



Interactive-predictive machine translation: towards the next generation of workbenches for translators, the CasMaCat project

Germán Sanchis-Trilles, Francisco Casacuberta
Pattern Recognition and Human Language Technology Group
Universitat Politècnica de València

Séminaire DGA sur le Traitement de l'Information Multimédia
DGA workshop on Multimedia Information Processing
Télécom ParisTech
1-3 juillet 2013



Interactive-predictive machine translation

June 26, 2013

Goal of the CasMaCat project

The **CASMACAT** project will build the next generation translator's workbench to improve productivity, quality, and work practises in the translation industry.

Novel types of assistance to human translators:

- **Interactive translation prediction** vs. traditional post-editing
- **Interactive editing** vs. plain text editing
- **Adaptive translation models** vs. "static" pre-trained translation models



The project

- Co-funded by the European Union under the Seventh Framework Programme Project 287576 (ICT-2011.4.2)
- EU contribution: 2.5 MEuros
- 327 person-months
- 8 work-packages
- November 2011 - October 2014

Partners



University of Edinburgh (UEDIN), United Kingdom.
Leader: Philipp Koehn (Main coordinator).



Universitat Politècnica de València (UPVLC), Spain.
Leader: Francisco Casacuberta.

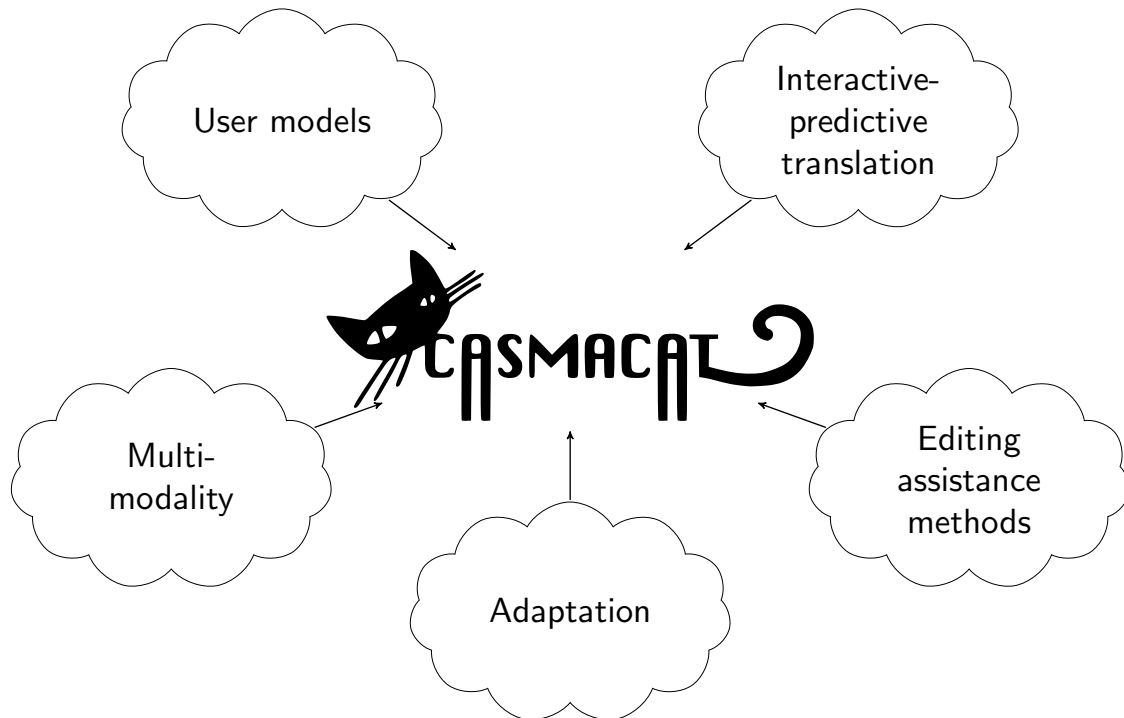


Copenhagen Business School (CBS), Denmark.
Leader: Michael Carl.

