



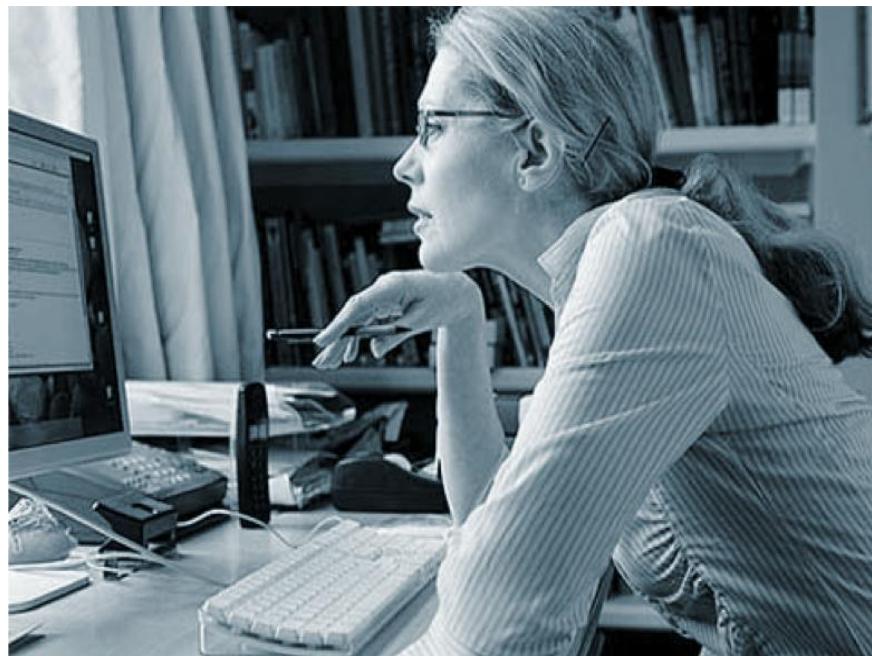
FONDAZIONE  
BRUNO KESSLER

# Machine Translation and Post-Editing

Marcello Federico

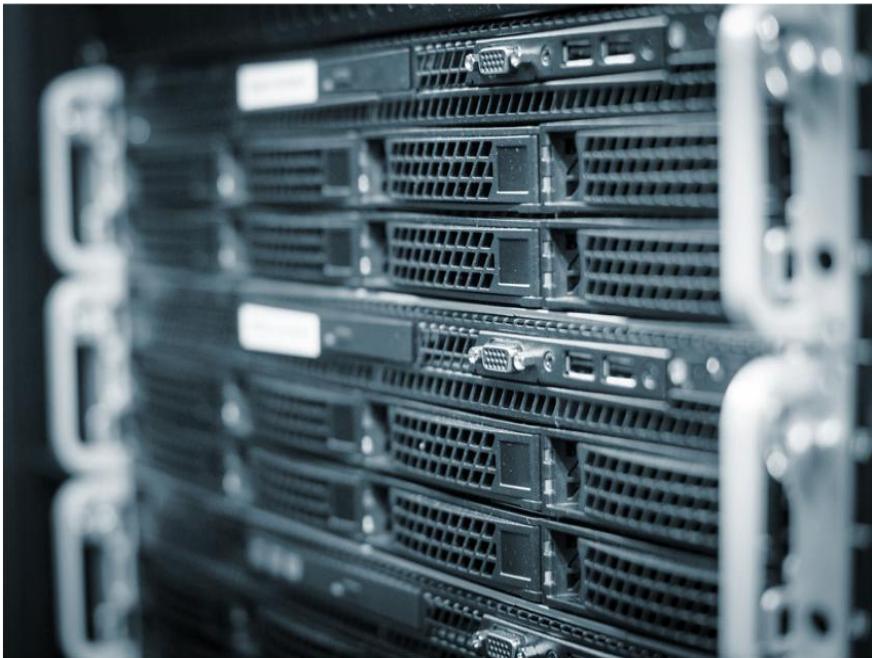
Workshop ELRC @ Italia, Rome, 15 March 2015

