



MT@EC: Working with Translators

29 November 2012

*Markus Foti
DG Translation
European Commission*

MT@EC

- **MOSES engine**
- **EURAMIS data**
 - 23 languages – multilingual data
 - Over 200 million segments
- **1750 Translators in 23 language departments**
 - But not all available all the time

Showing 30 hits / 30 (more results available)

TM: Archive **Doc. No.:** 002590000 **Req. Serv.:** EAC **Year:** 2005 **SL:** EN (Reverse) [Show doc.](#) [Download](#) [Feedback](#)

IT: L'inadempimento degli obblighi di cui all'articolo 1 è considerato come una grave **violazione** della convenzione e può comportare l'annullamento del finanziamento o della parte di esso ancora dovuta.

EN: Failure to meet the requirements set out in Article 1 is considered to be a serious breach of agreement and could result in the cancellation of the funding or part of the funding due at the time.

TM: Archive **Doc. No.:** 011180000 **Req. Serv.:** ENTR **Year:** 2005 **SL:** EN (Reverse) [Show doc.](#) [Download](#) [Feedback](#)

IT: Nonostante ciò, in questa fase della procedura resta valida l'obiezione relativa alla **violazione** dell'articolo 6, paragrafi 5 e 6, della direttiva 89/105/CEE prodotta dall'aggiornamento dell'elenco positivo a norma dell'articolo 9 della legge n. 178/2002 e dai decreti del 27 settembre e del 21 novembre 2002.

EN: In spite of this fact, the objection concerning the infringement of Article 6(5) and 6(6) of Directive 89/105/EEC by the review of the positive list pursuant to Article 9 of Law 178/2002 and the decrees of 27 September and 21 November 2002 has to be maintained at this stage of the procedure.

IT: Restano gli effetti della **violazione** per quanto riguarda i medicinali esclusi dal rimborso senza rispettare le prescrizioni procedurali della direttiva.

EN: The effects of the violation remain as regards those medicinal products excluded from reimbursement without respect of the procedural requirements of the directive.

TM: Archive **Doc. No.:** 001610200 **Req. Serv.:** TRADE **Year:** 2005 **SL:** EN (Reverse) [Show doc.](#) [Download](#) [Feedback](#)

IT: È inoltre opportuno specificare le conseguenze della **violazione** degli impegni assunti o il recesso da essi.

EN: it is also appropriate to lay down the consequences of breach or withdrawal of undertakings.

IT: In caso di **violazione** o di revoca di un impegno ad opera di una delle parti, può essere istituita una misura correttiva a norma dell'articolo 10, in base ai fatti disponibili nel quadro dell'inchiesta nel cui ambito è stato accettato l'impegno.

EN: In case of breach or withdrawal of undertakings by any party, a balancing measure shall be imposed in accordance with Article 10, on the basis of the facts available within the context of the investigation which led to the undertaking.

TM: Legis-Juris **Doc. No.:** 12004V/AFI/DCL **Year:** 2004 **SL:** EN (Reverse) [Show doc.](#) [Download](#) [Feedback](#)
Doc. Type: Constitution for Europe

IT: La morte non si considera cagionata in **violazione** del presente articolo se è il risultato di un ricorso alla forza resosi assolutamente

EN: Deprivation of life shall not be regarded as inflicted in contravention of this article when it results from the use of force

DGT Internal Usage

What do we offer?

Since May 2011

52 language pairs

- **English into all EU languages**
... and vice-versa
- **French into DE, EL, EN, IT, ES, NL, PT**
- **German and Spanish into FR and EN**

DGT Internal Usage

Machine Translation automatically generated for incoming translation requests in MS Office formats **from English** into 17 languages:

July 2011: 10 languages

- **BG, DA, ES, FR, IT, NL, PL, PT, RO, SV**

March 2012: 6 additional languages

- **CS, FI, LT, LV, MT, SL**

April 2012: 1 additional language joins in

- **EL**

Open to all other DGT translators on demand

DGT Internal Usage

Real-life trial for 17 language pairs from English

- Since 15 July 2011: 235K files, ~ 3.3M pages

On-demand translations for 52 LPs

- Since 16 May 2011: 12658 files, ~ 230K pages
- Used by more than 500 individual users



European
Commission

Usable through TWB

The image shows a screenshot of a computer screen with two windows open. The top window is Microsoft Word, displaying a document titled "AGRI-2011-62626-02-00-IT-ORI-00.DOC". The text in the document is as follows:

30 ai 60 cm, e quindi essere posti in cassette disposti in fascetti.

