# ANALYSIS OF THE SECOND FIELD TRIAL

Eva Marcos and Massimiliano Pellegrino, Celer Soluciones

November 25th, 2013

# COMPANY BACKGROUND

- The company has been using CAT tools and MT for more than a decade
- We mainly translate in the Life Sciences, Institutional and Technical fields.
- Our most common language pairs are English-Spanish, Spanish-English.
- Participation in many R&D European and Spanish projects (Transtype 2, SMART, CASMACAT, EXPERT) has been a key factor

# THE EXPERIENCE OF CELER SOLUCIONES

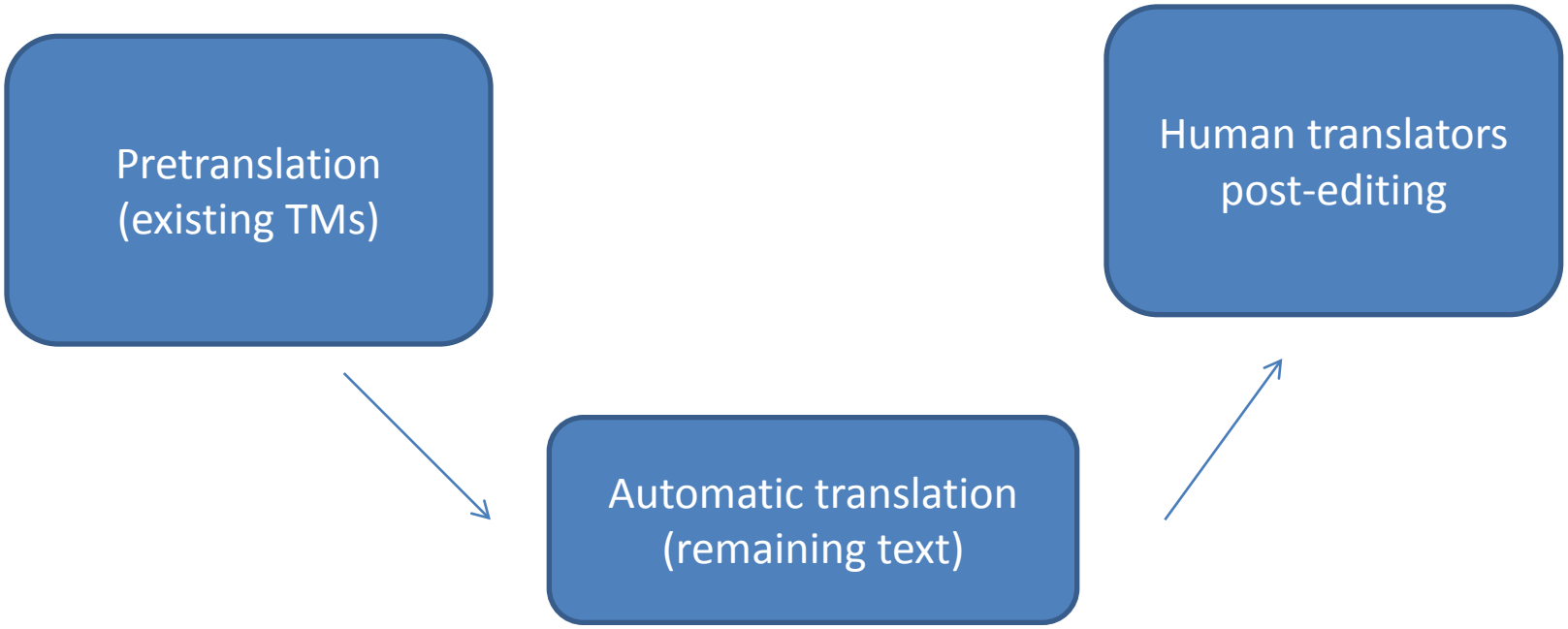
- Celer involved in working with MT and post-editing tools for a long time

**Some figures (post-edited words after MT)**

- 2010: 6,295,419 words
- 2011: 6,814,115 words
- 2012: 7,463,106 words

## MT AND POST-EDITING – OVERVIEW

- Gradually becoming a common practice within the localization industry as opposed to full human translation of new texts.
- Usual procedure:



# MOTIVATION

- Why pursue the development of a post-editing workbench?
  1- DEVELOPERS OF COMMERCIAL SYSTEMS: achieve business and financial gains
  2- RESEARCHERS: looking for a better translator-support tool to empower translation professionals.