# human translation



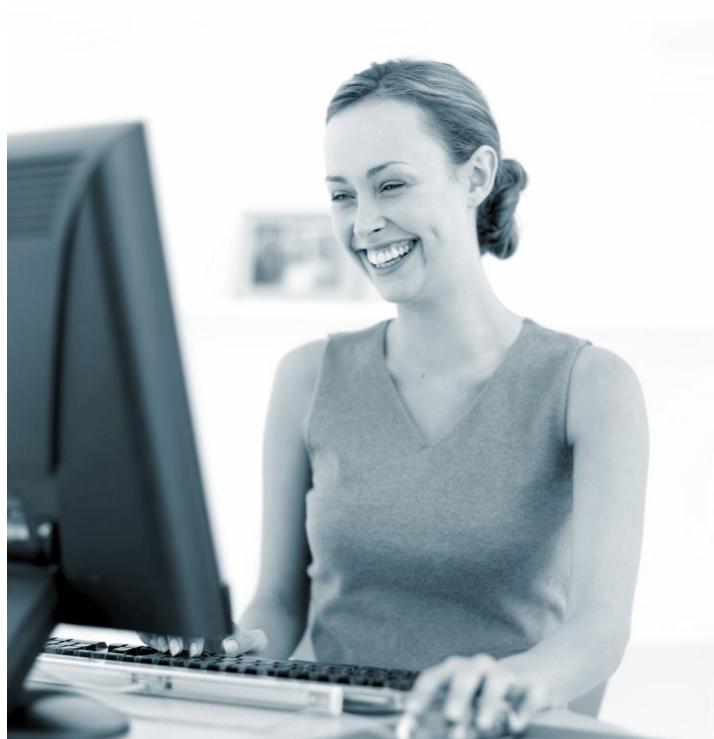
- accurate
- time consuming
- demanding

# machine translation



- less accurate
- quick
- inexpensive

# machine and human symbiosis



- better user experience
- MT learns from user
- enhanced productivity

# computer assisted translation

The screenshot shows a web browser window for 'Translate - PROJ-359BF7E47E19CC9E903E2484864C5926 - 352'. The URL is <http://demo.matecat.com/translate/PROJ-359BF7E47E19CC9E903E2484864C5926/en-it/361-sldfjw322d>. The page title is 'Translate - PROJ-359BF7E47E19CC9E903E2484864C5926 - 352'. The browser toolbar includes Reader mode and Google search.

The main content area displays a document with two columns: English on the left and Italian on the right. The English text is: "A CAT Tool for Your Business. Simple. Web-Based". The Italian text is: "Un CAT Tool per il tuo lavoro. Semplice. Web-based.". A progress bar at the top indicates 100% completion.

Three speech bubbles highlight specific features:

- A black speech bubble points to the Italian text: "Post-editing".
- A black speech bubble points to the Italian text: "Suggestion from TM".
- A black speech bubble points to the Italian text: "Suggestion from SMT".

Below the main content, there are two sections labeled 197959 and 197960, each containing text about MateCat's integration of Statistical Machine Translation (SMT) and Collaborative Translation.

At the bottom, a progress bar shows 31% completion, along with status metrics: Progress: 31%, Eq. words: --, Translated: --%, Approved: --%, Words last hour: --, Expected completion: -- Hours, and Project Statistics.

# translation memory

- Incrementally stores translated segments. Given a new source segment it looks for **perfect or fuzzy matches**
- Matches are ranked (100%-matches on top) and presented to the user as translation **candidates for post-editing**
- A TM can be shared among and simultaneously updated by several translators working on the same project
- TMs model the style and terminology of the customers

# translation memory

Segment to be translated:

*Pull the lever to engage Dark Forces*

TM fuzzy match: 62.5% [5 words out of 8]

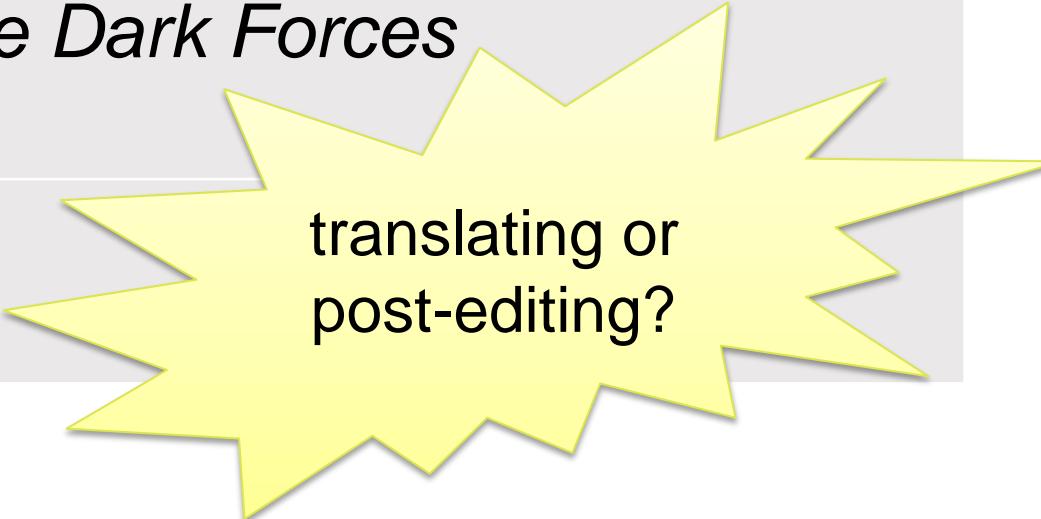
*Pull the lever to engage the drive motors.*

*Tirare la leva per attivare i motori di trazione.*

# machine translation

Segment to be translated:

*Pull the lever to engage Dark Forces*



A large, yellow, five-pointed starburst graphic with a black outline, positioned to the right of the English sentence. It contains the text "translating or post-editing?" in a black sans-serif font.

