D6.2: Analysis of the second field trial

Eva Marcos Iglesias, Massimiliano Pellegrino, Michael Carl,
Mercedes García Martínez, Bartolomé Mesa-Lao, Nancy Underwood

Distribution: Public

CasMaCat
Cognitive Analysis and Statistical Methods
for Advanced Computer Aided Translation

ICT Project 287576 Deliverable D6.2

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	D6.2
Deliverable title	Analysis of the second field trial
Type	Report
Status & version	Draft
Number of pages	13
Contributing WP(s)	WP7
WP / Task responsible	CS, CBS
Other contributors	
Internal reviewer	
Author(s)	Eva Marcos Iglesias, Massimiliano Pellegrino, Michael Carl, Mercedes García Martínez, Bartolomé Mesa-Lao, Nancy Underwoodl
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

In this work package, we evaluate the CASMACAT workbench in field trials to study the use of the workbench in a real-world environment. We will also integrate the workbench into community translation platforms and collect user activity data from both field trials and volunteer translators.

This Deliverable covers Tasks 6.1 and 6.2.

Task 6.1: Three field trials at a translation agency (Celer Soluciones SL)to evaluate the CASMACAT workbench in a real-world professional translation environment.

Task 6.2: Analysis of translator feedback and activity data. Collection of feedback of translators' self-estimation through retrospective interviews.

Contents

1   Introduction

Technology has played an increasingly important role within translation over the past six decades. Nowadays its impact is undisputedly extensive and has reached an unprecedented level that deserves careful consideration as a crucial factor which affects human translators. Post-editing machine translation (MT) output is gradually becoming a common practice within the localization industry as opposed to full human translation of new texts. Depending on the nature of the text, more and more language service providers (LSPs) pre-translate the source text using existing translation memories (TMs) 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 translators correct and adapt, i.e. post-edit, the output from both translation memories and machine translation to produce different levels of quality. Improving and maximising the potential of a post-editing workbench is thus one of the priorities set by both the industry and researchers when addressing the technological challenges faced by LSPs. The motivation behind this research comes from a desire to know how such tools can be of greater support to translation professionals, and how technology can even empower them to choose the translation methods, strategies and tools they feel comfortable with and which bring out the best of their skills. Hence the importance of developing a tool for better assisting human translators in the task of post-editing MT outputs.

2   Motivation and background

The motives for pursuing the development of a post-editing workbench can be diverse according to the interest of each stakeholder who is engaged in the language industry. Developers of commercial systems wish to improve their workbenches primarily in order to achieve business and financial gains from selling a high-quality product which helps to improve productivity. Researchers, on the other hand, may pursue the development and improvement of a post-editing workbench looking for better translator-support tools which empower translation professionals, regardless of their organisational environment. The CasMaCat workbench should help the translation community to overcome some of the challenges facing the profession in a more flexible way.

Since user satisfaction and translation productivity are at the core of the CasMaCat project, the aim of the field trials is to collect all the feedback possible from a sample of professional post-editors in order to implement their wish list in future prototypes of the system. Since a post-editing workbench is a tool in the hands of professionals rather than being simply a work of engineering, suggestions for improvement cannot be provided by any other means than the users themselves. The approach to be taken in order to improve the overall quality of an existing TM system is clear: optimising processes like segmentation, alignment and matching algorithms are at the top of the list. However, we assume that the needs and requirements of a post-editing workbench are of a different nature. Research that involves human evaluation of the implementation process (capable of establishing user-based evidence) is thus of paramount importance.

In general terms research on MT systems has traditionally focused on the tools themselves and, only recently, on productivity issues correlating with time and quality expectations. Not much research has been carried out yet on the use of a post-editing tool in a broader context.

The gap between social (user-focused) and scientific (technology-focused) research on MT systems is probably the main reason why the needs of users struggle to find their way into development proposals. The CasMaCat field trials attempt to bridge this gap by approaching the question of a post-editing workbench improvement from the social scientist's point of view and by offering an investigation of user needs which will supply the evidence needed by system developers before developing future versions of the CasMaCat workbench.