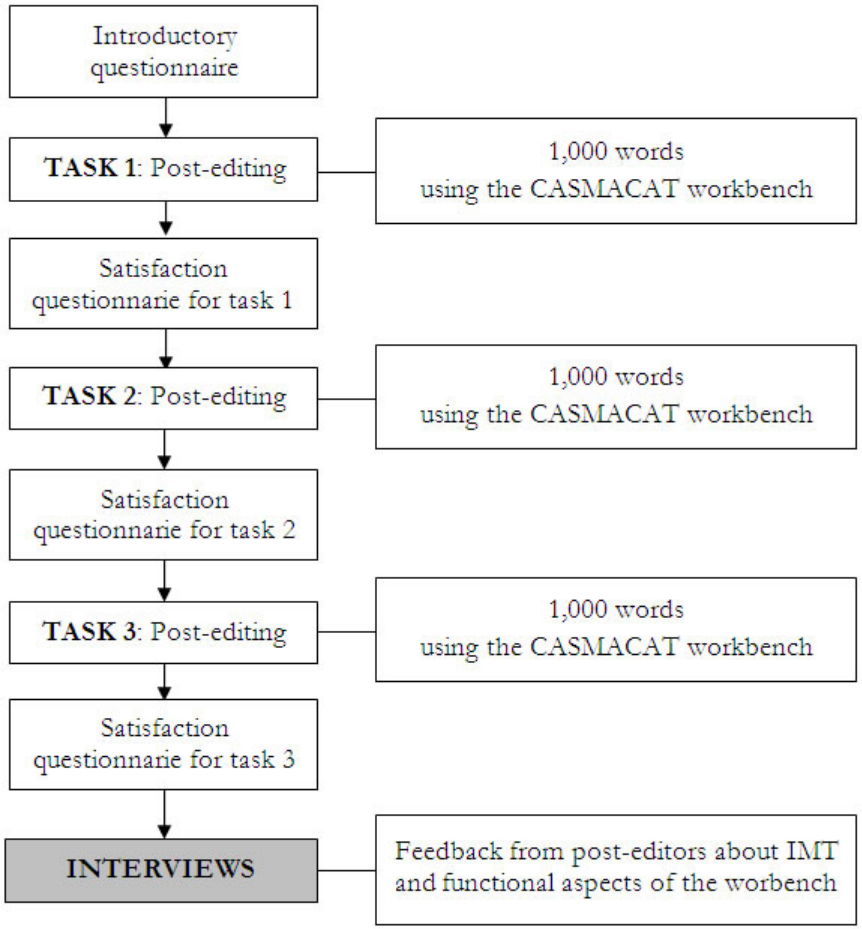
**CASMACAT PROJECT → USER SATISFACTION AND TRANSLATION PRODUCTIVITY**

# AIMS OF FIELD TRIAL

- Collect all feedback possible from professional post-editors in a real world environment to implement their wish list in future prototypes (Improve overall quality of TM system through optimising processes like segmentation, alignment and matching . Human evaluation is essential to get to know the requirements of post-editing workbench)
- Bridge the gap between social and scientific research on MT systems
- Identify, analyse and consolidate the users' feedback on the second prototype.
- Test interactive MT for post-editing purposes.
- Test different visualisation options of text being post-edited.