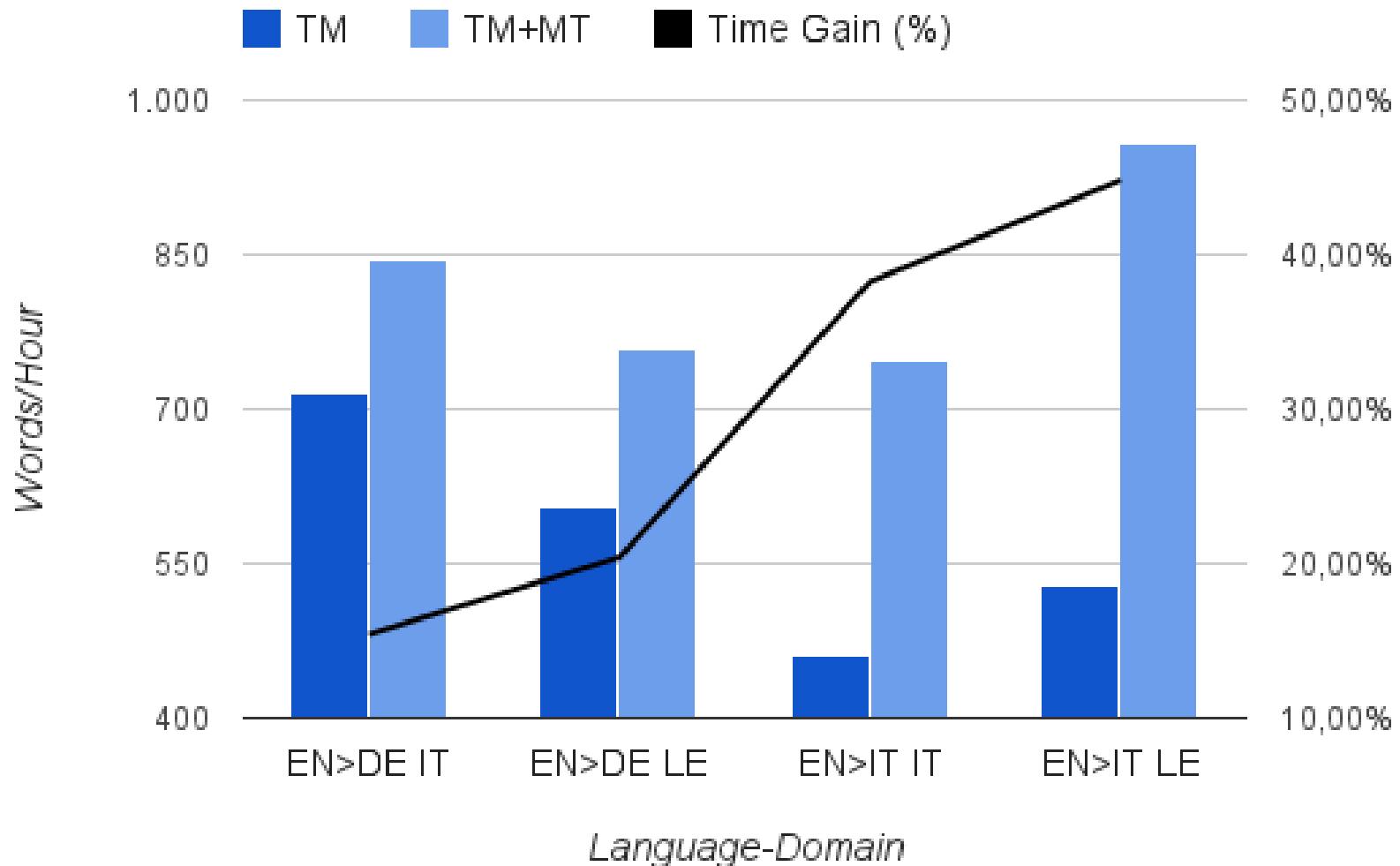
translating or  
post-editing?

