



---

## D7.2: Second report on dissemination activities

---

Philipp Koehn, Francisco Casacuberta, Michael Carl

Distribution: Public

---

CASMACAT  
Cognitive Analysis and Statistical Methods  
for Advanced Computer Aided Translation

ICT Project 287576 Deliverable D7.2



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



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, 2013
Actual date of delivery	November 4, 2013
Date of last update	November 4, 2013
Deliverable number	D7.2
Deliverable title	Second report on dissemination activities
Type	Report
Status & version	Final
Number of pages	8
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	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/>

© 2013, The Individual Authors

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

# 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

<b>1</b>	<b>Organization of Dissemination Events (Task 7.1)</b>	<b>4</b>
1.1	Shared Task on Quality Estimation at Workshop on Statistical Machine Translation	4
<b>2</b>	<b>Scientific Dissemination (Task 7.3)</b>	<b>4</b>
2.1	Academic Meetings . . . . .	4
2.2	Invited Talks . . . . .	5
2.3	Publications . . . . .	5
<b>3</b>	<b>Coordination of Dissemination Activities (Task 7.4)</b>	<b>7</b>

# 1 Organization of Dissemination Events (Task 7.1)

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

<http://www.statmt.org/wmt13/> In collaboration with the EU Projects QTLaunchPad and Matecat, 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. Both UPVLC and UEDIN participated in the shared task, the latter in a joint submission with FBK.

# 2 Scientific Dissemination (Task 7.3)

## 2.1 Academic Meetings

The CASMACAT project hosted an exhibitor booth at the **Machine Translation Summit**. The booth was open throughout the conference and the partners UEDIN, UPVLC, and CBS presented the project and demonstrated the workbench. The conference was well-attended by industry and academia alike and the CASMACAT project played a prominent role at the conference also with research paper presentations and invited talks.

UEDIN organized the **7th Workshop on Statistical Machine Translation**<sup>1</sup> in Sofia, Bulgaria, 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 generated by the first CASMACAT field trial. Several representatives of partners of the CASMACAT project participated in the challenge and the workshop at large. The CASMACAT workbench was also used by Charles University with assistance from UEDIN to gather post-editing data on machine translation output from several participants in the translation shared task.

CBS organized the **Third PhD Course in Translation Processes Research** from August 5 to 8, 2013. 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 **Speech and Gaze in Translation** (August 09, 2013, CBS, Denmark) was held which attracted around 40 participants<sup>2</sup>

CBS organized a **Special symposium on Empirical approaches to gaze data analysis in reading, writing and translation** in the context of the 17th European Conference on Eye Movements, August 11–16, 2013. The symposium featured four talks on various aspects of gaze data in reading and translation.<sup>3</sup>

In the Context of the 24th International Conference on Computational Linguistics (COLING 2012), CBS organized a workshop on **Eye-tracking and Natural Language Processing** on 15 December 2012, Mumbai, India.<sup>4</sup>

Following the “10th International Workshop on Natural Language Processing and Cognitive Science” (NLPCS), CBS conducted on October 17th, 2013 a Tutorial on **Empirical Translation Process Research: a methodology for investigating cognitive processes in translation and machine translation post-editing**.<sup>5</sup>

<sup>1</sup>more information at <http://www.statmt.org/wmt12/>

<sup>2</sup><http://www.cbs.dk/en/research/departments-and-centres/department-of-international-business-communication/events/international-workshop-on-speech-and-gaze-in-translation>

<sup>3</sup>[https://www.conftool.pro/ecem2013/index.php?page=browseSessions&presentations=hide&form\\_session=16&CTSID\\_ECEM2013=WGXL80ttzvjIA4BfrevmzgtgZ56](https://www.conftool.pro/ecem2013/index.php?page=browseSessions&presentations=hide&form_session=16&CTSID_ECEM2013=WGXL80ttzvjIA4BfrevmzgtgZ56)

