D7.3: Third report on dissemination activities

Philipp Koehn, Francisco Casacuberta, Michael Carl

Distribution: Public

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	D7.3
Deliverable title	Third report on dissemination activities
Type	Report
Status & version	Final
Number of pages	14
Contributing WP(s)	WP7
WP / Task responsible	UEDIN, UPVLC, CBS, CS
Other contributors	
Internal reviewer	Philipp Koehn
Author(s)	Philipp Koehn, Francisco Casacuberta, Michael Carl
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

Workpackage 7 comprises of dissemination activities of the CASMACAT project. In this report, we summarize the promotion of project goals, progress and outcomes to the larger academic research community, the commercial sector targeted by the work, and beyond.

Contents

1 Organization of Dissemination Events (Task 7.1)

1.1 Shared Task on Quality Estimation at Workshop on Statistical Machine Translation

In collaboration with the EU Project QTLaunchPad, we organized shared task on quality estimation, related to tasks in Workpackage 3 on confidence measures. The data gathered from the first field trial was prepared as training and test sets and used by 13 participating research groups. UPVLC and UEDIN participated in the shared task in a joint submission with FBK.

1.2 Organization of Workshops

The CASMACAT consortium organized a **Workshop on Humans and Computer-assisted Translation (HaCAT)** [1] in the context of EACL, 26-27 April 2014 in Gothembourg, Sweden. Several CASMACAT papers were presented, as well as a CASMACAT demo. The workshop attracted around 30 participants.

CBS organized the **Fourth PhD Course in Translation Processes Research**[2] from July 7 to 10, 2014. As in the previous years, the PhD course focused on theoretical aspects of translation process research, on experimental research design and methodology, on data visualization and human translation modeling, and on qualitative and quantitative data analysis.

Following the PhD course, a **Workshop on Translation, Bilingualism and Translation Technology**[3] was held on July 11 2014 which attracted around 30 participants.

The CASMACAT project organized jointly with the MATECAT project the **Workshop on Interactive and Adaptive Machine Translation (IAMT)**, co-located with the eleventh biennial conference of the Association for Machine Translation in the Americas in Vancouver, Canada. The workshop brought together researchers with interests in the core problems of the CASMACAT project, and featured an invited talk and paper presentations by the project.

In collaboration with the NICT/Japan, CBS organized the **2nd Workshop on Future Directions in Translation Research**[4] on October 8th, 2013 in Osaka, Japan. Philipp Koehn and Michael Carl gave invited lectures.

2 Scientific Dissemination (Task 7.3)

2.1 Academic Meetings

The CASMACAT project hosted an exhibitor booth at the **Annual Conference of the European Association for Machine Translation.** The booth was open throughout the conference and the partners UEDIN, UPVLC, and CBS presented the project and demonstrated the workbench.

UEDIN organized the **9th Workshop on Statistical Machine Translation**[5] in Baltimore, USA, August 8–9, 2013, collocated with the Annual Meeting of the Association for Computational Linguistics (ACL). The workshop is partly funded by the FP7 CSA MosesCore. The workshop featured for a second time a shared task on quality estimation, based on data

---

[1] https://sites.google.com/site/hacat2014/

[2] http://bridge.cbs.dk/platform/?q=TPR2014

[3] http://bridge.cbs.dk/platform/?q=TDAworkshop

[4] http://www.mastar.jp/wfdtr/index-e.html

[5] more information at http://www.statmt.org/wmt14/

generated by the first CASMACAT field trial. Several representatives of partners of the CASMACAT project participated in the challenge and the workshop at large.

CBS organized a summer project on **Translation Data Analytics (TDA)**[6] from July 13 to August 15, 2014. The goal of the summer project on which 19 students from all over the world participated, was to explore and analyse user activity data which is collected in advanced man-machine communication situations. A workshop report is attached in Appendix

2.2 Invited Talks and Lectures

- Francisco Casacuberta: "Interactive MT", Ninth MT Marathon, Trento, Italy, 2014.