3   Aims

The overall aim of this research is to identify, analyse and consolidate the users' feedback on the second prototype of the CASMACAT workbench, which has incorporated a number of new features which were requested by the users in the first field trial. The following specific aims were also foreseen in this second field trial:

- To test interactive machine translation for post-editing purposes.

- To test different visualisations options of the text being post-edited.

- To collect quantitative and qualitative data on user satisfaction when working with the second prototype of the CASMACAT workbench.

Apart from the productivity and performance data collected and reported on deliverable D1.2, we were also very much interested in collecting users' feedback about these two specific aims.

4   The second field trial

A detailed description of the set-up and performance of the second CASMACAT field trial is provided in Deliverable 2.1. This current deliverable reports on the user feedback elicited from the second field trial in which prototype-II of the workbench was trialled in the real-world environment provided by Celer Soluciones. The user feedback from the first field trial was reported in deliverable D6.1 and in the paper presented at ASLIB 2012:

Mesa-Lao, Bartolomé. 2012. "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.

Based on the feedback in that first field trial, a number of extra functionalities requested by the post-editors were incorporated into prototype-II (see section 4.2).

4.1   Field trial overview

Figure 1 presents an overview of the second CASMACAT field trial for the part that was carried out at the Celer Soluciones SL offices. After completing an introductory questionnaire to collect metadata (age, gender, training, years of experience, expertise in post-editing, etc.), participants were introduced to the CASMACAT workbench featuring interactive machine translation (IMT). Post-editing guidelines were also provided as well as a detailed description of the characteristics of IMT which they would see implemented in either task 1, 2 or 3 (P, PI or PIA). The order of presentation of these three tasks was randomised in order to avoid any order effects in the results.

4.2   CasMaCat prototype-II - new functionalities

The CASMACAT GUI has changed since the first field trial and a number of new functionalities have been added. Figure 2 shows the GUI which the users were presented with during the second field trial.

A description of the prototype and its GUI options can be found in deliverable D1.2.

Of the additional functionalities requested by the participants in the first field trial (see Deliverable D6.1), the following were incorporated into the second CASMACAT prototype.
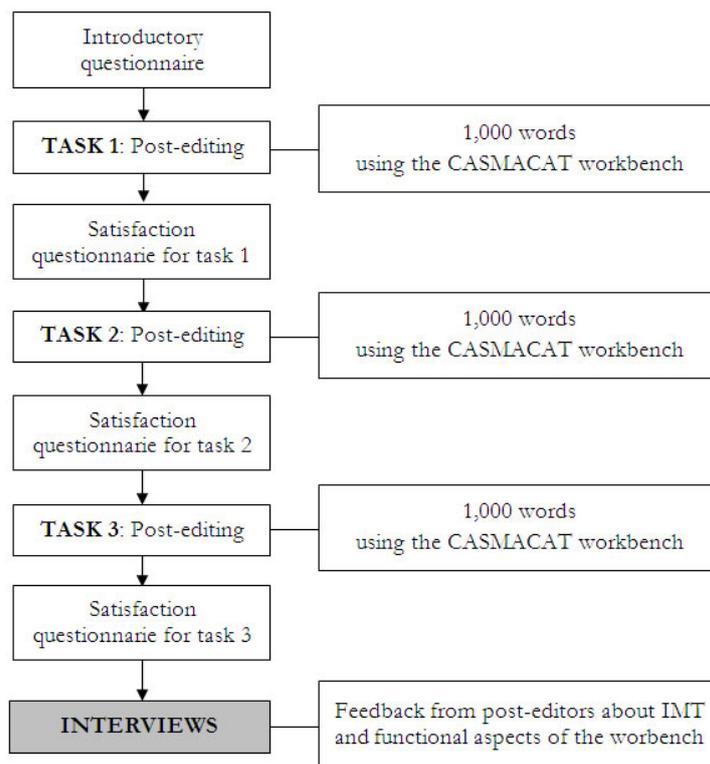
Figure 1: Overview of the second field trial - June 2013