<sup>4</sup><http://bridge.cbs.dk/platform/?q=ETNLP2012>

<sup>5</sup><https://sites.google.com/site/nlpcs2013/home/tutorials>

## 2.2 Invited Talks

- Philipp Koehn: "Machine Translation and Human Translation", keynote, TriKonf, Freiburg, 2013.
- Philipp Koehn: "New Types of Assistance for Translators", Second Workshop on Future Directions in Translation Research, Osaka, Japan, 2013.
- Philipp Koehn: "Advances in Machine Translation for Patent Translation and Beyond", keynote, The 5th Workshop on Patent Translation, Nice, France, 2013.
- Philipp Koehn: "Quality for Whom? Evaluation for Different Uses of Machine Translation", keynote, QTLaunchPad Workshop on Human-Centric Machine Translation and Evaluation, Nice, France, 2013
- Philipp Koehn: "New Tools for Translators", Imperial College, London, United Kingdom, 2013
- Philipp Koehn: "Over the Horizon: New Types of Assistance for Translators", Fondazione Bruno Kessler, Trento, Italy, 2013
- Germán Sanchis-Trilles, Francisco Casacuberta. "Interactive-predictive machine translation: towards the next generation of workbenches for translators, the CasMaCat project". DGA workshop on Multimedia Information Processing 2013. Paris France.
- Daniel Ortiz-Martínez. "Incremental Learning for Statistical Machine Translation". Workshop on Future Directions in Translation Research, Nara, Japan 2012.
- Vicent Alabau. "Special lectures on Multimodal and Interactive Machine Translation and Web Technologies". Workshop on Speech and Eye-tracking enables Computer-Assisted Translation (SEECAT). Copenhagen, Denmark 2013.
- Michael Carl: "Gaze Activity Patterns in Translation and Post-editing", Second Workshop on Future Directions in Translation Research, Osaka, Japan, 2013.
- Michael Carl: "Translation Process Research at the CRITT and the CRITT TPR-Database", Workshop on Future Directions in Translation Research, Nara, Japan 2012.
- Michael Carl: "Translation Process research at the CRITT and the SEECAT summer project", Workshop on Simultaneous Interpretation and Multimodal Translation, AT&T Research, NJ, USA, 2013

## 2.3 Publications

- Ondrej Bojar, Christian Buck, Chris Callison-Burch, Christian Federmann, Barry Haddow, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia: "Findings of the 2013 Workshop on Statistical Machine Translation", Proceedings of the Eighth Workshop on Statistical Machine Translation.
- Vicent Alabau, Ragnar Bonk, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes Garcia-Martinez, Philipp Koehn, Luis Leiva, Bartolome Mesa-Lao, Herve Saint-Amand, Chara Tsoukala, German Sanchis, Daniel Ortiz and Jesus Gonzalez: "Advanced Computer Aided Translation with a Web-Based Workbench", The 2nd Workshop on Post-Editing Technologies and Practice.

- Vicent Alabau, Ragnar Bonk, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes Garcia-Martinez, Jesus Gonzalez, Philipp Koehn, Luis Leiva, Bartolome Mesa-Lao, Daniel Ortiz, Herve Saint-Amand, German Sanchis, Chara Tsoukala: "CASMACAT: An Open Source Workbench for Advanced Computer Aided Translation", The Prague Bulletin of Mathematical Linguistics, Number 100, October 2013, pages 101-112.
- Turan Rustamli: "Paraphrasing in Context of Computer Aided Translation", Master of Science thesis in Artificial Intelligence, University of Edinburgh, 2013.
- Philipp Koehn; Michael Carl; Francisco Casacuberta; Eva Marcos. "CASMACAT : Cognitive Analysis and Statistical Methods for Advanced Computer Aided Translation". In: Proceedings of the XIV Machine Translation Summit. ed. /Andy Way; Khalil Sima'an; Mikel L. Forcada; Daniel Grasmick; Heidi Depraetere. Allschwil : European Association for Machine Translation 2013, p. 411
- Michael Carl. "Dynamic Programming for Re-Mapping Noisy Fixations in Translation Tasks", In: Journal of Eye Movement Research, Vol. 6, No. 2, 2013, p. 1-11
- Bartolomé Mesa-Lao. "The next generation translator's workbench: post-editing in CASMACAT v.1.0." Proceedings of the 34th Translating and the Computer Conference, ASLIB, 29 & 30 November, 2012.
- Barbara Dragsted; Michael Carl. "Towards a Classification of Translation Styles based on Eye-tracking and Keylogging Data", In: Journal of the Writing Research, Vol. 5, No. 1, 6. 2013, p. 133-158
- Laura Winther Balling; Michael Carl. "Scaling up Translation Process Research : Analysis of the Translation Process Using th CRITT TPR Database". In: 7th EST Congress Germersheim 2013, August 29 - September 1, 2013: Translation Studies: Centers and Peripheries. Abstracts. ed. Sebnem Bahadr; Dilek Dizdar; Zahra Samareh. Germersheim : European Society for Translation Studies 2013, p. 210-211
- Michael Carl. "The Casmacat Post-editing Workbench Prototype-II : A Research Tool to Investigate Human Translation Processes for Advanced Computer Aided Translation". In: 7th EST Congress Germersheim 2013, August 29 - September 1, 2013: Translation Studies: Centers and Peripheries. Abstracts. . ed. Sebnem Bahadr; Dilek Dizdar; Zahra Samareh. Germersheim : European Association for Machine Translation 2013, p. 230-231
- Balling Winther, Laura and Michael Carl. 2013.Production Time Across Languages and Tasks: A Large-scale Analysis using the CRITT Translation Process Database, in Aline Ferreira and John Schwieter (eds) The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science, in print
- Carl, Michael, Silke Gutermuth and Silvia Hansen-Schirra. Forthcoming. "Post-editing machine translation a usability test for professional translation settings". In Psycholinguistic and cognitive inquiries in translation and interpretation studies, edited by John W. Schwieter and Aline Ferreira. Cambridge Scholars Publishing
- Elming, Jakob, Michael Carl, and Laura Winther Balling. Forthcoming. "Investigating User Behaviour in Post-editing and Translation Using the CASMACAT Workbench". In Expertise in Post-editing: Processes, Technology and Applications, edited by Sharon OBrien, Michael Simard, Lucia Specia, Michael Carl and Laura Winther Balling. Cambridge Scholars Publishing.
- Moritz Schaeffer and Michael Carl. "Shared Representations and the Translation Process", In Describing Cognitive Processes in Translation: Acts and events, by Ehrensberger-Dow, Maureen, Birgitta Englund Dimitrova, Sverine Hubscher-Davidson and Ulf Norberg (eds.) [TIS 8:2] , scheduled to be published in December 2013.