- Jesús González-Rubio: "Online and Active Learning for Machine Translation and Computer-Assisted Translation", AMTA Workshop on Interactive and Adaptive Machine Translation, Vancouver, Canada, 2014.

- Michael Carl: "From CasMaCat to SEECAT: Patterns of Interaction in Advanced Computer-Assisted Translation", 9 April, Yandex, Moscow, 2014. `https://events.yandex.ru/events/science-seminars/carl-09apr/`

- Philipp Koehn: "Machine Translation and Human Translation", Institute of Translation & Interpreting, Scottish Network, Edinburgh, Scotland, March 2014.

- Robin Hill: "Eye movements: A Window on Mind and Brain", British Science Association, Grasmmarket, Edinburgh, June 2014 [invited public talk featuring the translation process].

- Frank Keller & Robin Hill: module at the RefNet Summer School on Psychological and Computational Models of Language Production, Edinburgh, August 2014.

- Robin Hill: guest lecture for Case Studies in Design Informatics, Edinburgh, November 2014.

2.3 Publications

- Vicent Alabau, Jesús González-Rubio, Daniel Ortiz-Martínez, Germán Sanchis-Trilles, Francisco Casacuberta, Mercedes García-Martínez, Bartolomé Mesa-Lao, Dan Cheung Petersen, Barbara Dragsted and Michael Carl: "Integrating Online and Active Learning in a Computer-Assisted Translation Workbench". Proceedings of the Workshop on Interactive and Adaptive Machine Translation (IAMT) at the 11th conference of the Association for Machine Translation in the Americas (AMTA), 2014. pp. 1-8.

- Germán Sanchis-Trilles, Luis A. Leiva. "A Systematic Comparison of 3 Phrase Sampling Methods for Text Entry Experiments in 10 Languages". Proceedings of the 16th international conference on Human-computer interaction with mobile devices and services (MobileHCI), 2014. pp. 537-542.

- Germán Sanchis-Trilles, Daniel Ortiz-Martínez, Francisco Casacuberta. "Efficient Word-graph Pruning for Interactive Translation Prediction". Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT), 2014. pp. 27-34.

- Philipp Koehn, Michael Carl, Francisco Casacuberta, Eva Marcos. "CASMACAT: Cognitive Analysis and Statistical Methods for Advanced Computer Aided Translation". Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT), 2014. pp. 57.

---

[6]`http://bridge.cbs.dk/platform/?q=TDA2014`

- José G.C. Souza, Jesús González-Rubio, Christian Buck, Marco Turchi, Matteo Negri. "FBK-UPV-UEdin participation in the WMT14 Quality Estimation shared-task". Proceedings of the 9th workshop on statistical machine translation (WMT), 2014. pp. 322-328.

- Luis A. Leiva, Vicent Alabau, Verónica Romero, Alejandro H. Toselli, Enrique Vidal. "Context-aware Gestures for Mixed-Initiative Text Editing UIs". Interacting with Computers, 2014.

- Mara Chinea-Rios, Germán Sanchis-Trilles, Francisco Casacuberta. "Online Optimisation of Log-linear Weights in Interactive Machine Translation". Proceedings of the 9th Language Resources and Evaluation Conference (LREC), 2014. pp. 3556-3559

- Nancy Underwood, Bartolomé Mesa-Lao, Mercedes García-Martínez, Michael Carl, Vicent Alabau, Jesús González-Rubio, Luis A. Leiva, Germán Sanchis-Trilles, Daniel Ortiz-Martínez and Francisco Casacuberta. "Evaluating the Effects of Interactivity in a Post-Editing Workbench". Proceedings of the 9th Language Resources and Evaluation Conference (LREC), 2014. pp. 553-559.

