



D7.1: First report on dissemination activities

Philipp Koehn, Michael Carl, Jorge González

Distribution: Public

CASMACAT
Cognitive Analysis and Statistical Methods
for Advanced Computer Aided Translation

ICT Project 287576 Deliverable D7.1



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, 2012
Actual date of delivery	October 30, 2012
Date of last update	October 30, 2012
Deliverable number	D7.1
Deliverable title	First report on dissemination activities
Type	Report
Status & version	Draft
Number of pages	7
Contributing WP(s)	WP7
WP / Task responsible	UEDIN, UPVLC, CBS, CS
Other contributors	
Internal reviewer	
Author(s)	Philipp Koehn, Michael Carl, Jorge González
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/>

© 2012, 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

1	Organization of Dissemination Events (Task 7.1)	4
1.1	Second international PhD course on Translation Process Research	4
1.2	International Workshop on Expertise in Translation and Post-editing Research and Application	4
2	Scientific Dissemination (Task 7.3)	4
2.1	Academic Meetings	4
2.2	Publications	5
3	Coordination of Dissemination Activities (Task 7.4)	6

1 Organization of Dissemination Events (Task 7.1)

1.1 Second international PhD course on Translation Process Research

<http://www.cbs.dk/TPR2012>

Following up on last year's success, the CRITT group (CBS) has organised the second international PhD course in Translation Process Research (TPR) from August 13 to 16, 2012. The course focused primarily on theoretical aspects of translation process research, on experimental research design and methodology as well as on data visualization and human translation modeling. As an addition to last year's course we introduced this year a first release of the TPR Database (CRITT TPR-DB V1.0), which was developed in context of CSMACAT project. In future releases of the summer school (August 2013) we will develop in more detail the practical hands-on sessions based on releases of the TPR-DB material. Participants from UPVLC participated at the summer course.

1.2 International Workshop on Expertise in Translation and Post-editing Research and Application

<http://www.cbs.dk/ETP>

The 2-day ETP workshop followed immediately the TPR PhD summer course with the aim to extend it into a broader context. The goal of the workshop was to bring together translation industry, translation process researchers, as well as translation system designer to examine **Expertise in Translation and Post-editing** from different possible angles. Among other tools, the CSMACAT workbench was demonstrated and several CSMACAT related papers were presented. The workshop got a very positive feed-back, so that the organizers decided to repeat the same schema next year (PhD course followed by a workshop). Particularly positive was the large interest and participation of contributions from the translation industry. Several representatives from UPVLC, UEDIN, Celer, and Translated (Matecat) were present at the workshop, so that a closed meeting was scheduled to discuss integration of the MATECAT and CSMACAT GUIs.

2 Scientific Dissemination (Task 7.3)

2.1 Academic Meetings

CBS organized a symposium on **Eye tracking in Translation and Post-editing Studies** at the Tobii conference **EyeTrackingBehaviour 2012**¹. The symposium brings together speakers from research groups in Ireland (Sharon O'Brien and Stephen Doherty, Dublin City University), Germany (Silke Gutermuth and Silvia Hansen-Schirra, University of Mainz), England (Moritz Schaeffer, University of Leicester) and Denmark (Michael Carl, Copenhagen Business School) using eye-trackers for translation process studies to present and discuss their recent research in the field.

CBS organizes a workshop **Eye tracking and Natural Language Processing**² in the context of COLING 2012, Mumbai. The workshop will bring together interested parties in using eye-tracking devices in NLP-related research, including representation, implementation

¹more information at

<http://www.tobii.com/en/eye-tracking-research/global/news-and-events/tobii-conference/ETB2012/>

²more information at <http://www.cbs.dk/ETNLP>

and algorithms for gaze detection as well as assessment and evaluation of possibilities and quality of gaze data in natural language applications

UEDIN organized the **7th Workshop on Statistical Machine Translation**³ in Montréal, Canada, June 7–8, 2012, collocated with the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT). The workshop is partly funded by the FP7 CSA MOSESCORE. While the workshop has a long history to promote research and development in statistical machine translation, this year a special challenge on quality estimation was posed. Several representatives of partners of the CASMACAT project participated in the challenge and the workshop at large.

UEDIN organized the **7th Machine Translation Marathon**⁴ in Edinburgh, September , partly funded by the FP7 CSA MOSESCORE. The CASMACAT project used this forum to promote research in computer aided translation, including a hands-on project on the integration of word-level confidence measures into the CAT editor that ran during the duration of the marathon. Since representatives from CBS, UEDIN, and UPVLC as well as from the MATECAT project attended this meeting, we also discussed the further development of the common CAT workbench and the streamlining of activities.

