverable gives an overview of the data collected in TPR-DB version 2.0. A more detailed description of the TPR-DB can be found on the TPR-DB website. A description of the structure and the features is provided in a document on the same site from the link `http://bridge.cbs.dk/resources/tpr-db/TPR-DB1.4.pdf`.

Contents

[1] `http://bridge.cbs.dk/platform/?q=CRITT_TPR-db`

1 Overview of the Translation Process Database (TPR-DB)

The CRITT TPR Database is a publicly available database of recorded translation (and other text production) sessions. It contains user activity data (UAD) of translators behaviour collected in almost 30 studies of translation, post-editing, revision, authoring and copying tasks, recorded with the CasMaCat workbench and with Translog-II. Each study consists of between 8 and more than 100 recording sessions. The data amounts currently to more than 430 hours of text production gathered in more than 1,400 sessions and amounts to more than 600.000 translated words in more than 10 different target languages. An overview over the studies is given in section 2. The website makes available the raw raw logging data (> 20GB), as well as a post-processed translation process research database (TPR-DB, zipped 170MB), both under a creative commons license.

Alll studies in the TPR-DB are based on key logging, and a large number also contains eye-tracking data. Each study in the TPR-DB was conducted with a (set of) research question(s) in mind, which con be roughly summarized as follows:

- The TPR-DB contains nine studies conducted with the three different CasMaCat workbences as follows:
  - ALG14: This study compares professional translator and bilinguals while post-editing with the third prototype of the CasMaCat workbench featuring visualization of word alignments.
  - CEMPT13: This study cantains post-editing recordings with the second prototype of the CasMaCat workbench, featuring interactive machine translation.
  - CFT12: This study contains data of the first CasMaCat field trial from June 2012, comparing post-editing with from-scratch translation.
  - CFT13: This study contains data of the second CasMaCat field trial from June 2013, comparing post-editing and interactive machine translation.
  - CFT14: This study contains data of the second CasMaCat field trial from June 2014, comparing interactive machine translation and online learning.
  - JN13: This study is recorded with the second prototype of the CasMaCat workbench featuring interactive machine translation and word alignments.
  - LS14: This study investigates learning effects with interactive post-editing over a period of six week (longitudinal study) with the third prototype of the CasMaCat workbench.
  - PFT13: This study is a pre-field trial test prior to the second CasMaCat field trial.
  - PFT14: This study is a pre-field trial test prior to the third CasMaCat field trial.
- The aim of the MultiLingual experiment is to compare from-scratch translation (T), post-editing (P) and monolingual post-editing (E), for different translators and for different languages. The six English source texts are translated by student and experienced translators; three texts (1-3) are news, three texts (4-5) sociological texts from an ececlopeda. 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 deliverable D1.1, section 2.1.
  - BML12: This study contains translating, post-editing and editing data of six texts from English into Spanish.
  - KTHJ08: This study contains only translation data for the news text 1-3.

  - MS12: This study contains translating, post-editing and editing of the six texts English into Chinese.
  - NJ12: This study contains translating, post-editing and editing of the six texts English into Hindi by professional translators.
  - SG12: This study contains translating, post-editing and editing of the six texts English into German.
  - TDA14: In this study participants were asked to copying the six English texts.
  - WARDHA13: This study contains translating, post-editing and editing of the six texts English into Hindi by students.

- In addition, the TPR-DB contains a few individual experiments that were conducted with Translog-II:

  - ACS08: This study explores the way in which translators process the meaning of non-literal expressions by investigating the gaze times associated with these expressions.
  - BD08: This study involves Danish professional translators working from English into Danish.
  - BD13: This study involves secondary school students translating and post-editing from English into Danish.
  - GS12: This study contains post-editing data of four pieces of news from Spanish into English.
  - HLR13: This is a translation study from English into Estonian (5 participants translating 3 different texts).
  - JLG10: This study investigates L1 and L2 translations from/to English and Brazilian Portuguese.
  - LWB09: This study reports on an eye tracking experiment in which professional translators were asked to translate two texts from L1 Danish into L2 English.
  - MS13: This study is an investigation of translator's behaviour when translating and post-editing Portuguese and Chinese in both language directions.
  - RH12: This is and authoring study for te production of news by two Spanish journalists.
  - ZHPT12: This study investigates translator's behaviour when translating journalistic texts. The specific aim is to explore translation process research while processing non-literal (metaphoric) expressions.