Per la cipolla da consumo fresco i bulbi privati dalla tunica esterna vengono sottoposti alla spuntatura delle code, con taglio variabile dai 35 ai 60 cm, e poi riuniti in fasci di 1,5-6 kg e posti in cassoni o cassette.”

For onions for fresh consumption bulbs private from the outer skin are submitted to the clipping the Code, cut of 35 to 60 cm and then assembled in bundles of 1,5-6 kg and placed in trays or crates. ”

Article 9, comma 2, paragrafo 4

The bottom window is SDL Trados Translator's Workbench, showing the same text translated into English. The interface includes a menu bar (File, Settings, View, Options, Tools, Help) and a toolbar. The text in the workbench is:

Per la cipolla da consumo fresco i bulbi privati dalla tunica esterna vengono sottoposti alla spuntatura delle code, con taglio variabile dai 35 ai 60 cm, e poi riuniti in fasci di 1,5-6 kg e posti in cassoni o cassette.”

Per la cipolla da consumo fresco i bulbi privati dalla tunica esterna vengono sottoposti alla spuntatura delle code, con taglio variabile dai 35 ai 60 cm, e poi riuniti in fasci di 1,5-6 kg e posti in cassoni o cassette.”

For onions for fresh consumption bulbs private from the outer skin are submitted to the clipping the Code, cut of 35 to 60 cm and then assembled in bundles of 1,5-6 kg and placed in trays or crates. ”

The bottom status bar of the workbench shows a zoom level of 75% and a match count of "Match 1 of 5".

Machine Translation User Group

Representatives from all language departments

Task Forces on:

- **Conferences**
- **Best Practices**
- **Evaluation Methodology**

Evaluation

What questions need answering?

- **Is the next generation better?**
- **Paper cuts: small but irritating flaws**
- **Are experiments paying off?**



European
Commission

Engines (generations)

My tasks Welcome Markus [[Logout](#) | [Profile](#) | [Password](#)]

Rating sentence 2 from task 2. Source: 2011 vs. 2012, en-it. 0 of 100 sentences rated.

[<previous](#) [<previous unrated](#) [\[exit task\]](#) [next unrated](#) [> next](#)

Source:
CBRN risk mitigation actions enable the EU to collectively contribute to reducing the threat of malevolent uses of weapons of mass destruction (WMD) materials and know-how.

Translation 1:
Materia di attenuazione dei rischi CBRN azioni consentire all' UE di contribuire collettivamente a ridurre la minaccia di loro usi delle armi di distruzione di massa (ADM) materiali e competenze tecniche.