- Luis A. Leiva, Enrique Vidal. Warped K-Means: An algorithm to cluster sequentially-distributed data. *Information Sciences*, 2013. Vol. 237 (10), pp. 196-210.
- Jesús González-Rubio, Jos R. Navarro-Cerdan, Francisco Casacuberta. Partial Least Squares for Word Confidence Estimation in Machine Translation. 6th Iberian Conference on Pattern Recognition and Image Analysis, (IbPRIA) LNCS 7887, 2013. pp. 500-508. Springer.
- Luis A. Leiva, Vicent Alabau, Enrique Vidal. Error-proof, High-performance, and Context-aware Gestures for Interactive Text Edition. Proceedings of the 2013 annual conference extended abstracts on Human factors in computing systems (CHI EA), 2013. pp. 1227-1232.
- Francisco-Javier López-Salcedo, Germán Sanchis-Trilles, Francisco Casacuberta. On-line Learning of Log-Linear Weights in Interactive Machine Translation. Proceedings of Advances in Speech and Language Technologies for Iberian Languages 2012. Communications in Computer and Information Science. Vol 328. Springer. pp 277-286. 2012
- Daniel Ortiz-Martnez, Germán Sanchis-Trilles, Francisco Casacuberta, Vicent Alabau, Enrique Vidal, Jos-Miguel Bened, et al. The CASMACAT Project: The Next Generation Translator's Workbench. Proceedings of iberSPEECH 2012, pp. 326-334. Madrid, November 21-23 2012.
- Germán Sanchis-Trilles, Francisco Casacuberta. Building task-oriented machine translation systems. Proceedings of iberSPEECH Madrid, pp. 480-489. November 21-23 2012.
- Jesús González-Rubio, Daniel Ortiz, José Miguel Benedí, Francisco Casacuberta. Interactive Machine Translation using Hierarchical Translation Models. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP13), pp. 244-254. October 18-21, 2013 Seattle, USA.
- Jesús González-Rubio, Jos R. Navarro-Cerdan, Francisco Casacuberta. Dimensionality reduction methods for machine translation quality estimation. *Machine Translation*, In press.
- Vicent Alabau, Alberto Sanchis, Francisco Casacuberta. Improving on-line handwritten recognition in interactive machine translation. *Pattern Recognition*. In press.
- Jesús González-Rubio, Francisco Casacuberta. Cost-Sensitive Active Learning for Computer-Assisted Translation. *Pattern Recognition Letters*. In press.

### 3 Coordination of Dissemination Activities (Task 7.4)

Dissemination activities are co-ordinated by UEDIN. The main goals for the second year was to create awareness for the project, the solutions developed so far, and the initial stage of promoting the workbench.

Promotion of the workbench was done in three main venues: Installation instructions on the CASMACAT web site, a booth and demonstration at the Machine Translation Summit, and a paper on the tool and demonstration at the Machine Translation Marathon.

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.

A keynote presentation which features the CASMACAT project prominently at the Trikonf conference that is attended mainly by professional translators was very effective in reaching out to this community and should lead to collaborations for the remainder of the project.

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 2, we have gathered 87 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.