Google Translate hint:

*Pull the lever to engage the drive motors.*

*Tirare la leva di impegnarsi Dark Forces*

# impact of MT post-editing



M. Federico, A. Cattelan, M. Trombetti. "Measuring user productivity in machine translation enhanced computer assisted translation". *Proc. of AMTA*, San Diego, CA, October 28 – November 1, 2012.

a crash course on  
machine translation

# machine translation

... is like deciphering a secret code



"Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography - methods which I believe succeed even when one does not know what language has been coded - one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Warren Weaver (1894-1978)

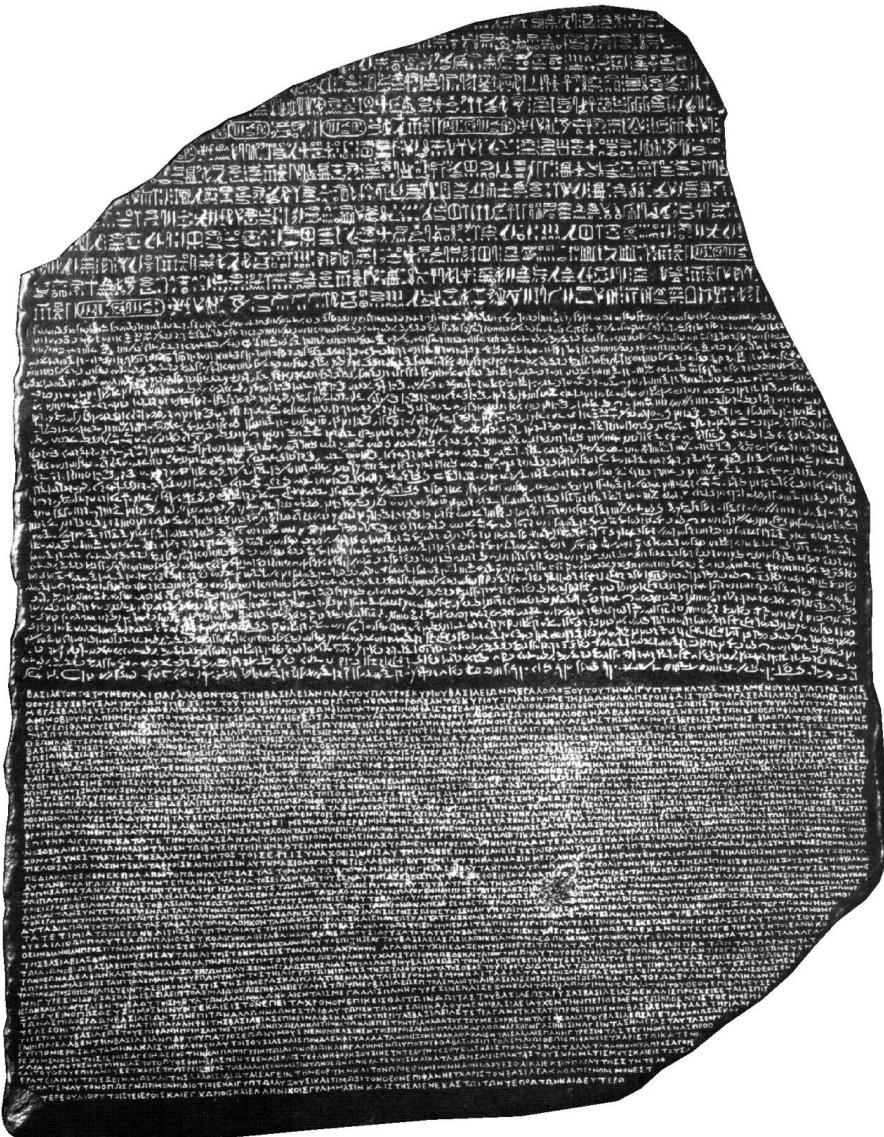
The computer learns how to translate by analyzing many human translations

# translation and archaeology

*With machine learning  
algorithms a computer  
can learn to translate*

...

*... by analysing a  
large sample of  
translation examples.*



# statistical machine translation

- Training Phase
  - sentence alignment
  - word alignment
  - phrase extraction
  - ....
- Translation Phase
  - choice of phrases and their position

# statistical machine translation

- **Training Phase**
  - sentence alignment
  - word alignment
  - extraction of phrases
  - ....
- Translation Phase
  - choice of phrases and their position

# sentence alignment

## EXECUTIVE SUMMARY

### The Bologna Process and its objectives for 2020

The Bologna Process has transformed the face of European higher education. Indeed all countries have made significant changes that have enabled the European Higher Education Area to emerge, and which have laid the ground for higher education that is serving an increasing range of societal demands; higher education structures have been modified, quality assurance systems developed, mechanisms to facilitate mobility established, and a range of issues for the social dimension of higher education identified. The scale of a project that, on the basis of voluntary cooperation, agrees and implements common objectives for the higher education systems of 47 countries is unprecedented.