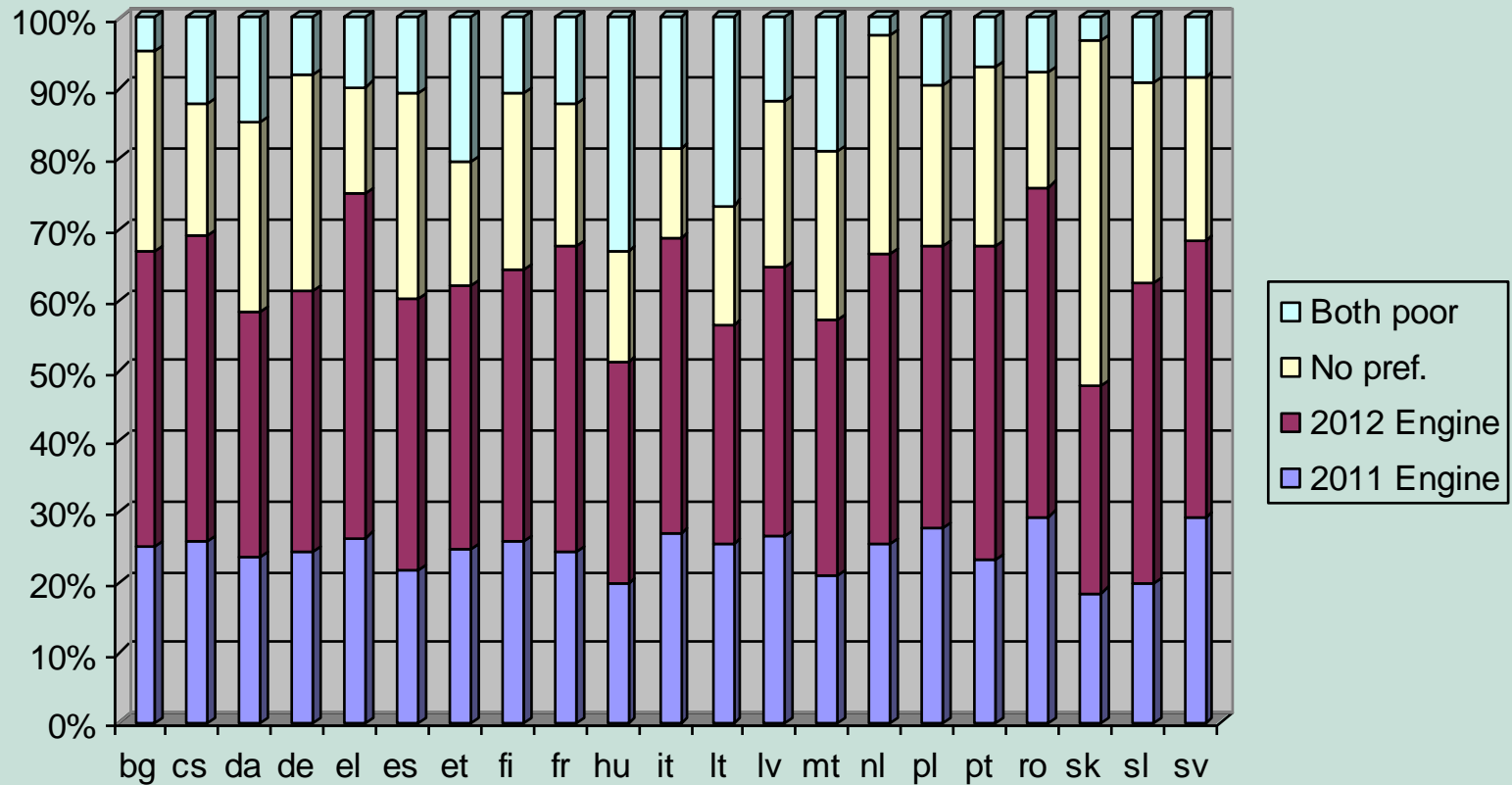
Translation 2:
Le azioni di attenuazione dei rischi CBRN consentire all' UE di contribuire collettivamente a ridurre la minaccia di usi malevoli delle armi di distruzione di massa (ADM) materiali e di know-how.

[\[remove highlighting\]](#)

Which translation would you prefer for post-editing?

- Translation 1: Translation 1 is better than Translation 2.
- Translation 2: Translation 2 is better than Translation 1.
- No preference: Both translations are of comparable quality.
- Both poor: The low quality of the translations makes it impossible to compare them.

2011 vs 2012 Engines, Human Evaluation



Improving Quality

Typographical Errors Survey

Aim: obtain from users the most frequent and irritating – but easy-to-fix – errors

58 errors reported by 16 of 23 LDs

Improving Quality

35 of 58 errors fixed, e.g.

- **Extraneous spaces around quotation marks**
- **Corruption of characters in LT (a, č, e, è, ĺ, š, u, ū, ž)**
- **Portuguese spelling reform**
- **Maltese spelling reform**

Normalisation

Spelling - a few examples:

- **German:** replace incorrect beta character by “ß”
- **Greek:** correct some frequent “Greeklish” abbreviations mixing Latin and Greek characters, e.g. “EOK” (3 Greek letters) instead of “EOK” (Latin E, Greek Omikron, Latin K)
- **Italian:** correct grave accents on some common words, e.g. “piu`” -> “più”
- **Dutch:** correct capitalisation of IJ, e.g. “IJsland” instead of “Ijsland”
- repair incorrectly encoded characters
- etc.

Improving Quality

11 errors require a more complex approach.

Letter case in references (court cases, official journal):

- **COMP/M.727 --> COMP/m/727,**
- **SA.33618 --> sa.33618**
- **OJ C307 --> EYVL c307**

Improving Quality

12 responses were out-of-scope (TWB problems, Word problems, etc.).