# OVERVIEW OF THE SECOND CASMACAT FIELD TRIAL

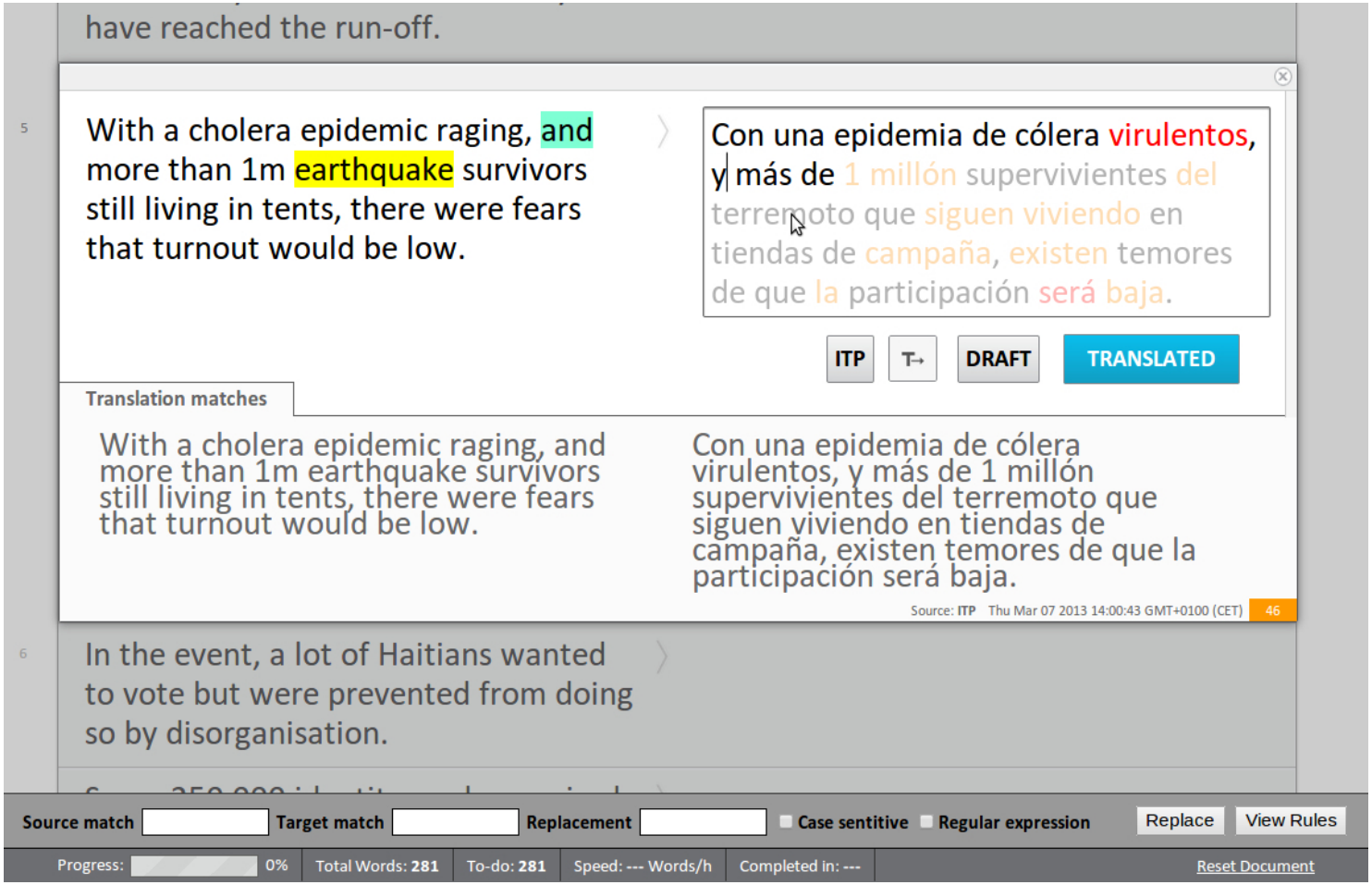
In Madrid - offices of Celer Soluciones - June 2013



# SYSTEM AND TASKS

- MT system used based on Moses (UEDIN WMT 2013 evaluation campaign). Best constraint system at the evaluation campaign. Tied for second place (with two others) overall.
- TASKS→ Translation of new stories (EN-ES) collected from CNN, Fox News, NY Times…
- System was trained
  - 4.5 Million word News Commentary
  - 57 million word Europarl
  - 319 million word United Nations parallell Corpora
  - 45 million word CommonCrawl parallel corpus
  - Additional 386 million words of monolingual news language model data on 1,062 million words from the Spanish Gigaword corpus.
- It achieved a BLEU score of 34.8 (case –sensitive) on the 2012 test set and 30.4 on 2013 test set
- Analysis by professional translator at Celer Soluciones = mostly useful for post-editing