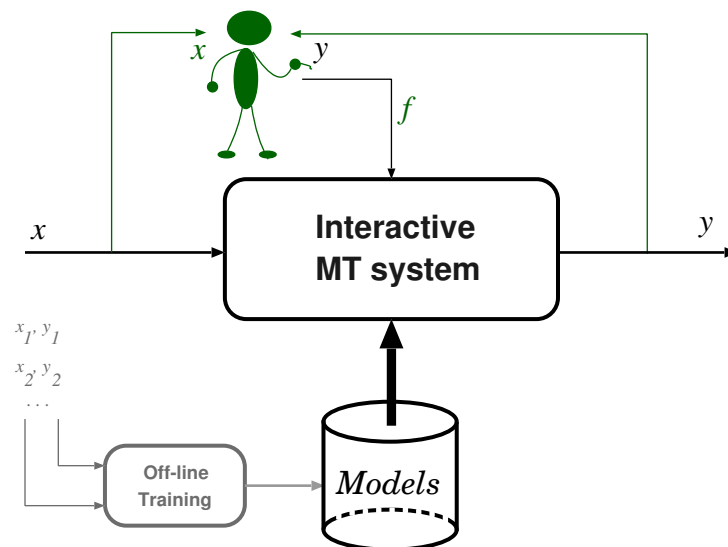


Celer Soluciones (CL), Spain.
Leader: Eva Marcos.

Main tasks



Interactive-predictive translation



[Foster et al., MT 1997][Barrachina et al., CL 2008][Casacuberta et al., CACM 2008].



Interactive-predictive translation: an example

Translating the source sentence (x) "*Click OK to close the print dialogue*" into Spanish:

System: Haga clic para cerrar el diálogo de impresión



Interactive-predictive translation: an example

Translating the source sentence (x) "*Click OK to close the print dialogue*" into Spanish:

System: Haga clic para cerrar el diálogo de impresión

User: Haga clic en





Interactive-predictive translation: an example

Translating the source sentence (x) "*Click OK to close the print dialogue*" into Spanish:

System: Haga clic para cerrar el diálogo de impresión

User: Haga clic **en**

System: **ACEPTAR** para cerrar el diálogo de impresión



Interactive-predictive translation: an example

Translating the source sentence (x) "*Click OK to close the print dialogue*" into Spanish:

System: Haga clic para cerrar el diálogo de impresión

User: Haga clic **en**

System: **ACEPTAR** para cerrar el diálogo de impresión

User: Haga clic en **ACEPTAR** para cerrar el **cuadro**





Interactive-predictive translation: an example

Translating the source sentence (x) *"Click OK to close the print dialogue"* into Spanish:

System: Haga clic para cerrar el diálogo de impresión

User: Haga clic **en**

System: ACEPTAR para cerrar el diálogo de impresión

User: Haga clic en ACEPTAR para cerrar el **cuadro**

System: de diálogo de impresión



Interactive-predictive translation: an example

Translating the source sentence (x) *"Click OK to close the print dialogue"* into Spanish:

System: Haga clic para cerrar el diálogo de impresión

User: Haga clic **en**

System: ACEPTAR para cerrar el diálogo de impresión

User: Haga clic en ACEPTAR para cerrar el **cuadro**

System: de diálogo de impresión

User: Haga clic en ACEPTAR para cerrar el cuadro de diálogo de impresión



Interactive-predictive translation: an example

Translating the source sentence (x) “Click OK to close the print dialogue” into Spanish:

System: Haga clic para cerrar el diálogo de impresión

User: Haga clic **en**

System: ACEPTAR para cerrar el diálogo de impresión

User: Haga clic en ACEPTAR para cerrar el **cuadro**

System: de diálogo de impresión

User: Haga clic en ACEPTAR para cerrar el cuadro de diálogo de impresión

Result (y): Haga clic **en** ACEPTAR para cerrar el **cuadro** de diálogo de impresión