- Michael Carl; Mercedes García-Martínez; Bartolom Mesa-Lao, Nancy Underwood. 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

- Michael Carl. Produkt- und Prozesseinheiten in der CRITT Translation Process Research Database In: Translationswissenschaftliches Kolloquium III: Beitrge zur bersetzungs- und Dolmetschwissenschaft (Kln/Germersheim). . ed. /Barbara Ahrens; Silvia Hansen-Schirra; Monika Krein-Khle; Michael Schreiber; Ursula Wienen. Frankfurt am Main : Peter Lang 2014, p. 247-266 (F T S K Reihe A: Abhandlungen und Sammelbaende, No. 65)

- Pascual Martnez Gmez; Akshay Minocha; Jin Huang; Michael Carl; Srinivas Bangalore; Akiko Aizawa. Recognition of Translator Expertise using Sequences of Fixations and Keystrokes In: Procedings of Symposium on Eye Tracking Research and Applications. ed. /Pernilla Qvarfordt; Dan Witzner Hansen. New York : Association for Computing Machinery 2014, p. 299-302

- Laura Winther Balling; Michael Carl. Production Time Across Languages and Tasks : A Large-Scale Analysis Using the Critt Translation Process Database. In: The Development of Translation Competence: Theories and Methodologies from Psycholinguistics and Cognitive Science. . ed. /John W. Schwieter; Aline Ferreira. Newcastle upon Tyne : Cambridge Scholars Publishing 2014, p. 239-268

- Ulrich Germann (Editor) ; Michael Carl (Editor) ; Philipp Koehn (Editor) ; Germn Sanchis-Trilles (Editor) ; Francisco Casacuberta (Editor) ; Robin Hill (Editor) ; Sharon OBrien (Editor) / Proceedings of the Workshop on Humans and Computer-assisted Translation (HaCaT) Stroudsburg, PA : Association for Computational Linguistics 2014, 105 p.

- Sharon O'Brien (Editor) ; Laura Winther Balling (Editor) ; Michael Carl (Editor) ; Michel Simard (Editor) ; Lucia Specia (Editor). Post-editing of Machine Translation : Processes and Applications. Newcastle upon Tyne : Cambridge Scholars Publishing 2014, 335 p.

- Moritz Schaeffer; Michael Carl. Measuring the Cognitive Effort of Literal Translation Processes In: Proceedings of the Workshop on Humans and Computer-assisted Translation

(HaCaT). ed. /Ulrich Germann; Michael Carl; Philipp Koehn; Germn Sanchis-Trilles; Francisco Casacuberta; Robin Hill; Sharon OBrien. Stroudsburg, PA : Association for Computational Linguistics 2014, p. 29-37

- Michael Carl; Moritz Schaeffer. Word Transition Entropy as an Indicator for Expected Machine Translation Quality In: Proceedings of the Workshop on Automatic and Manual Metrics for Operational Translation Evaluation. MTE 2014. ed. /Keith J. Miller; Lucia Specia; Kim Harris; Stacey Bailey. Paris : European Language Resources Association 2014, p. 45-50

- Jakob Elming; Laura Winther Balling; Michael Carl. Investigating User Behaviour in Post-editing and Translation using the CASMACAT Workbench In: Post-editing of Machine Translation: Processes and Applications. . ed. /Sharon O'Brien; Laura Winther Balling; Michael Carl; Michel Simard; Lucia Specia. Newcastle upon Tyne : Cambridge Scholars Publishing 2014,

- German Sanchis-Trilles; Vicent Alabau; Christian Buck; Michael Carl; Francisco Casacuberta; Mercedes Garcia Martinez; Ulrich Germann; Jesus Gonzalez-Rubio; Robin L. Hill; Philipp Koehn; Luis A. Leiva; Bartolom Mesa-Lao; Daniel Ortiz-Martnez; Herve Saint-Amand; Chara Tsoukala / Interactive Translation Prediction vs. Conventional : Post-editing in Practice - A Study with the CasMaCat Workbench. Machine Translation. 2014

- Bartolomé Mesa-Lao. Gaze Behaviour on Source Texts : An Exploratory Study Comparing Translation and Post-editing. In: Post-editing of Machine Translation: Processes and Applications. . ed. /Sharon O'Brien; Laura Winther Balling; Michael Carl; Michel Simard; Lucia Specia. Newcastle upon Tyne : Cambridge Scholars Publishing 2014, p. 219-245

