European Commission
machine translation
collaboration across linguistic barriers

Spyridon Pilos
Head of language applications
Directorate-General for Translation

Riga, 28.04.2015
Machine translation for the EU

- EU and languages
- MT@EC and its users
- Automated translation platform
EU official languages over time
EU translation services

1200 European Parliament, Luxembourg
141 Court of Auditors, Luxembourg
47 European Investment Bank, Luxembourg
600 Council Secretariat, Brussels
2500 European Commission, Luxembourg – Brussels
530 Committee of the Regions and Economic and Social Committee, Brussels
65 European Central Bank, Frankfurt
200 Translation Centre, Luxembourg
851 Court of Justice, Luxembourg
47 Interpretation Services
65 Lawyer-linguists

Interinstitutional cooperation on translation and interpretation
Why does the Commission need MT?

The Commission...

- DGT has **1700** translators
- **Over 2 M pages** translated in 2014

But...

...just to make europa.eu fully multilingual
more than **7 M documents** would have to be translated
or **8 500 translators/year**!

The result:

**Thousands of non-translated documents**
(and this does **not** include
user generated content and social media)**
There are also interactions with and between actors in the Member States.
Wouldn’t it be great if I could start using a public service in any Member State from any place and obtain the information in my mother tongue?
Interoperability addresses the need for:

- **Cooperation** among public administrations, to establish public services
- **Exchanging information** among public administrations to fulfil legal requirements or political commitments
- **Sharing and reusing information** among public administrations to increase administrative efficiency and cut ‘red tape’ for citizens and businesses

The results are:

- **Improved public service delivery** to citizens and businesses, by facilitating the ‘one-stop-shop’ delivery of public services
- **Lower costs** for public administrations, businesses and citizens due to the efficient delivery of public services
A framework: EIF
European Interoperability Framework

A programme: ISA
Interoperability Solutions for public Administrations
http://ec.europa.eu/isa/

Sharing results: Joinup
An open collaborative platform supporting interoperability in Europe
http://joinup.ec.europa.eu/
**Recommendation 4.** Public administrations should use information systems and technical architectures that cater for multilingualism when establishing a European public service.
The role of Machine Translation

MT is the only viable solution for:

- quick and cheap access to information in foreign languages.
- understanding information received in a foreign language that otherwise could not be used or would require substantial time and costs to translate.
- making multilingual use of websites possible.
- facilitating cross-lingual information search and analytics.

That is why machine translation (MT) is a critically important technology for multilingual Europe.
MT@EC: a European Commission product

Released: 26 June 2013

Languages: All directions between 24 EU official languages

Technology: Statistical machine translation using open source software Moses co-funded by EU Framework Programmes for research and innovation

Development by DGT between 2010-2013 co-funded by the ISA programme (action 2.8)

http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm
MT@EC description

Delivery:  - web user interface (human to machine)
          - web services (machine to machine)

Special features:

• **User interface** in 24 languages
• Source document format/formatting **maintained** [not for pdf]
• Specific output **formats for translation:** tmx and xlf
• Translation can also be returned by **email**
• Can translate **multiple** documents to **multiple** languages
• Indication of **quality** for language pairs (using BLEU Scores)
• **Feedback** mechanism (using EU Survey)
MT@EC - Machine Translation

Select at least one source file:  
OR

Translate from...

Translate into...

Output format:  [Same as source]  [TMX]  [XLIFF]

[ ] E-mail me my translation.
[ ] Delete after download.

Translate document

Selected files

- DG Briefing docx (34 kb) 100% Remove
Translate documents | Translate text | My translation requests

Text to translate:
Good morning to everyone!

Translate from...
EN

Translate into...
BG CS DA DE
EL ES ET FI
FR GA HR HU
IT LT LV MT
NB NL PL PT
RO SK SL SV

E-mail me my translation.
Delete after download.

Translated text:
Buongiorno a tutti!
Who can use MT@EC today?

- European institutions and bodies:
  - Commission
  - Parliament
  - Council
  - Court of Justice
  - Court of Auditors
  - Economic and Social Committee
  - Committee of the Regions
  - European Central Bank,
  - European Investment Bank etc.

- Online services funded or supported by the EU
- Public administrations in the EU Member States
<table>
<thead>
<tr>
<th>Country</th>
<th>reg'd</th>
<th>using</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Belgium</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Croatia</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cyprus</td>
<td>81</td>
<td>55</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Estonia</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Finland</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>34</td>
<td>25</td>
</tr>
<tr>
<td>Germany</td>
<td>47</td>
<td>37</td>
</tr>
<tr>
<td>Greece</td>
<td>46</td>
<td>34</td>
</tr>
<tr>
<td>Hungary</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ireland</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Users in public administrations**

registered **520**

using **359**

69.4%

If you are wish to try it request access to DGT-MT@ec.europa.eu

<table>
<thead>
<tr>
<th>Country</th>
<th>reg'd</th>
<th>using</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>Latvia</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Lithuania</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Malta</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Netherlands</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Poland</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Portugal</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>Romania</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Slovakia</td>
<td>91</td>
<td>42</td>
</tr>
<tr>
<td>Slovenia</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Spain</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>Sweden</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>UK</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
# Online services connected to MT@EC

<table>
<thead>
<tr>
<th>Service</th>
<th>Description/URL/use</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMI</td>
<td>Internal Market Information System – an intranet for more than 7000 public administrations in the EU Member States <a href="http://ec.europa.eu/internal_market/imi-net/index_en.html">http://ec.europa.eu/internal_market/imi-net/index_en.html</a> ➔ free-text message exchange</td>
</tr>
</tbody>
</table>
What next?

from MT@EC... to the CEF automated translation platform

**CEF.AT** will:
- build on the *existing* MT@EC service - but not be limited to it
- put emphasis on *secure, quality, customisable* MT for pan-European online services - but not be limited to them
- be a *multilingualism enabler* – not only MT
Users and Services

DISPATCHER
managing MT requests

MT engines by language

MT data
language resources specific for each MT engine

Language resources
Euramis (EU)

Customised interfaces

MT@EC SERVICE

ENGINES & DATA

USER FEEDBACK

Generic "EU" MT
CEF.AT platform

Outline

Digital Service Infrastructures

- Dispatcher managing MT requests
- MT engines by language, domain...

MT service

Engines factory

Language resources

- Multilingual corpora
- Monolingual corpora
- NLP Tools
- Other

From data to engines

SECURE (and PERFORMING)

QUALITY

CUSTOMISABLE

Collect and clear
For CEF.AT we will be looking for:

- even better *language technologies*  
  *both open source and commercial products*
- more quality *language resources*  
  *both open and licenced data*
- even better *ways to enable multilingualism*  
  *both from research and market*

(and appropriate infrastructure to *scale up*)
DGT contributes to this effort...

... as *solution provider* for the CEF.AT platform

... publishing **DGT-TM** a *public* translation memory with almost **1.5 billion words** in **24 languages**

- on the EU Open Data Portal

- on the EC Joint Research Centre Language Technology web site

... supporting networking of translation stakeholders across the EU through the **Translating Europe** initiative, aimed to create a community of interest on translation


---

The time for making Digital Europe multilingual is **NOW**.
DGT events for your agenda

Annual Translating Europe Forum
Brussels, 29-30 October 2015
Focus: Youth

Annual MT@Work Conference
Brussels, 11 December 2015
Focus: MT in translation practice

spyridon.pilos@ec.europa.eu
dgt-mt@ec.europa.eu