TOTAL: Two word-strokes



Interactive-predictive translation

- Given a source text x and a “correct” *prefix* y_p of the target text, search for a *suffix* \hat{y}_s , that maximises the posterior probability over all possible suffixes of a given length:

$$\hat{y}_s = \operatorname{argmax}_{y_s} \Pr(y_s \mid x, y_p) \equiv \operatorname{argmax}_{y_s} \Pr(y_p y_s \mid x)$$

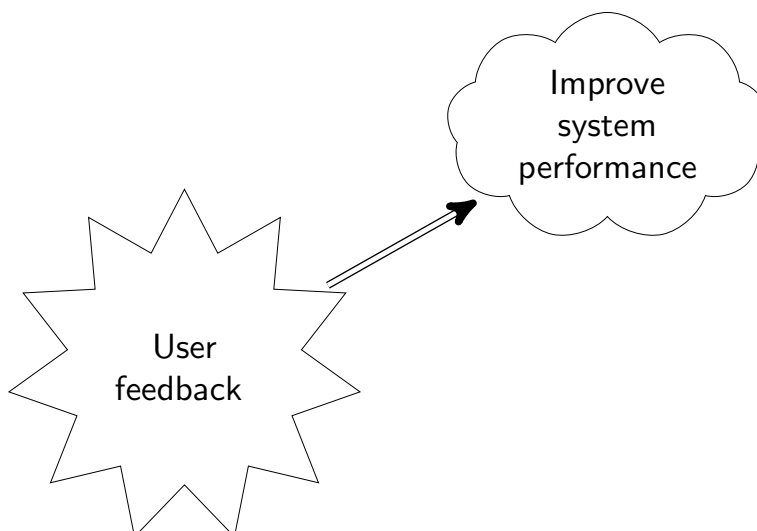
- Main difference between IPMT & MT: **search over the set of suffixes**
- **~20%** productivity increase with IPMT vs PE in field trials
- Ongoing field trials to evaluate IPMT vs. PE



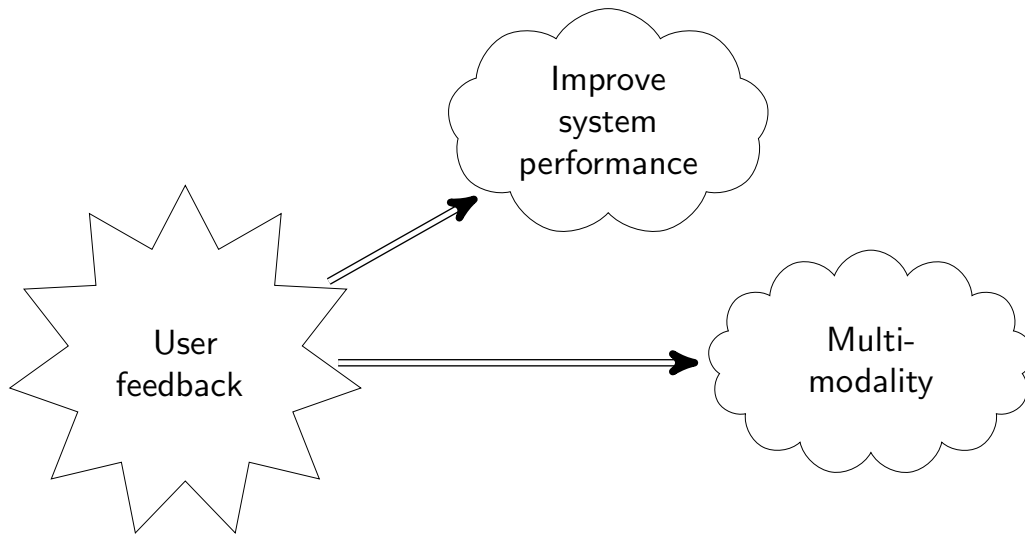
Interactive-predictive translation: Challenges & opportunities