- Bartolomé Mesa-Lao, Francisco Casacuberta, Michael Carl, Robin Hill, Moritz Schaeffer Advanced Post-editing with CasMaCat Post-editing mit CasMaCat ein Profisystem Les progrès de la post-édition avec CasMaCat, XXth FIT World Congress, 2014

- Jesús González-Rubio. "On the Effective Deployment of Currrent Machine Translation Technology". PhD thesis, Universitat Politcnica de Valncia. 2014. Supervised by Dr. Daniel Ortiz-Martínez and Prof. Francisco Casacuberta.

- Vicent Alabau, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Ulrich Germann, Jesús González-Rubio, Robin Hill, Philipp Koehn, Luis A. Leiva, Bartolomé Mesa-Lao, Daniel Ortiz-Martínez, Hervé Saint-Amand, Germán Sanchis-Trilles, Chara Tsoukala. "CASMACAT: A Computer-assisted Translation Workbench". Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2014. pp. 25-28.

- Vicent Alabau, Luis A. Leiva. "Proofreading Human Translations with an E-pen". Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation, 2014. pp. 10-15.

- Luis A. Leiva, Germán Sanchis-Trilles. "Representatively Memorable: Sampling the Right Phrase Set to Get the Text Entry Experiment Right". Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI), 2014. pp. 1709-1712.

- Daniel Ortiz-Martínez, Francisco Casacuberta. "The New Thot Toolkit for Fully-Automatic and Interactive Statistical Machine Translation". Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2014. pp. 45-48.

- Jesús Gonzlez-Rubio, Francisco Casacuberta. "Inference of Phrase-Based Translation Models via Minimum Description Length". Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2014. pp. 90-94.

- Vicent Alabau. "Multimodal Interactive Structured Prediction". PhD thesis. Universitat Politcnica de Valncia. 2014. Supervised by Prof. Francisco Casacuberta and Dr. Alberto Sanchis

- Vicent Alabau, Alberto Sanchis, Francisco Casacuberta. "Improving on-line handwritten recognition in interactive machine translation". Pattern Recognition, 2014. Vol. 47 (3) pp. 1217-1228.

- Jesús González-Rubio, Francisco Casacuberta. "Cost-Sensitive Active Learning for Computer-Assisted Translation". Pattern Recognition Letters, 2014. Vol. 37 pp. 124-134.

- Jesús González-Rubio, José R. Navarro-Cerdan and Francisco Casacuberta. "Empirical Study of a Two-Step Approach to Estimate Translation Quality". Proceedings of the International Workshop on Spoken Language Translation (IWSLT), 2013. pp. 227-234.

- Jesús González-Rubio, Francisco Casacuberta. "Improving the Minimum Bayes Risk Combination of Machine Translation Systems". Proceedings of the International Workshop on Spoken Language Translation (IWSLT), 2013. pp. 219-226.

- Jesús González-Rubio, José R. Navarro-Cerdan, Francisco Casacuberta. "Dimensionality reduction methods for machine translation quality estimation". Machine Translation, 2013. Vol. 27 (3), pp. 281-301.

- Philipp Koehn and Ulrich Germann. "The Impact of Machine Translation Quality on Human Post-Editing", Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation, 2014.

- Philipp Koehn and Chara Tsoukala and Herve Saint-Amand. "Refinements to Interactive Translation Prediction Based on Search Graphs", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014.

- Ulrich Germann. "Dynamic Phrase Tables for Machine Translation in an Interactive Post-editing Scenario", Proceedings of the Workshop on Interactive and Adaptive Machine Translation, pages 20-31, 2014.

- Robin Hill & Frank Keller. "Error detection in native and non-native speakers provides evidence for a Noisy Channel Model of sentence processing", 27th Annual Conference on Human Sentence Processing (CUNY 2014), Columbus, OH, US, 2014.