However, conscious of the fact that the second decade of the present millennium has given rise to new challenges, the ministers, gathering at Leuven/Louvain-la-Neuve in 2009, broadly stated the issues that need to be addressed in a changing environment. They called for a quality higher education and set the following four main goals for the present decade:

- finalizing the structural reform and deepening its implementation through a consistent understanding and use of the developed tools;
- implementing quality higher education, connected with research and lifelong learning and promoting employability;
- making the social dimension become a reality by ensuring that the student body entering and completing higher education reflects the diverse student body of Europe's populations;
- ensuring that at least 20 % of those graduating in the European Higher Education Area (EHEA) have had a study or training period abroad<sup>(1)</sup>.

### The report

The report, which reflects the framework of the Leuven/Louvain-la-Neuve Communiqué, is the result of a joint effort by Eurostat, Eurydice as well as by Eurostudent and has been overseen by the Bologna Follow-up Group and more specifically by a working group established by the latter. In line with the specific mission and methodology of the aforementioned data collectors, the report describes the state of implementation of the Bologna Process in 2012 from various perspectives and with data ranging from 2010 to 2011 as well as with earlier trends data for some statistical figures. Thus the report provides statistical data as well as contextualized information and it compares social and economic data on student life. Statistical evidence is complemented by normative system descriptors as well as by an analysis of how the system works. The former scorecard indicators have been newly revised by the Bologna Follow-up Group and integrated into the report as Bologna indicators.

Those former scorecard indicators carry value judgements expressed through the use of the dark green, light green, yellow, orange and red colour scheme. As compared with previous exercises, the colour dark green is less prevalent in some action lines than before. This is due to the fact that a more

## SOMMARIO ESECUTIVO

### Il Processo di Bologna e i suoi obiettivi per il 2020

Il Processo di Bologna ha cambiato il volto dell'istruzione superiore in Europa. Infatti tutti i paesi hanno apportato cambiamenti significativi che hanno reso possibile la nascita dello Spazio europeo dell'istruzione superiore, e che hanno gettato le basi per un'istruzione superiore capace di soddisfare una crescente serie di esigenze sociali; la struttura dell'istruzione superiore è stata modificata, sono stati sviluppati sistemi di assicurazione di qualità, si sono stabiliti meccanismi per favorire la mobilità, e sono stati individuate varie questioni legate alla dimensione sociale dell'istruzione superiore. La portata di un progetto, basato sulla cooperazione volontaria, che stabilisce e attua obiettivi comuni per i sistemi di istruzione superiore di 47 paesi non ha precedenti.

Tuttavia, consci del fatto che il secondo decennio del millennio in corso ha dato origine a nuove sfide, i ministri, riunitisi a Lovanio/Louvain-La-Neuve 2009, hanno stabilito in generale le questioni da affrontare in un ambiente in evoluzione. Hanno affermato l'esigenza di un'istruzione superiore di qualità e posto i seguenti quattro obiettivi fondamentali per il decennio in corso:

- definire la riforma strutturale e migliorare la sua implementazione attraverso una comprensione e un utilizzo costante degli strumenti sviluppati;
- attuare un'istruzione superiore di qualità, legata alla ricerca e all'apprendimento permanente e che favorisca l'occupabilità;
- dare concretezza alla dimensione sociale, garantendo che il corpo studentesco che accede all'istruzione superiore e la porta a termine rispecchi il variegato complesso degli studenti delle popolazioni europee;
- fare sì che almeno il 20% dei diplomati nello Spazio europeo dell'istruzione superiore (EHEA) abbia trascorso un periodo di studio o tirocinio all'estero<sup>(1)</sup>.

### Il rapporto

Il rapporto, che rispecchia la struttura del Comunicato di Lovanio/Louvain-la-Neuve, è il risultato dell'impegno congiunto di Eurostat, Eurydice ed Eurostudent sotto la supervisione del Gruppo di follow-up di Bologna, e più precisamente di un Gruppo di lavoro da quest'ultimo istituito. In linea con la missione specifica e la metodologia degli organismi sopraindicati, il rapporto descrive lo stato di implementazione del Processo di Bologna nel 2012 da vari punti di vista e con dati dal 2010 al 2011 integrati da dati sulle tendenze precedenti per alcuni valori. Il rapporto fornisce quindi sia dati statistici sia informazioni contestualizzate, e mette a confronto dati sociali ed economici inerenti alla vita degli studenti. I risultati statistici sono integrati da descrittori del sistema normativo e da un'analisi del funzionamento del sistema. I precedenti indicatori della scheda di valutazione sono stati rivisti dal Gruppo di follow-up di Bologna e inseriti nel rapporto come indicatori di Bologna.