- Visual tracking of changes (through interactive machine translation features)

- Monitoring the post-editing progress by real-time word counters and progress bars

- Systematic search and replace

- Copying text from source to target segments

- Autowrite functions (This autowrite function as described by the translators last year is what has been implemented as interactive machine translation, i.e. the system dynamically predicts alternative TM/MT translations) - This feature has been the independent variable of most of our analyses during this second field trial.

- Translation memory (TM) module as a result of the collaborative work done with the researchers of the MateCat project.

4.3   Post-editors' profiles

Of the nine participants in the second field trial only three had been involved in the first field trial. The nine interviewees in the evaluation of the second CASMACAT prototype had the following profiles as professional translators/post-editors working for Celer Soluciones SL (Madrid):

5   Eliciting user feedback

User feedback was elicited from the post-editors after they had worked with all three system configurations at the Celer Soluciones offices. As in the first field trial this took the form of
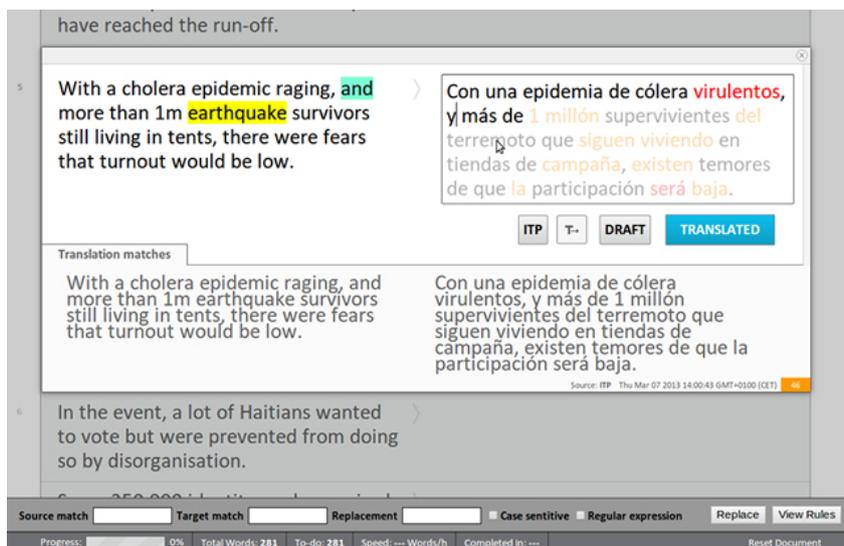
Figure 2: Graphical User Interface of the CasMaCat workbench - prototype-II

Participants	P01	P02	P03	P04	P05	P06	P07	P08	P09
Gender	F	M	M	F	F	M	F	F	M
Years of translator training	2	1	2	1	5	1	3	4	5
Years of professional experience	8	20+	15	2	5	20+	13	13	7
Previous experience in post-editing	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Participated the first CasMaCat field trial	Yes	No	Yes	No	Yes	No	No	No	No

Table 1: Profiles of the participants in the second CasMaCat field trial

individual semi-structured interviews with each post-editor. The interviewer spent time with the users of the CasMaCat workbench finding out about their personal opinions about the system. In the structured part of the interview, the interviewer used a standardised interview schedule with a pre-defined script (see Appendix 7.1). The questions tended to be asked in a similar order and format to facilitate comparisons between all answers. These questions were, however, intended to guide the discussion to any relevant information, rather than restricting it to specific types of information. There was also scope for pursuing and probing for novel, relevant information, through additional questions often noted as prompts on the schedule. The interviewer frequently had to formulate impromptu questions in order to follow up leads that emerged during the interview.