2 TPR-DB tables

The core of the TPR-DB consists of a number of tables, which project the raw logging data into a number of product - and process units in the form of tab separated columnw. These tables can be further processed by various visualization and analysis tools, such as Excel, R, Weka, etc. The TPR-DB can be downloaded free of charge from the TPR-DB website `http://bridge.cbs.dk/platform/?q=CRITT_TPR-db`. A description of the structure and its features is provided on the same site from the link `http://bridge.cbs.dk/resources/tpr-db/TPR-DB1.4.pdf`.

Table 1 and the list in section 1 summarizes the most important studies in the TPR-DB. It shows only studies with one hour or more of total recording duration (*FDur*). Each *Study* is a coherent collection of translation or text production sessions, which can consist of different tasks. The number of sessions per task is given in the *Sess* column. The Table lists in **bold**

Study	Sess	SL	TL	Task	Texts	Part	FDur	KDur	PDur	SLen	TLen
ACS08	30	en	da	T	4	17	3.94	2.8779	1.9332	170	5085
ACS08	30	en	en	C	4	17	1.81	1.7840	1.6013	170	5099
ALG14	8	en	es	P	2	8	2.57	0.4909	0.1747	558	4460
ALG14	8	en	es	PA	2	8	2.77	0.4517	0.1692	558	4460
BD08	10	en	da	T	1	10	1.34	0.7303	0.4480	110	1100
BD13	8	en	da	T	2	8	0.75	0.5221	0.3213	100	786
BD13	10	en	da	P	2	10	0.24	0.1062	0.0569	100	970
BML12	64	en	es	P	6	32	2.31	0.8774	0.4418	141	9012
BML12	63	en	es	T	6	32	8.20	5.7491	3.8062	141	8936
BML12	60	en	es	E	6	30	1.98	0.9259	0.4729	141	8468
CEMPT13	20	en	pt	PIA	2	20	5.96	1.6028	0.5279	330	6323
CEMPT13	20	en	pt	P	2	20	5.29	1.3740	0.5732	330	6494
CFT12	40	en	es	P	22	5	4.01	0.0025	0.0004	784	30930
CFT12	49	en	es	T	24	5	12.78	0.0042	0.0008	814	43514
CFT13	27	en	es	R	26	4	7.11	1.9191	0.4413	1002	26919
CFT13	27	en	es	PI	9	9	28.99	9.6174	3.3044	1176	31752
CFT13	27	en	es	P	9	9	27.07	8.3450	3.5100	1179	31294
CFT13	27	en	es	PIA	9	9	35.04	10.7897	3.9125	1183	31838
CFT14	14	en	es	R	14	4	4.54	1.2171	0.2828	2901	40614
CFT14	7	en	es	P	2	7	16.51	7.9017	3.4180	2901	20273
CFT14	7	en	es	PIO	2	7	15.68	7.9835	3.4917	2901	20341
GS12	8	es	en	P	4	4	1.05	0.3499	0.1909	307	2458
HLR13	15	en	et	T	3	5	2.24	1.1063	0.6730	102	1535
JLG10	10	en	pt	T	3	5	4.52	2.0716	1.2302	241	2577
JLG10	10	pt	en	T	3	5	4.58	2.0442	1.1718	246	2611
JN13	4	en	de	PIA	2	4	2.69	0.6779	0.2735	648	2590
JN13	4	en	de	P	2	4	2.33	0.5869	0.2189	648	2590
KTHJ08	69	en	da	T	3	24	6.45	5.4536	3.8183	153	10571
LS14	60	en	es	PI	24	5	53.81	21.1103	9.5166	1202	72109
LS14	60	en	es	P	24	5	51.43	17.1049	7.4178	1202	72126
LWB09	40	da	en	T	3	18	3.21	2.7843	2.0511	142	5652
MS12	19	en	zh	P	6	11	1.76	0.4634	0.0497	141	2708
MS12	15	en	zh	T	5	10	2.98	1.0167	0.1088	137	2061
MS12	10	en	zh	E	5	8	0.48	0.1491	0.0183	137	1295
MS13	16	zh	pt	P	2	16	2.10	0.9024	0.4443	88	1410
MS13	16	pt	zh	T	2	16	1.90	0.7261	0.1161	86	1386
MS13	22	zh	pt	T	2	22	3.75	2.1343	1.2265	88	1938
MS13	18	pt	zh	P	2	18	1.85	0.6443	0.0934	86	1555
NJ12	39	en	hi	T	6	20	13.04	7.4243	3.3156	141	5505
NJ12	61	en	hi	P	6	20	14.20	6.6184	3.0615	141	8581
PFT13	9	en	es	P	1	9	1.19	0.3110	0.1406	339	3035
PFT13	19	en	es	PI	1	19	2.74	0.8291	0.4267	355	6689
PFT13	16	en	es	PIC	3	16	2.06	0.7679	0.1518	272	5396
PFT13	15	en	es	PIO	3	15	1.82	0.4683	0.0669	363	4611
PFT13	16	en	es	PIL	3	16	1.57	0.3929	0.1511	363	5572
PFT14	3	en	da	PIVO	2	3	2.15	0.6775	0.1622	1051	3245
PFT14	2	en	da	PIVA	1	2	2.02	0.7255	0.1843	1143	2286
PFT14	2	en	da	PIV	2	2	1.98	0.7667	0.1905	1080	2161
RH12	2	es	es	A	2	2	2.02	0.9443	0.6398	604	1207
SG12	46	en	de	E	6	23	4.29	1.7912	0.9342	142	6522
SG12	45	en	de	P	6	23	5.60	1.9265	1.0550	142	6352
SG12	47	en	de	T	6	24	9.39	4.6145	2.9421	142	6632
TDA14	48	en	en	C	6	8	3.60	3.4924	2.6617	142	6792
WARDHA13	34	en	hi	T	6	18	12.72	3.5562	0.5553	142	4832
WARDHA13	31	hi	hi	C	6	18	10.83	5.1386	0.7569	141	4365
WARDHA13	27	en	hi	P	6	15	6.43	1.8991	0.4418	141	3780
ZHPT12	12	zh	pt	T	1	12	3.16	1.4560	0.8510	92	1104
Total	1422	5	15	9	338	319	427h	165h	75h	518	606996