Quei primi indicatori della scheda di valutazione esprimono giudizi di valore espressi attraverso l'utilizzo dello schema di colori verde scuro, verde chiaro, giallo, arancione e rosso. Rispetto alle indagini precedenti, il colore verde scuro non è più prevalente come prima in certe linee d'azione. Ciò

(1) The Bologna Process 2020 – The European Higher Education Area in the new decade. Communiqué of the Conference of European Ministers Responsible for Higher Education, Leuven and Louvain-la-Neuve, 28-29 April 2009.

(1') Il processo di Bologna 2020 – Lo Spazio europeo dell'istruzione superiore nel nuovo decennio. Comunicato della Conferenza dei ministri europei responsabili dell'istruzione superiore, Lovanio/Louvain-la-Neuve, 28-29 aprile 2009.

# sentence alignment

## The Bologna Process and its objectives for 2020

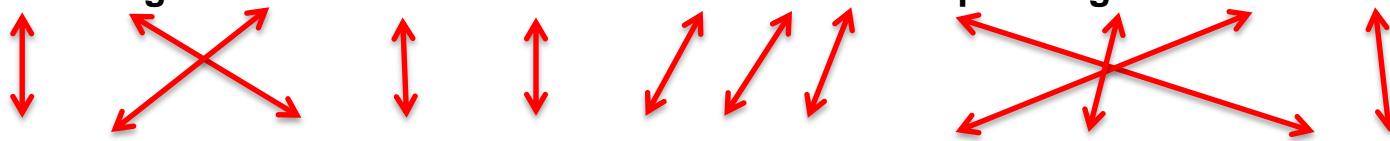
The Bologna Process has transformed the face of European higher education. Indeed all countries have made significant changes that have enabled the European Higher Education Area to emerge, and which have laid the ground for higher education that is serving an increasing range of societal demands; higher education structures have been modified, quality assurance systems developed, mechanisms to facilitate mobility established, and a range of issues for the social dimension of higher education identified. The scale of a project that, on the basis of voluntary cooperation, agrees and implements common objectives for the higher education systems of 47 countries is unprecedented.

## Il Processo di Bologna e i suoi obiettivi per il 2020

Il Processo di Bologna ha cambiato il volto dell'istruzione superiore in Europa. Infatti tutti i paesi hanno apportato cambiamenti significativi che hanno reso possibile la nascita dello Spazio europeo dell'istruzione superiore, e che hanno gettato le basi per un'istruzione superiore capace di soddisfare una crescente serie di esigenze sociali; la struttura dell'istruzione superiore è stata modificata, sono stati sviluppati sistemi di assicurazione di qualità, si sono stabiliti meccanismi per favorire la mobilità, e sono stati individuate varie questioni legate alla dimensione sociale dell'istruzione superiore. La portata di un progetto, basato sulla cooperazione volontaria, che stabilisce e attua obiettivi comuni per i sistemi di istruzione superiore di 47 paesi non ha precedenti.

# word alignment

The Bologna Process has transformed the face of European higher education .



Il Processo di Bologna ha cambiato il volto dell' istruzione superiore in Europa .

# phrase extraction

The Bologna Process has transformed the face of European higher education .  
Il Processo di Bologna ha cambiato il volto dell' istruzione superiore in Europa .



The | Process | Bologna | has | transformed | the | face | of | education | higher | European  
II | Processo | Bologna | ha | cambiato | il | volto | dell' | istruzione | superiore | in | Europa

The Bologna Process | has transformed | the face | higher education | European  
Il Processo di Bologna | ha cambiato | il volto | istruzione superiore | in Europa

The Bologna Process has | the face of | European higher education  
Il Processo di Bologna ha | il volto dell' | istruzione superiore in Europa

The Bologna Process has transformed | the face of European higher education  
Il Processo di Bologna ha cambiato | the face of 'higher education in Europe'

| has transformed the face  
| ha cambiato il volto

The Bologna Process has transformed the | has transformed the face of European higher education  
Il Processo di Bologna ha cambiato il | ha cambiato il volto dell' istruzione superiore in Europa