As the post-editors described their experiences with the CasMaCat workbench, the interviewer kept probing, searching for better and more complete descriptions of problems, comments and suggestions. Again, notes were taken of the key points made by the users in addition to the answers given to the questions. During the interview stage, the interview script was being revised continuously through the feedback loop created by the participants and the interviewer.

Feedback in term of user satisfaction for each of the three systems was also collected in the form of questionnaires after each of the post-editing session (see Figure 1).

6 Findings

6.1 Questionnaires on user satisfaction

After each session, the post-editors were asked to rate their satisfaction with their output and the CASMACAT tool on a 1-5 Likert scale, where 5 corresponded the highest positive reply and 1 the lowest. They were asked to respond to the following questions:

- How satisfied are you with the translations you have produced ? (*Satisfaction*)

- How would you rate the workbench you have just used in terms of usefulness/aids to perform a post-editing task? (*Tool*)

- Would you have preferred to work on your translation from scratch? (*From scratch*)

- Would you have preferred to work on the machine translation output without the interactivity provided by the system? (*No IMT*)

The post-editors responses for the three different system configurations are shown in the Tables 2, 3 and 4.

Participant	Satisfaction	Tool	From scratch
P01	3	3	No
P02	4	3	Yes
P03	3	3	Yes
P04	4	3	No
P05	4	4	No
P06	5	3	No
P07	3	2	Yes
P08	4	2	Yes
P09	4	1	Yes

Table 2: Satisfaction ratings for traditional post-editing (P)

Participant	Satisfaction	Tool	From scratch	No IMT
P01	4	4	No	No
P02	4	2	Yes	Yes
P03	3	3	No	No
P04	4	4	No	Yes
P05	3	4	No	No
P06	5	3	No	Yes
P07	4	1	Yes	Yes
P08	4	2	No	Yes
P09	4	4	Yes	Yes

Table 3: Satisfaction ratings for post-editing with interactivity (PI)

The results from the questionnaire show varying levels of satisfaction for the different systems. Some participants (i.e. P01, P03 and P07) appear to be more satisfied with the interactive systems (PI and PIA) than with traditional post-editing alone (P).

Participant	Satisfaction	Tool	From scratch	No IMT
P01	4	4	No	No
P02	4	4	Yes	No
P03	4	4	No	No
P04	5	4	No	No
P05	4	3	No	No
P06	5	2	No	Yes
P07	3	2	Yes	No
P08	3	3	No	Yes
P09	4	3	Yes	No

Table 4: Satisfaction ratings for post-editing with advanced interactivity (PIA)

6.2 Interviews

Due to the deliberately open format and the manner in which the questions were posed, the interviews in themselves provide information that is not always easy to assess. The interviewees discuss feelings rather than facts but in general they liked the tool regardless of the nature of the work. Interactivity has positively surprised them and almost all of them consider the visualisation aids to be extremely useful. They would also like to take part in next year's trials, despite having maintained a generally sceptical tone.

The need to expand the CASMACAT workbench functions was not so evident in the interviews in this year's trial. On some occasions, the subject only arose when the interviewer mentioned a potential extra function. The repeated correcting of errors was not mentioned by any of the participants interviewed in the interviews this year. However, interactivity caused several complaints among some post-editors who found it very frustrating whenever they introduced minor changes while finalising their work (i.e. correcting a typo) and the interactivity provided by the system changed their text again.

All the field trial participants are regular users of translation memories and once again indicated that they felt the lack of some functionalities from their own systems, such as concordances or the system learning as post-editing progresses.

The functions of the prototype tested in this second field trial and the best proposals made by the participants are discussed in further detail below.

6.2.1 Post-editing environment (the editor)

Apart from one of the people interviewed, who was more attached to older translation procedures, all of the post-editors taking part in the trials said that they worked comfortably with the current interface of the prototype. One of the interviewees mentioned that he particularly liked the fact that the original translation is displayed at the bottom of the text for reference purposes. The layout of the segments (horizontal, combined or vertical, one next to the other) did not seem overly important to the interviewees and they continue to like the appearance of the format in two columns. New versions of computer-assisted translation tools follow similar lines and most of the post-editors now work with this type of interface.