Table 1: Summary information of the TPR-DB v.2.0: number of sessions, different texts, participants in study, translation direction, task and production duration (FDur, KDur, PDur) as well as average source text length (SLen) and total produced target language words (TLen)

studies recorded with the CasMaCat GUI. All other studies are recorded with Translog-II. Study names in *italics* are part of the multilingual translation collection (see deliverable D1.1, Appendix 4.1), in which three to six short English source texts were translated into various different languages by a large number of different translators.

During each session a particular *Task* is conducted, as follows:

- A: Authoring of a journalistic text. Source and target languages are identical.
- C: Copying a text (manually) from the source window into the target window. Source and target languages are identical.
- E: Editing of post-editing of MT output without access to the source text (monolingual post-editing).
- P: Traditional post-editing of MT output (no additional help is provided during the process).

  Within the CasMaCat context, a large number of different post-editing settings were investigated:

  - PA: Traditional post-editing visualizing source (ST) and target (TT) alignment links (triggered by mouse or cursor).
  - PI: Advanced post-editing through interactive translation prediction (ITP) / interactive machine translation.
  - PIA: Advanced post-editing through ITP showing ST-TT alignments (visualization option).
  - PIC: Advanced post-editing through ITP showing ST-TT alignments (visualization option).
  - PIO: Advanced post-editing through ITP and online learning techniques.
  - PIL: Advanced post-editing through ITP showing the unpost-edited text (suffix) in grey (visualization option).
  - PIV: Advanced post-editing through ITP showing Search&Replace bar, alignments and mouse-triggered alternative ITP options.
  - PIVA: Advanced post-editing through ITP and active learning techniques.
  - PIVO: Advanced post-editing through ITP and online learning techniques.

- R: Review of post-edited text.
- T: Translation 'from-scratch'.