- Robin Hill & Frank Keller. "Human detection of translation errors in text: unwrapping the dynamic process through eye-tracking", Translation in Transition: between cognition, computing and technology; Copenhagen, Denmark, 2014.

- Robin Hill, Frank Keller, Moritz Schaeffer, et al. "CASMACAT: A live demonstration", Architectures and Mechanisms in Language Processing (AMLaP XX); Edinburgh, 2014.

- Robin Hill, Frank Keller, et al. "CASMACAT: A Computer-Assisted Translation Workbench", DEMOfest, Edinburgh 2014.

3 Coordination of Dissemination Activities (Task 7.4)

Dissemination activities are co-ordinated by UEDIN. The main goals for the third year was to engage potential users of the workbench and to promote the project outcomes.

Promotion of the workbench was done in three main venues: Installation instructions on the CASMACAT web site, a booth and demonstration at the Annual Conference of the European Association for Machine Translation, and the organization of the Workshop on Interactive and Adaptive Machine Translation.

In addition, the installation of the tool by outside parties was promoted in two targeted campaigns, where special hand-holding was provided to early adopters by email. The first such campaign took place in July 2013 with five active participants. We also reached out to the MT@EC project at the European Commission in January 2013 with a visit to their facilities in Luxembourg.

The website maintained by the project (at http://www.casmacat.eu/) is used to promote the public mailing list to harvest email addresses from interested targeted outsiders. At the end of year 3, we have gathered 135 subscribers, and attracted interest from commercial contacts from companies such as Audi, Oracle, Symantec, Autodesk, Asia Online, Crosslang, Locstar, Welocalize, Translate Plus, Abbyy Language Services, Tilde, and also the European Commission.

4 Appendix

Translation Data Analytics 2014
Workshop Report

Michael Carl and Srinivas Bangalore

The Translation Data Analytics workshop 2014 (TDA 2014) was organized from July 14, 2014 to August 15, 2014 at the Spejdercenter Holmen in Copenhagen, Denmark. The goals of the five week workshop was to analyze, model and gain actionable insights from data tracked from multimodal human-machine interaction during language translation activity.

Participation: The workshop had an active international participation from BS, MS and PhD students and researchers from many countries – India (7 students), Germany (2), Canada (2), Russia (2), Belgium (1), Brazil (1), Denmark (1), Norway (1) Spain (1), Switzerland (1), UK (1). The technical background of the participants varied between Computer Science students with Machine Translation and Natural Language Processing knowledge and researchers and practitioners in human language translation. This mixed group of participants provided a balance between translation theorists and software engineers.

Projects: The ten projects pursued during the course of the workshop is listed below.

1. Analysis of translator's behavior based on text genre and user interface variations
2. Cross-lingual study (English-to-Danish, German, Hindi and Spanish) of correlating syntactic variations among translators with gaze fixations
3. Graph-theoretic methods for identifying comprehension and translation segmentation based on eye-gaze transitions
4. Machine learning models for predicting the gaze fixation durations on source text using lexical, syntactic and semantic information
5. Identifying translators from profiles of translator's activity, translator's target text production and session summary statistics
6. Correlating cross-translator variations against variations in output of machine translation
7. Rendering translator activities as images for visual analytics
8. Evaluating efficiency gains between Typing and Speech input
9. Understanding translation errors in machine translation and post-edited texts for English-Dutch translations.
10. Training a Hindi speech recognition system using Kaldi -- an open source speech recognition toolkit.

The assignment of the participants to projects was based on their proficiency and interest for that topic. However, the assignment was broadly left fluid that allowed for the participants expertise to be harnessed by other projects.

Background on data: From the approximately 26 studies in the TPR-DB, participants investigated (mainly) two distinct subsets.