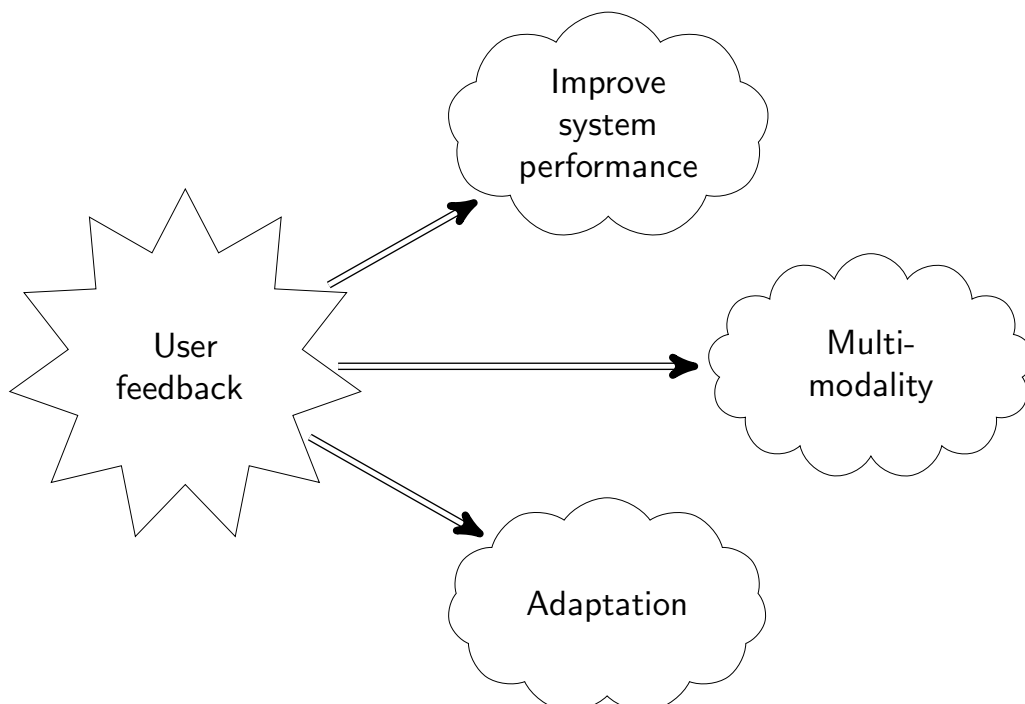
Interactive-predictive translation: Challenges & opportunities



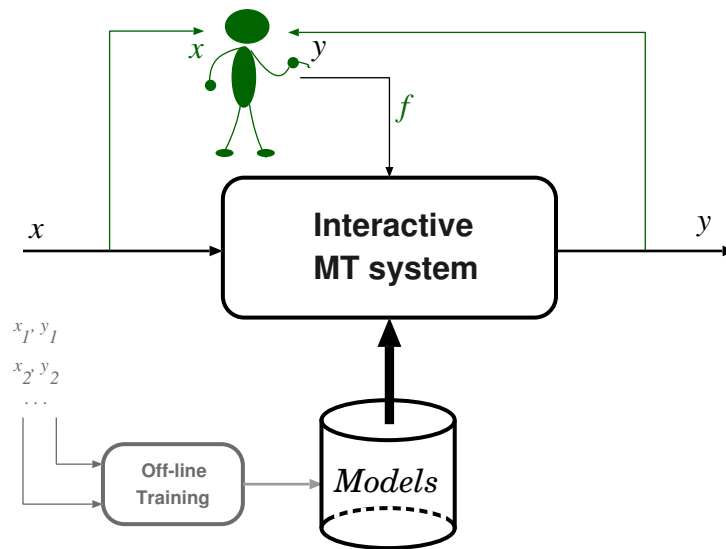
Interactive-predictive translation: Challenges & opportunities



