MT@EC
European Commission machine translation for Public Administrations in the EU Member States

Contact: DGT-MT@ec.europa.eu
Updated: 26.09.2014
European Commission machine translation for public administrations

- European Commission and languages
- MT@EC: machine translation for the EU
- Access to MT@EC for public administrations
- What next?
EU official languages over time

- 1958: Dutch, French, German
- 1963: Greek
- 1973: English, Danish
- 1978: Swedish, Portuguese, Finnish, Slovak, Hungarian
- 1981: Croatian
- 1986: Lithuanian, Maltese
- 1995: Irish
- 2004: Slovak, Czech, Estonian
- 2007: Polish, Romanian
- 2013: Bulgarian
EU translation services

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

Interinstitutional Committee for Translation and Interpretation

Interpretation Services

Lawyer-linguists
Why do we need machine translation?

*The Commission...*
- DGT has 1700 translators
- Over 2 M pages translated in 2013

*But...*

...just to make europa.eu fully multilingual
almost 6.8 M documents to be translated
or 8 500 translators/year!

*The result:*
Thousands of non-translated documents
(and this does not include user generated content)
There are also interactions...
MT and EC: a long history

**Started in the 1970s**
- Eurotra (78-92): research, high expectations
- Rule-based ECMT (75-97), costly to develop – not scalable
  (18 language pairs in 20 years - coverage of post-2004 languages never attempted- system shut down in 2010)

**Data-driven systems (Statistical MT):**
- Economic and quick to develop... if you have good data
- EC needs solution for all EU languages... and has good data

**EC action plan (2009), Inter-service task force (2010)**
- The goal: MT@EC offering machine translation for all languages to and from English, operational in July 2013
Basic Architecture for Statistical MT
Two Usage Scenarios for MT@EC

**MT for understanding (inbound)**

Robustness, Coverage

Practically unlimited demand; free web-based services cover much of it

---

**Requirements for MT@EC**

- Provide MT as a (simple and robust) service
- Optimise quality for understandability (gisting)
- Deal with many domains, document types, formats, ...
- Scale to huge volumes
**Two Usage Scenarios for MT@EC**

**MT for dissemination (outbound)**

*Textual quality*

Publishable quality can only be authored by humans; Translation Memories & CAT-Tools used by professional translators

**Requirements for MT@EC**

- Provide MT as a tool within a CAT workflow
- Develop new ways to incorporate feedback
  - explicit feedback on MT quality, implicit feedback via TM
  - improvements requiring language-specific knowledge
  - towards hybrid approaches
- Optimise quality for post-editing
MT@EC: a European Commission product

Released: 26 June 2013 (version 1.0)
Version 2.0 released on 3 July 2014

Languages: All 24 EU official languages
552 language pairs (62 direct)

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)

* Interoperability solutions for public administrations
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)

Security:  Host (EC data centre) + access (ECAS)
           + transfer (sTESTA)

Special features:
  • Source document format/formatting maintained [not for pdf]
  • Specific output formats for translation: tmx and xlf
  • Can translate multiple documents to multiple languages
  • Translation can also be returned by email
  • Indication of quality for language pairs
  • Feedback mechanism
  • User interface in 24 languages
MT@EC is already available to...

... the staff of European institutions and bodies: Commission, Parliament, Council, Court of Justice, Court of Auditors, Committees, Central Bank, Investment Bank, Translation Centre, etc.

... online services funded or supported by EC

Currently: IMI, SOLVIT

In the pipeline: eJustice portal, JoinUp, TED, nLEX, Open Dispute Resolution portal, CircaBC, etc.
MT@EC for public administrations

Free real-life trial until end of 2015:
- **Staff** can have direct free access to the standard MT@EC service (upon request)
- Your **Organisation** can participate in a customisation pilot project, where DGT can also build specific engines with your data (bilateral cooperation agreement)*

Two levels of **Security** for document transfers possible:
- over sTESTA**, a very secure private network between public administrations in the EU, separate from the internet
- over the internet through a secure (https) connection

* Administrative Agreements signed only until end of June 2015
** You can check if your organisation has access to sTESTA on: https://portal.testa.eu/jetspeed/portal/homepage/about.psm.l
Customisation pilots

Pilot A: Connect an information system of yours to the standard MT@EC service.