6.2.2 Post-editing process

The post-editing process in the second CASMACAT field trial focused on the use of interactive machine translation (IMT) for post-editing purposes. Three systems were used for post-editing:

- System 1 (P): Traditional post-editing.

- System 2 (PI): Post-editing that included interactive translation (changing the text to the right of the cursor depending on the corrections made by the post-editor).

- System 3 (PIA): Post-editing with interactive translation and visual aids (word alignments, suffix length prediction, etc.).

A detailed description of the systems involved in the second CASMACAT field trial can be found in deliverable D1.2.

The third option was most widely accepted by those interviewed and most confirmed having gained time in comparison with traditional post-editing thanks to the visualisation aids. Only in one of the interviews did the matter of segmentation arise, although the general conclusion (given that most of the interviewees work full-time in translation and, therefore, are more used to translation memories) is that, at present, there is no other way of organising texts (although the possibility of seeing the source and target documents in context is always a help). Some of the functions that the interviewees said they lacked during the trials are:

- Spellchecker: Apart from one of the post-editors, they all mentioned the importance of including a spellchecker in the tool.

- Comments: A space for notes, comments or some kind of mark so that the post-editor can return to it when reviewing.

- Automatic correction of obvious errors (typos): This is one of the functions that is included in many text editors and that two of the interviewees said would be extremely useful (e.g. in the case of common misspellings).

6.2.3 Productivity enhancement features

As well as the existing functions in the workbench, many mentioned other tools that could significantly enhance productivity:

- Search/replace tool: This would allow for global corrections to be made without having to go through the text segment by segment, which would create errors more likely. This function exists in the workbench now, however in the field trial itself when post-editing short news items, it was only used by one participant. For longer, more repetitive and technical texts it would be used much more often.

- Formatting options: Functions relating to formatting, such as bold or italics (in the case of words remaining in the original language, words to be highlighted, etc.). All texts have some kind of formatting and commercial CAT tools all have the means to preserve this format, either during the translation process itself or at the end, when the text is exported to Microsoft Word, Adobe Acrobat or another external program.

- Glossaries and dictionaries: In the interviews, the general and specific need arose in each document translation using the workbench to include query tools that learned from what was being post-edited. This is a request which was also voiced in the first field trial.

- Quality controls including the verification of numbers, parentheses, format, excessive differences in length between source and target segments, etc. Research is also required to determine which of these verification functions should automatically correct the errors detected and which should indicate any mistakes without correcting them. Along these lines, it would also be of interest to be able to mark the segments as imposed by the client, reviewed by a supervisor, etc.

- Automatic propagation of translations: This issue was not covered directly in the interviews, although it is an option that is present in many of the current CAT tools and is obviously useful in a post-editing system because it avoids the need to correct the same segments several times. An option that makes propagated translations overwrite already translated segments and even mark them as checked is required.

- Undo option on several levels and with predictable behaviours, such as in Microsoft Word.

In informal conversations with one of the reviewers, the possibility also arose of including a voice recognition system to select the option proposed by the system that seems most correct to the post-editor. This would greatly simplify the interface, as the post-editor would only have to read out the selected option from the list of proposals presented by the systems.

6.2.4 Non-functional aspects

One of the most common complaints among users of the prototype is the fact that the editor completely changes the segment when the post-editor modifies just one word. An attempt has apparently been made to correct this situation so that the segments which have been post-edited can then be reviewed and minor corrections made without the system again suggesting modifications and changing the whole segment. However, the result has not been entirely satisfactory. This could be connected to quality control so that, when a segment is considered verified and approved by a supervisor, the tool does not try to modify it.