The first set of data consists of the two most recent CasMaCat studies that were conducted from April - June 2014 by Celer and involved together 8 professional post-editors. In a longitudinal study (LS14), five post-editors post-edited 24 texts (4 texts per week, each text with approximately 1000 words) over a duration of six weeks under two different conditions: 'traditional' post-editing (P) and interactive post-editing (PI). The aim of this LS14 study was to assess whether post-editors learn to adapt to the somewhat unusual interactivity mode when using the post-editing tool over a period of six weeks, and compare this with their post-editing progress when using 'traditional' post-editing. An extension of this LS14 data is the second CasMaCat field trial 2014 (CFT14) which was also conducted by Celer with the aim to assess whether post-editors profit from interactive translation as compared, again, to 'traditional' post-editing. Seven post-editors participated in the CFT14 study from which four post-editors were also present at the LS14 study. The CFT14 study is, different from the LS14 study, with the following changes: instead of a news text (as in LS14) in CFT14 medical texts from the EMEA corpus were to be post-edited. Each post-editor post-edited only two texts, one in the conventional post-editing mode, the other with interactive online learning. Texts were much longer, with each text about 4500 words, so as to observe the online learning effect with tokens that would repeatedly be observed. Both studied in combination contain more than 205.000 target text tokens.

The second data set that was investigated during the TDA project consists of six short English source texts that were translated and post-edited into four very different target languages, Danish, German, Hindi and Spanish. The data set was collected over the past six years, and consists of 390 target language texts, so that there are, on average, about 8 translations for each English source text per language and per translation condition (post-editing vs. from-scratch translation). The data was collected with the Translog-II user interface and more than 100 different translators participated in the translation sessions with very different background, from translation students to (occasional) freelance- and full-time professional translators. The number of produced target text tokens in all 390 translations is almost 60.000.

Significant results from projects: The projects produced a number of interesting results, some of which have been presented as technical abstracts and papers that have been submitted to conferences. Other results will be submitted as papers and disseminated at upcoming conferences. For the purpose of this report, we summarize the key results here.

1. Increased correlation between translator's gaze fixation duration with the syntactic variations in the translation. The larger the syntactic variety of target language sentences produced by translators for a source sentence, the larger is the gaze fixation duration on the source text. This correlation effect has been shown for English-Danish and English-German.

2. Novel graph visualization of the translator's activity that helps identify the segments of the sentences where gaze tends to be localized, suggesting that translation proceeds in chunks of source (and target) translation units.
3. Predictive models for source gaze fixation using machine learning models; prediction accuracy is close to 60% when gaze fixations are discretized into 4 groups. Words with significant variability in translated texts cause high gaze fixations.
4. Demonstrated quantitatively that dictating text can be upto 350% faster than typing texts. This result has efficiency implications for speech-based interfaces for translators.
5. Implemented an Open source software (Kaldi) based Hindi Speech Recognizer performing at over 60% word accuracy.
6. Automatic identification of similarity between translators matches with the meta-information provided by the translators. Furthermore, the identification of a translator among 5 translators can be achieved at about 50% accuracy using machine learning-based classification methods leveraging the activity profile, the number of edits during a session, and the translation units produced by a translator.
7. Created generative machine learning models based on Activity sequences and PoS sequences to automatically cluster post-editors which were later verified as accurate based on translator's metadata
8. Created discriminative models based on the features from activity n-grams (bucketed to various sub-activities according to duration), segment summary and production units to make a segment level post-editor identification model that performs at 51% accuracy.
9. Demonstrated a learning effect over a 4 week period. Classification models that could separate experienced post-editors from inexperienced post-editors in week 1 at 71.7% accuracy, could distinguish them only at 57.1% accuracy, suggesting that the user behavior of the two groups of post-editors were more similar by week 5.
10. Quantified the savings in typing effort during post-editing with on-line terminology learning relative to post-editing without terminology learning. Significant time savings in these two post-editing conditions, if time spent to consult resources is factored out.