2.2 Publications

- Vicent Alabau, Luis A. Leiva, Daniel Ortiz-Martínez and Francisco Casacuberta. User Evaluation of Interactive Machine Translation Systems. Proceedings of the 16th Annual Conference of the European Association for Machine Translation (EAMT), pp. 20-23, 2012.
- Vicent Alabau and Luis A. Leiva. Transcribing Handwritten Text Images with a Word Soup Game. Proceedings of the ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA), pp. 2273-2278, 2012.
- Vicent Alabau and Francisco Casacuberta. Study of Electronic Pen Commands for Interactive-Predictive Machine Translation. Proceedings of the International Workshop on Expertise in Translation and Post-editing Research and Application, 2012.
- Jesús González-Rubio, Alberto Sanchis and Francisco Casacuberta. PRHLT Submission to the WMT12 Quality Estimation Task. Proceedings of the Seventh Workshop on Statistical Machine Translation, North American Chapter of the Association for Computational Linguistics (NAACL), pp. 104-108, Association for Computational Linguistics, 2012.
- Pascual Martínez-Gómez, Germán Sanchis-Trilles and Francisco Casacuberta. Online adaptation strategies for statistical machine translation in post-editing scenarios. Pattern Recognition, 45:9, pp. 3193-3203, 2012.
- Vicent Alabau, Alberto Sanchis and Francisco Casacuberta. On the Optimal Decision Rule for Sequential Interactive Structured Prediction. Pattern Recognition Letters, 33:6, pp. 2226-2231, 2012.
- Germán Sanchis-Trilles. Building task-oriented machine translation systems. PhD Thesis, Universitat Politècnica de València, 2012.
- Jesús González-Rubio, Daniel Ortiz-Martínez and Francisco Casacuberta. Active learning for interactive machine translation. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp. 245-254, Avignon, France, April 2012.

³more information at <http://www.statmt.org/wmt12/>

⁴more information at <http://www.statmt.org/mtm12/>

- Jorge González. A finite-state approach to phrase-based statistical machine translation. Proceedings of the 10th edition of the International Workshop on Finite State Methods and Natural Language Processing (FSMNLP), pp. 95-103, 2012.
- Daniel Martín-Albo Simón, Verónica Romero, Alejandro H. Toselli and Enrique Vidal. Multimodal Computer-Assisted Transcription of Text Images at Character-level Interaction. International Journal of Pattern Recognition and Artificial Intelligence, 26:5, 2012.
- Elming and Bonk, The CASMACAT workbench: a tool for investigating the integration of technology in translation, In Proceedings of the International Workshop on Expertise in Translation and Post-editing - Research and Application, Copenhagen Business School, 2012
- Mesa-Lao, Translating vs Post-editing: A pilot study on eye movement behaviour across source texts, In Proceedings of the International Workshop on Expertise in Translation and Post-editing - Research and Application, Copenhagen Business School, 2012
- Silva, Post-editing integration in a translation agency workflow, In Proceedings of the International Workshop on Expertise in Translation and Post-editing - Research and Application, Copenhagen Business School, 2012
- Michael Carl and Barbara Dragsted. Inside the Monitor Model: Processes of Default and Challenged Translation Production Translation: Computation, Corpora, Cognition, Vol 2, No 1 (2012), 2012.
- 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 Translog-II: a Program for Recording User Activity Data for Empirical Translation Process Research. Paper presented at The Eighth International Conference on Language Resources and Evaluation. 2012, Istanbul, Tyrkiet.
- Michael Carl Translog-II: a Program for Recording User Activity Data for Empirical Translation Process Research. Poster session presented at 13th International Conference on Intelligent Text Processing and Computational Linguistics. 2012, New Delhi, Indien
- Michael Carl and Henrik Høeg Müller, Integrating Treebank Annotation and User Activity in TranslationResearch, In Christian Chiarcos, Sebastian Hellmann, and Sebastian Nordhoff (eds), Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata, pp. 77-85, Springer, 2012.
- Laura Winther Balling, Michael Carl, and Arnt Lykke Jakobsen (eds.) Proceedings of the International Workshop on Expertise in Translation and Post-editing - Research and Application, Copenhagen Business School, 2012 <http://www.cbs.dk/content/download/190452/2416914/file/2012-Proceedings.pdf>
- Christian Buck. Black Box Features for the WMT 2012 Quality Estimation Shared Task, Proceedings of the 7th Workshop on Statistical Machine Translation, pages 91-95, Montréal, Canada, June 7-8, 2012.
- Sittikorn Assavanives. Sentence Completion Prediction Based on Hierarchical Models, Honours project report, University of Edinburgh, 2012.

3 Coordination of Dissemination Activities (Task 7.4)

Dissemination activities were co-ordinated by UEDIN. The main goals for the first year was to create awareness for the project and the problems to be addressed by the started work.

The website maintained by the project (at <http://www.casmacat.eu/>) was set up to promote the public mailing list to harvest email addresses from interested targeted outsiders. While we have not yet broadly promoted the project (we will only have a publicly released prototype in Year 2), we have already gather 39 subscribers, and attracted interest from commercial contacts from companies such as Audi, Oracle, and Symantec.