# phrase extraction

# machine translation

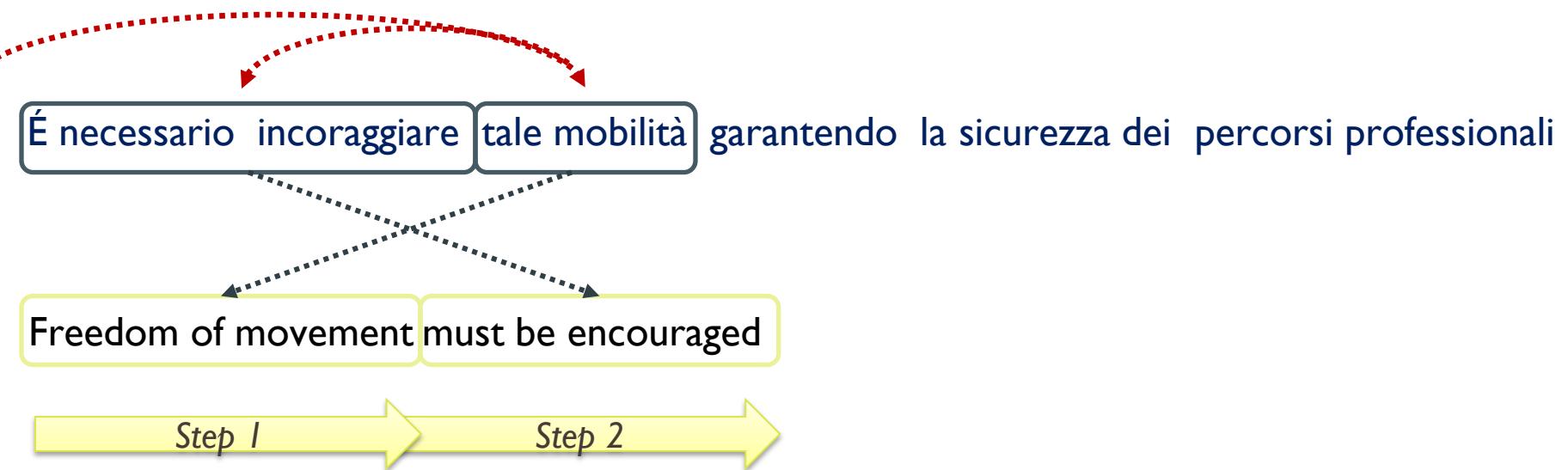
- training phase
  - sentence alignment
  - word alignment
  - extraction of phrases
  - ....
- **translation phase**
  - choice of phrases and position

# translation algorithm

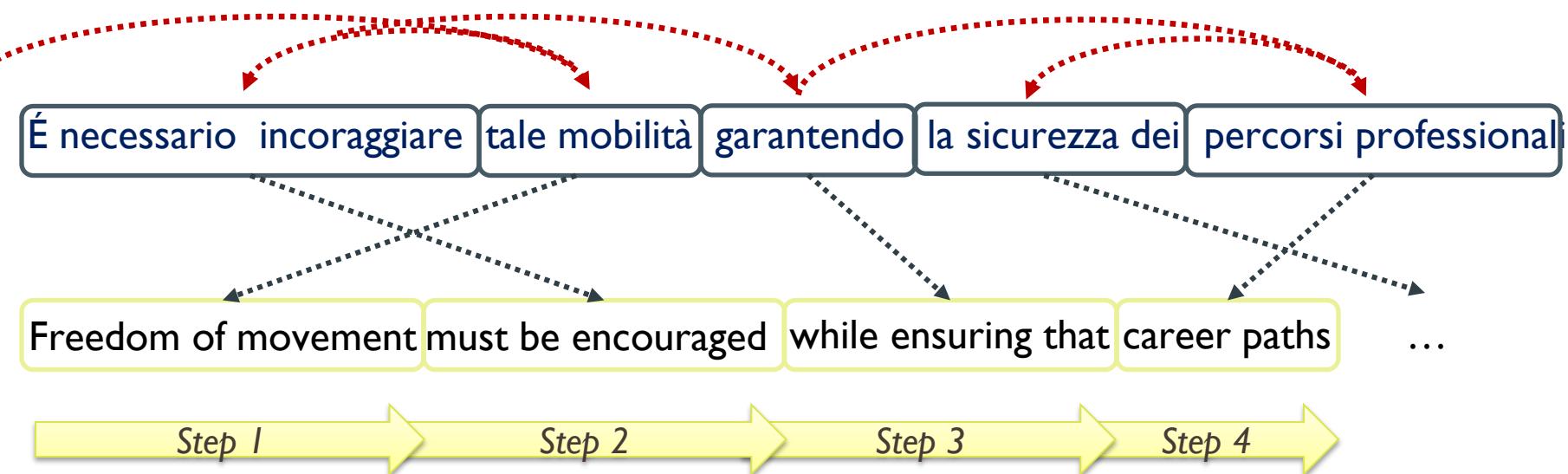
È necessario incoraggiare tale mobilità garantendo la sicurezza dei percorsi professionali