# NEW GRAPHIC USER INTERFACE



# FUNCTIONALITIES ADDED

Functionalities requested by participants in the first Field Trial incorporated to version 2.0 of prototype:

- Visual track of changes
- Real-time PE progress
- Search and replace
- Copy-paste source to target
- Autowrite functions
- Translation Memory module

# POST-EDITORS' PROFILE

| ***Participants*** | ***P1*** | ***P2*** | ***P3*** | ***P4*** | ***P5*** | ***P6*** | ***P7*** | ***P8*** | ***P9*** |
|---|---|---|---|---|---|---|---|---|---|
| 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 as translators | *8* | *+20* | *15* | *2* | *5* | *+20* | *13* | *13* | *7* |
| Previous experience in post-editing | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Participated in last year's field trial | *Yes* | *No* | *Yes* | *No* | *Yes* | *No* | *No* | *No* | *No* |

Table 1. *Profile of the post-editors interviewed*

# OBTAINING USER FEEDBACK

- Obtained after post-editors had worked with all three GUI configurations

- Semi-structured interviews :

**UNSTRUCTURED PART OF THE INTERVIEW**

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

- Interviewer kept searching for better descriptions of problems + comments and suggestions

# OBTAINING USER FEEDBACK

- 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 seconf fieldf trial performed with the CASMACAT workbench prior to the interviews
- Interactive machine translation
- CASMACAT workbench: Protoype 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: and 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.

# USER SATISFACTION

- Rated after each session on a 1-5 Likert scale (5 highest positive reply, 1 lowest).
- 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)

# USER SATISFACTION (traditional post-editing)

| 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)

# USER SATISFACTION (post-editing with interactivity)

| 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)

# USER SATISFACTION (post-editing with advanced interactivity)

| 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)

# FINDINGS

- In general they liked the tool
- INTERACTIVITY – VERY POSITIVE (Visualisation Aids).
- All of them would like to participate in future Field Trials
- Need to expand functionalities not so evident this year (some of them very much related to TMs: *learning capacity, concordance….*)
- Change of text by system based on the corrections made by the post-editor – very frustrating
- Post-editing environment good, user-friendly interface (layout of segments not very important).

# Process

- Focused on interactivity (with and without visualisation aids) and divided in three blocks:

  1. Traditional post-editing
  2. Post-editing with interactive translation
  3. **Post-editing with interactive translation and visual aids**

- Third option to be the best accepted, they stated they gained time compared to traditional post-editing thanks to visualisation aids.

# Functions missed by participants

- Spell checker
- Comments: space for notes, marks, highlights…
- Automatic correction of obvious errors (typos)

# Productivity enhancement features

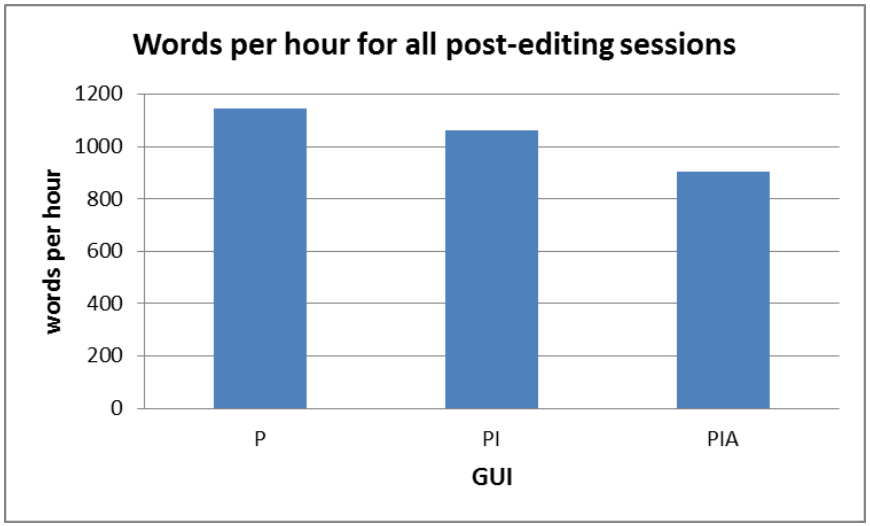
- Formatting options
- Glossaries / Dictionaries
- Quality controls
- Automatic propagation of translations
- Undo option
- Voice recognition system