The flexibility of the system's responses - quickly skipping from one segment to another, response to commands without delay is important and, in general, the level of satisfaction in this regard was high, so that the post-editing or reviewing work progresses smoothly and without frustration. The system must inspire trust that the work which has already been completed will not be lost as a result of inevitable hardware or software faults, the instability of the internet connection, etc. Although discussing the comfort of the final interface may seem a little premature at this stage, it should be considered because there are some potentially valuable functional aspects that could go unnoticed or become useless if they are not duly included in the interface.

It is clear in this sense that the post-editor must be able to adapt the colours of the interface, its fonts and font size, etc. to suit his/her taste. The segment being worked on should be in the centre of the screen so that both the previous segment and the next segment can be seen. Not all projects or all translators require the same functions and, therefore, it should be possible to enable and disable these to ensure they never hinder the post-editing work. For the purposes of the second field trial the participants were not allowed to switch off interaction or any of the visualisation aids which they had chosen during the trial. However it is possible to switch these options on and off. In the next trial we intend to investigate the effects of this possibility on post-editors' performance and acceptance of the tool.

Nowadays almost everyone, translators or otherwise, must use a considerable number of computer systems and we only reject those with unintuitive interfaces or with interfaces that are different from the norm. The tool should, where possible, respect the generally-accepted design rules for interfaces, particularly translation tool interfaces, as the freelance translator is often forced to work with several of them.

Many of the functions of the tool will depend on the browser and, given that this is subject to continuous and sometimes automatic updates, such updates must not affect functionality. One aspect that was tangentially covered is the internal logic of the tool. The layout in segments made the translators feel that the machine was working like another translation memory rather than an interactive automatic translation system. Furthermore, several indicated that

they lacked a concordance function, which suggests that they thought that CASMACAT was a translation memory. We believe it is important for the translator to have an idea of how the tool works and to be able to interpret and understand the way in which it responds to each situation. One matter that was not discussed in the interviews is the capacity of the tool to run on laptops, which is extremely important in the opinion of the post-editors. This capacity does not only depend on the power of the machine but also, and above all, on the interface. In fact, the keyboard of a laptop rarely includes a separate numeric keypad, for example, which could make translation and post-editing work much more frustrating if this was not considered when formulating the key combinations enabling the different functions.

In conclusion, in order to be useful and acceptable, a translation assistance tool must be just that: assistance, and not an annoying, difficult to control obstacle. We believe that the second CASMACAT prototype has met with expectations and it is especially worth noting the optimistic acceptance of translators with regard to the predictive system (especially with visualisation aids), although some repeat problems in operations did arise. The good design of the interface is probably as important as (if not more important than) the internal technology of the tool. Ergonomics, solidity, stability, response flexibility and controllability are essential qualities.

7 Appendix

7.1 Interview script

Date: June 10-19, 2013
Venue: Celer Soluciones SL (Madrid, Spain)
Time assigned: 30 minutes for each of the nine post-editors
Language: Spanish (mother tongue of all the post-editors)

[STRUCTURED PART OF THE INTERVIEW]

- Welcome and introduction.

- Signature of informed consent in order to be able to record the interview.

- General comments on the second field trial performed with the CASMACAT workbench prior to the interviews.

- Interactive machine translation

- CASMACAT workbench: Prototype II (second year of the project)

  - Functional aspects:
    * The (new) GUI (*for post-editors who already saw Prototype I*)
    * Workflow functionalities.
    * Comments on interactive machine translation (IMT):
      · Productivity as perceived by the post-editor: an aids or a hindrance?
      · User satisfaction feedback.
      · Room for a different view after more hours of interaction?
      · Suggestions about new ways of implementing IMT?
    * Comments on the desired functionalities to be implemented in future versions: *Departing from previous experiences in any TM system and any other post-editing workbench you may know, which specific functions would you like to see implemented in a post-editing tool?*

- Non-functional aspects:
    * Report on the usability, customisability, learn ability and supportability of the GUI.

[UNSTRUCTURED PART OF THE INTERVIEW]

- Room for open comments and feedback from the participants regarding any post-editing issues that may arouse.