Limit of Bold, Italics:

- **Bruselas, el 29 de marzo de 2012 – hoy, la Comisión Europea puso en marcha***
- *Should be:* **Bruselas, el 29 de marzo de 2012** – hoy, la Comisión Europea puso en marcha

Improving Quality

More, better data

3rd generation of engines being built for end-of-year

- Additional data up to August 2012
- Incorporation of fixes for 38 common typographical errors

Ongoing testing

- **German/ Hungarian – word reordering**
Positive results
- **Stemming and adding endings for Finnish/Hungarian**
In progress
- **“Turing smoothing” statistical analysis**
Somewhat positive
- **Inclusion of IATE (terminology database)**
Poor results

Other Potential Tools

PET post-editing tool

- **Measures keystrokes, time taken, distance between MT and final version**

(Wilker Aziz, U. of Wolverhampton and Lucia Specia, U. of Sheffield)

- **Great for evaluation:**

But for producing a final translation?

Looking for comparison in effort across languages, potentially over time for same language (but how to account for variation between individuals and texts)



European
Commission

editing...		part 46	revisions: 0	total 0s
"Consulté aún a mis colegas, pero parece que nosotros realmente no tenemos nada que le parezca eso".				1/50
According to the testees, there is no such groovy glossary of slang terms in Social Democrats Party.	Según los evaluadores, no existe ningún glosario de términos coloquiales tan bueno del Partido socialdemócrata.			
"I even consulted my colleagues, but it seems that we really do not have anything like that."	"Consulté aún a mis colegas, pero parece que nosotros realmente no tenemos nada que le parezca eso".			
"We rather address one another with names, sometimes we use abbreviations and name garbles," said Prague CSSD chairman Peter Hůlinský, addressed by party members as Hůlida.	"Estamos en lugar de abordar con nombres, a veces uso abreviaturas y nombre garbles", dijo el presidente peter hulinsky, dirigida por los miembros del partido como hulida.			
Accusations of witchcraft are also common in other African countries.	Las acusaciones de brujería también son comunes en otros países africanos.			
While in Burkina Faso the government is trying hard to moderate the witch hunt, the Gambia's President himself participated in the hunt last year.	Mientras que en Burkina Faso el gobierno está intentando difícilmente moderar la caza de brujas, el presidente mismo de la Gambia participó en la caza el año pasado.			
"Gambian President Jámí after the death of his aunt unleashed a large-scale witch-hunt for witches allegedly involved in her death and trying to use black magic to kill the president."	"Presidente de Gambia Jámí después de la muerte de su tía decretó una caza de brujas a gran escala para presuntamente implicados en su muerte y tratando de utilizar la magia negra para asesinar al presidente".			
The arrested people had to drink a potion of herbs with hallucinogenic effects, which also causes intestinal and kidney problems.				
				0 saved

Figure 1: Annotation window

The Holy Grail...

*Automatic feedback harvested through
integration with Trados Studio*

Translation distance measured using HTER(?)



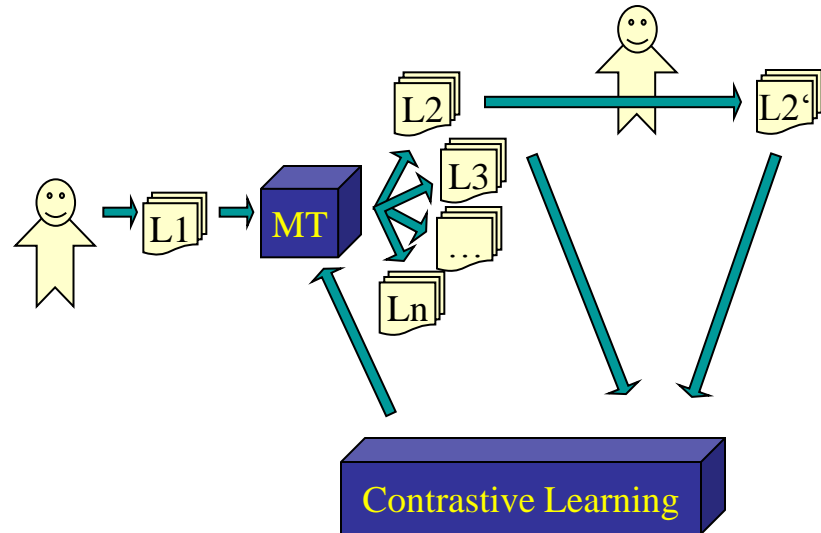
Harvesting Feed-Back from Translators

Improvements during Post-Editing are collected

Models are updated with these improvements

Translators who provide feed-back are rewarded

MT quality will improve in the direction that is most important to translators





European
Commission



Thank you