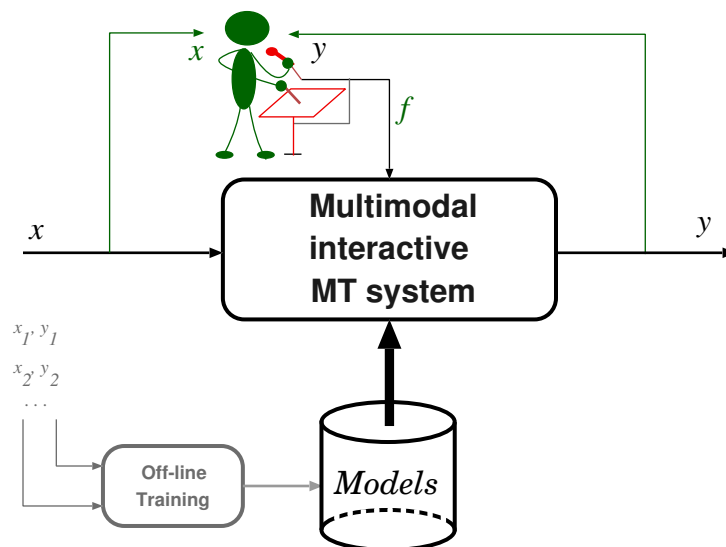
Interactive-predictive translation: Challenges & opportunities



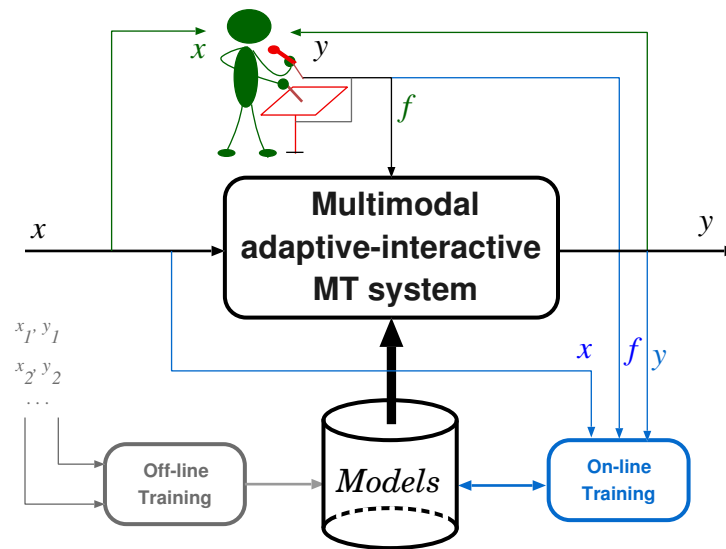
Additional features (I)



Additional features (I)



Additional features (I)



Additional features (II)

- On-line learning
 - Learn from user feedback
 - Models updated incrementally in real time
 - Effort reductions of up to 20%



- Domain and user adaptation
 - Most important: system stabilisation
 - Around 2% precision gain w.r.t. non-adapted baseline
 - Improvements with few data w.r.t. re-estimation
 - Still on-going work for adapting the methods to IPMT

Additional features (III)



- Cognitive studies and user modelling
 - Questionnaires and interviews
 - Use of an eye-tracker
- New translation assistance methods
 - Highlight possible incorrect words
 - Review only certain sentences:
+33% efficiency

- Multi-modality
 - Virtually error free set of gestures
 - Production-ready performance with HTR



11

Features to be shown in the demo

- Word alignments
- Word-level confidence measures
- Validated words
- Suffix length limitation
- IPMT vs PE



12

[Live demo](#)

[Demo video](#)

The UPVLC team

Francisco Casacuberta (team leader), Vicent Alabau
José-Miguel Benedí, Jorge González, Jesús González-Rubio
Alfons Juan, Luis A. Leiva, Daniel Martín-Albo
Daniel Ortiz-Martínez, Alberto Sanchis
Germán Sanchis-Trilles and Enrique Vidal

Thank you!