- Find the best sequence of steps:
  - **Step 1** Choose a phrase and its translation
  - **Step 2** Choose a phrase and its translation
  - ....
  - **Stop** when the translation is complete

# translation algorithm



# translation algorithm



# ModernMT



- > one-button installation
- > zero training time
- > instant context adaptation
- > learns from user data streams
- > scales with data and users
- > build a real product



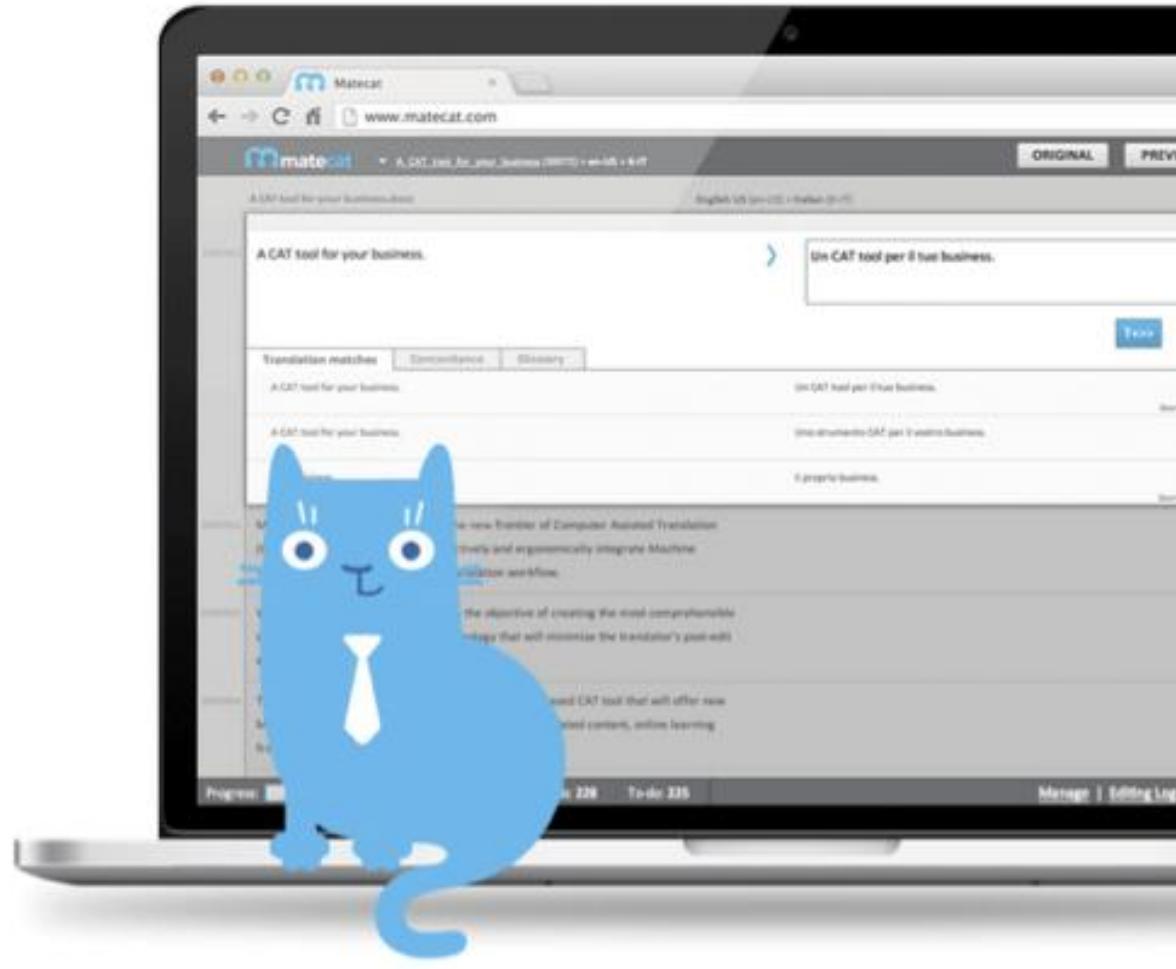
**First release 2Q 2016**

# conclusion

- translation from scratch vs. post-editing suggestions
- boring stuff left to machines
  - repetitive portions, technical terms, formatting issues
- keep translation skills for creative/difficult passages
- improvement of overall efficiency and user experience

integration of machine and human translation has gained traction in research and industry: lots of work in progress

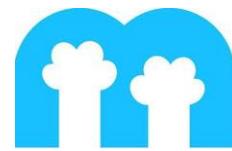
# Thank You!



## Credits:

ModernMT

Next Generation  
Machine Translation

 matecat™