As can be seen from Table 1, within one study there can be various different tasks. Each task is often conducted with several different (source) *Texts*, and in most cases each source text is processed (i.e. translated or post-edited, etc.) by several different participants *Part*. For instance, the CFT14 study consists of three tasks (P, PIO, R), where the P and the PIO tasks are based on two different English source texts and 7 post-editors. Each of the conditions (P and PIO) counts 7 sessions, meaning that 7 translations were produced. 4 participants were involved in the revision (R) task, who reviewed the 14 post-edited texts. That is, 7 post-editors produced 14 translations for 2 different English source texts under two conditions which were subsequently reviewed by 4 reviewers.

Table 1 lists three different ways of measuring total production and typing duration of all the translations produced in hours as computed in the TPR-DB:

1. *FDur*: production time of segment, excluding pauses > 200 seconds.
2. *KDur*: duration of coherent keyboard activity excluding keystroke pauses > 5 seconds.
3. *PDur*: duration of coherent keyboard activity excluding keystroke pauses > 1 second.

For example, in study CFT14 for each of the P and PIO tasks were needed 16.5 hours and 15.68 hours *FDur* duration, respectively, for the production of the 7 translations in either condition. Durations of coherent typing activity are much shorter, 7.9 hours or 3.4 hours depending on whether *KDur* or *PDur* is considered. *FDur* includes typing and thinking time while *PDur* and *KDur* aim at measuring typing effort.

The column *SLen* shows the average length in words of the source text; the column *TLen* the total number of produced target text words[2]. For instance, the two English source texts in the CFT14 study have on average each 2,901 words, and are thus the longest source texts in the TPR-DB. The 7 translations of these texts resulted in 20,273 target text words in the P condition and 20,341 words in the PIO condition. These 14 post-edited texts were then subsequently revised which yields a total of 40,616 target text words.

Overall the TPR-DB contains more than 430 hours of production time in terms of *Fdur* duration. In the 1,422 sessions were involved 694 translators producing all together more than 600,000 words in 15 different languages (including those from studies that are not shown in the Table 1).

The language pair en → es is the by far the largest language represented in TPR-DB, with 630 sessions, 490,000 target words and more than 290 hours of *FDur* production time. The second most represented language pair is en → hi with 161 sessions, more than 20,000 tokens in the Hindi translations and more than 46 hours of *FDur* production time. The third language pair is en → de with 146 sessions, more than 24,000 tokens in the German translations and more than 24 hours of *FDur* production time production time, followed by en → da with 127 sessions, more than 18,000 tokens in the Danish translations and 12 hours of *FDur* production time. The rest of the language pairs in the TPR-DB involve more than 20 translation directions in 7 different source and 16 target languages.

3 References

- Michael Carl, Moritz Schaeffer. "The CRITT Translation Process Research Database v1.4", *The Bridge: Research Platform. Department of Department of International Business Communication (IBC)*. Copenhagen Business School, 2014. Available at: `http://bridge.cbs.dk/resources/tpr-db/TPR-DB1.4.pdf`.
- Michael Carl. "The CRITT TPR-DB 1.0: A Database For Empirical Human Translation Process Research", *Workshop on Post-editing Technology and Practice*, AMTA 2012.
- Michael Carl; Mercedes Garca Martnez; Bartolom Mesa-Lao. "CFT13: A Resource for Research into the Post-editing Process". In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC'14). (eds.) Nicoletta Calzolari; Khalid Choukri; Thierry Declerck; Hrafn Loftsson; Bente Maegaard; Joseph Mariani; Asuncion Moreno; Jan Odijk; Stelios Piperidis. Paris: ELRA, 2014, p. 1757-1764.

[2]It is important to note that the total number of tokens as computed by the TPR-DB and the CasMaCat workbench itself may vary slightly due to tokenization differences.

- Nancy Underwood, Bartolom Mesa-Lao, Mercedes Garca-Martnez, Michael Carl, Vicent Alabau, Jess Gonzlez-Rubio, et al. Evaluating the Effects of Interactivity in a Post-Editing Workbench". In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC'14). (eds.) Nicoletta Calzolari; Khalid Choukri; Thierry Declerck; Hrafn Loftsson; Bente Maegaard; Joseph Mariani; Asuncion Moreno; Jan Odijk; Stelios Piperidis. Paris: ELRA, 2014, pp. 553-559, 2014.