References

- [Foster et al.(1997)] George Foster, Pierre Isabelle, and Pierre Plamondon. Target-text mediated interactive machine translation. *Machine Translation*, 12(1–2):175–194, 1997.
- [Barrachina et al.(2009)] Sergio Barrachina, Oliver Bender, Francisco Casacuberta, Jorge Civera, Elsa Cubel, Shahram Khadivi, Antonio Lagarda, Hermann Ney, Jesús Tomás, and Enrique Vidal. Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1):3–28, 2009.
- [Casacuberta et al.(2009)] Francisco Casacuberta, Jorge Civera, Elsa Cubel, Antonio Lagarda, Guy Lapalme, Elliott Macklovitch, Enrique Vidal: Human interaction for high quality machine translation. *Communications of the ACM*, 52(10):135–138, 2009.
- [Ortiz-Martínez et al.(2010)] Daniel Ortiz-Martínez, Ismael García-Varea, and Francisco Casacuberta. Online learning for interactive statistical machine translation. In *Proceedings of Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 546–554, June 2–4 2010.
- [Alabau et al.(2013)] Vicent Alabau, Alberto Sanchis, and Francisco Casacuberta. Improving On-line Handwritten Recognition in Interactive Machine Translation. To be published in *Pattern Recognition*. 2013.

Features of the workbench (I)

Keyboard and mouse based interaction

The screenshot displays a translation workbench interface. At the top, there is a source text: "With a cholera epidemic raging, and more than 1m earthquake survivors still living in tents, there were fears that turnout would be low." This is translated into Spanish: "Con una epidemia de cólera virulentos, y más de 1 millón supervivientes del terremoto que siguen viviendo en tiendas de campaña, existen temores de que la participación será baja." The interface includes a 'TRANSLATED' button and a 'DRAFT' button. Below the main text, there is a 'Translation matches' section showing the source and target text side-by-side. At the bottom, there is a search bar with fields for 'Source match', 'Target match', and 'Replacement', along with checkboxes for 'Case sensitive' and 'Regular expression', and a 'Replace' button. The status bar at the very bottom shows 'Progress: 100%', 'Total Words: 121', 'To-do: 121', 'Speed: 1 Word/s', and 'Completed in: 1m 10s 100ms'.

Features of the workbench (II)

E-pen based interaction

The screenshot shows the workbench interface with a focus on E-pen based interaction. The source text is "The metropolis has had a permanent anti-corruption unit since 1870." and the target text is "La metrópoli ha tenido una unidad anticorrupción permanente desde 1870." A 'visualization' window is open, showing the source text with a red dashed box around the phrase "No obstante, el martes anterior sorprendió a más de una persona." Below this, there are buttons for 'ITP', 'T-', 'DRAFT', and 'TRANSLATED'. The 'TRANSLATED' button is highlighted in blue. At the bottom, there is a search bar with fields for 'Source match', 'Target match', and 'Replacement', along with checkboxes for 'Case sensitive' and 'Regular expression', and a 'Replace' button. The status bar at the bottom shows 'Progress: 34%', 'Total Words: 314', 'To-do: 207', 'Speed: 1 Word/s', and 'Completed in: 1m 10s 100ms'.

Features of the workbench (III)

Constrained length prediction / Confidence measures / Word alignment

Lisbon and Madrid wish to embark on a path different from that taken by Greece and Ireland.

Lisboa y Madrid quieren emprender un camino diferente del adoptado por Grecia e Irlanda.

ITP T- DRAFT TRANSLATED

Lisbon and Madrid wish to embark on a path different from that taken by Greece and Ireland.

Lisboa y Madrid **desea embarcarse en un camino diferente del adoptado por Grecia e Irlanda.**

ITP T- DRAFT TRANSLATED

Lisbon and Madrid wish to embark on a path different from that taken by Greece and Ireland.

Lisboa y Madrid **desea embarcarse en un camino diferente del adoptado por Grecia e Irlanda.**

ITP T- DRAFT TRANSLATED