End-of-Project Presentations: The results of the working groups were presented during the last week of the TDA project in public morning sessions from August 11 to August 14. A number of colleagues for IBC and KU attended the these presentations:

Monday 11/08/2014:

9:30 - 10:15 Annegret Sturm, Bergljot Behrens, Moritz Schaeffer, Arndt Heilmann, Maheshwar Ghankot: *Syntactic entropy*

10:15 - 11:00 Tanik Saikh: *Predicting source text gaze fixation durations - a machine learning approach*

Tuesday 12/08/2014:

9:30 - 10:15 Karan Singla, David Orrego Carmona, Ashleigh Gonzales: *Predicting translator behaviour*

10:15 - 11:00 Akshay Minocha, Alena Konina: *Segmentation in translation: analysis and visualisation*

<u>Wednesday 13/08/2014</u>:
	9:30 - 10:15 Tatiana Fedorchenko, Arlene Koglin, Bartolomé Mesa-Lao, Mercedes García Martinez, Julián Zapata: *CasMaCat Field Trial data analysis 2014*
	10:15 - 11:00 Andreas Søeborg Kirkedal, Dipti Pandey: *Portable ASR systems*

<u>Thursday 14/08/2014</u>:
	9:30 - 10:15 Pintu Lohar, Ambarish Jadhav: *Correlation between human translation entropy and machine translation entropy*
	10:15 - 11:00 Joke Daems: *The usage of external resources in post-editing vs. human translation*

Papers resulting from the workshop: In addition to the above end-of-project presentations, papers and extended abstracts from this workshop have been submitted to several conferences, while other results will be reported in papers, book chapters and journal contributions that are under preparation.

1. Two extended abstracts were submitted to I*nnovation Paths in Translation and Intercultural Studies (IATIS 2015)* in Belo Horizonte, Brazil, and
2. A paper was submitted to *The Third Workshop on Post-editing Technology and Practice (WPTP 2014)*, organized as part of the Association for Machine Translation in the Americas 2014, Vancouver, Canada.
3. A paper on Hindi ASR was submitted to ICON
4. An abstract and an intent to submit a contribution to the *Special Issue of Translation Spaces: Translation as a Cognitive Activity*. deadline beginning of October
5. A submission is planned as a book chapter in "*Translation in Transition: Between Cognition, Computing and Technology*" to be published in the Translation Library Series of John Benjamins Publishing Company, deadline beginning of September
6. An independent book project is planned on "*New directions in Empirical Translation Process Research*", perhaps with Springer

Extra-curricular activities: The participants had a number of organized and ad-hoc social activities during the course of this workshop. A trip to Roskilde at the end of Week 1 -- the Royal Church, Viking boat museum and hands on rowing of Viking boats provided a rich Danish cultural experience to the participants. The participants regularly organized barbeques on most weekends and traded international culinary recipes and had friendly competitions to outdo each other in cooking extensive dinners. The participants benefitted greatly from a number of cultural events (Jazz Festival, KultureHavn) organized during July and August in Copenhagen.

Contrasting 2013 and 2014 workshops: While the duration of the TDA 2014 workshop was nearly half the duration of the Seecat 2013 workshop, the number of participants in the TDA workshop increased by about 50%. Moreover, the background of the participants in the TDA workshop was more varied, translators and software developers, with greater international diversity. The technical diversity presented challenges in terms of their ability to use and manipulate data using computational tools. The expectation of the work culture between these

two diverse groups presented some challenges in the execution of various projects. Despite these challenges, we believe that the diversity enabled the synergistic exploration of novel open-ended ideas than either of the groups could have achieved; in contrast to the Seecat 2013 workshop where concrete implementation projects were explored due to the background of the participants.

Finally, the logistics for the TDA 2014 workshop was significantly simpler than the complex arrangement of Seecat 2013. Having all the workshop participants resident at the Spejdercenter Holmen obviated the need to travel to/from multiple locations which saved time, cost and frustrations for organizers and participants of the workshop. Furthermore, the close proximity of Holmen to Copenhagen city center allowed the participants to explore evening and weekend activities as they chose to. The participants in the TDA 2014 workshop, despite given the choice for light lunches chose to prepare elaborate meals which meant far greater time was spent in procuring groceries, cooking and cleaning, in contrast, to the Seecat 2013 workshop.