# Non-functional aspects

- Most common complaint: system completely changing the segment even if only one word has been changed by post-editor (could be linked to quality control).
- Flexibility of system responses very important and positive
- The post-editor should be able to adapt colours, fonts, font sizes of interface, situation of segment working on
- Post-editors COULD NOT switch off interaction during field trial but it is possible and its effects will be investigated in next field trial.
- The prototype should respect the generally accepted design rules for translation tool interfaces and try to make its functions non-dependant on browser updates.

# Productivity

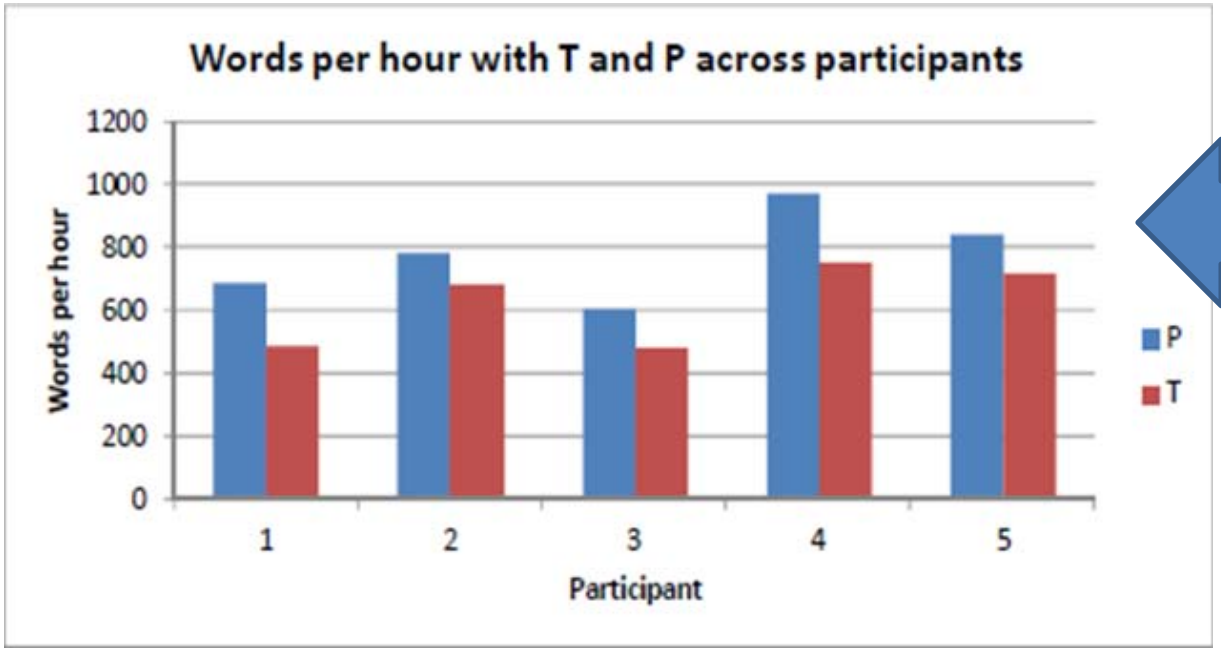
- Detailed on WP 1 presentation




In terms of average number of words per hour calculated on the time taken but excluding pauses of over 200 seconds

- This graph seems to indicate PI and PIA interfaces reduce productivity BUT there is a learning effect to be shown on next presentation

# Productivity (compared to last year's figures)




First Field Trial

*This comparison shows that the overall figures have increased from the range of 609-975 words per hour (2012, first prototype) to 904-1114 words per hour (2013, second prototype)*

# FINAL CONCLUSIONS

- ***A translator assistance system must be helpful and not annoying or difficult to control (according to the feedback elicited, the second prototype meets expectations).***
- ***Optimistic acceptance of translators of predictive system***
- ***Good design of interface***



# Thank you!