Pilot B: DGT builds custom engines (your data) available to all through MT@EC

Pilot C: DGT builds custom engines (your data) available only to you through MT@EC

Pilot D: DGT builds custom engines (your data) for you to run in your premises

Pilot E: DGT assists you to build your own custom engines for you to run in your premises

If you are interested email DGT-MT@ec.europa.eu
Staff access to MT@EC

• Get an individual **ECAS** user name and password (self-registration) using your **work email address**. [go to https://webgate.ec.europa.eu/cas/eim/external/register.cgi and follow the instructions]

• Send an **email** to **DGT-MT@ec.europa.eu** asking for access either over sTESTA or over the internet.

• DGT will activate your access and inform you by email.
You can ask for documents...

Machine Translation

<table>
<thead>
<tr>
<th>Translate documents</th>
<th>Translate text</th>
<th>My translation requests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Translate documents**

Select at least one source file:

Translate from...

Translate into...

Output format:

- Same as source
- TMX
- XLIFF

E-mail me my translation.

**Translate document**

**Selected files**

- MT at the Commission.doc (209 kb) **100%** Remove
- Budget flows.xlsx (8 kb) **100%** Remove
Tab: Translate documents

• Click on "Choose files..." [notice that only the acceptable formats are proposed]
• Select the file(s) you wish to translate [notice that these appear in the box "Selected files"]
• Select language to "Translate from..." [no language detection for the time being...]
• Select language(s) to "Translate into..." [notice the quality indication (medals)]
• Select output format ["same as source", tmx or xlli]
• Decide on reception by email
Tab: Translate documents

- Click on "Submit documents to translate"  
  [Notice message: "Your request has been submitted... "]
- Wait
- Result (or Notification) sent by email  
  [also available to download under "My Translation requests"]
Tab: Translate text

- Type/paste the source text in box "Text to translate" (max: 4000 chars)
- Select language to "Translate from..."
  [automatic language detection also available]
- Select language to "Translate into..."
- Decide on reception by email
- Click on "Translate text"
- Wait
- Result appears in box "Translated text"
  [It may also be sent by email. It is also available for download under "my translation requests"]
You can then download the result ...

Machine Translation

<table>
<thead>
<tr>
<th>History</th>
<th>Original file name</th>
<th>Status</th>
<th>Message</th>
<th>Document format</th>
<th>Target format</th>
<th>Source language</th>
<th>Target language</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DDB - Budget flows</td>
<td>Ongoing</td>
<td></td>
<td>xlsx</td>
<td>EN</td>
<td>EN</td>
<td>EL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MT at the Commission.doc</td>
<td>Translated</td>
<td></td>
<td>doc</td>
<td>EN</td>
<td>EN</td>
<td>EL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Text</td>
<td>Translated</td>
<td></td>
<td>bit</td>
<td>EN</td>
<td>EN</td>
<td>HR</td>
<td></td>
</tr>
</tbody>
</table>
Tab: My translation requests

- Monitor progress of translation requests
- Links to download each translation (document or text)

NOTE:

Entries (and download links) remain available for 24 h
Translation done for request [110349_1]

Please find attached the machine translation of your document 000 - Budget flows.xls into Greek.

Should you have specific questions or comments on the attached translated document, we kindly ask you to forward this email to the DGT-MT@ec.europa.eu mailbox. Please include your comments in that email.

In general we would be happy to know to what extent you are satisfied with the machine translation service. Please consider completing our feedback form at http://ec.europa.eu/eusurvey/runner/97ca34c6-6154-59fe-d0ca-3af7aa75725b

Your feedback will help us improve MT@EC.

Thank you very much.
Email (one per document/request)

The email includes:

- The translated document or text *(if email option selected)*
- the notification and link for downloading document *(if email delivery option is not selected)*
- Link for providing feedback
- Link to functional mailbox *DGT-MT@ec.europa.eu*
"Maturity Check" (April-May 2011)
- Can baseline MT engines already be used as such?
- Identify main sources of problems for various languages, cluster them across languages

"Real-life trial" (July 2011-June 2013)
- Make first MT results available to translators
- Auto-MT for 10..19 “best” language pairs from EN (now: all)
- On-demand MT for others

Automatic scores
- BLEU scores for internal tuning and regression testing.
- Can help to identify domains/document types where MT is most useful, but also point to systematic difficulties.
Maturity check 2011 (EN->X)

DGT’s SMT maturity check outcome as a (useful useless) sentences ratio + morphology
From the translator's point of view

+ Aid for typing
+ time savings
+ “original” proposed solution
+ guides the terminological research

— gender/numbers and order of words
— can be "fluent", but with mistranslations
— omissions and additions
— risk of error when incorrect terminology suggested
— quality dependent on the quality of the originals
MT@EC: right for the EU

Quality:
- built on data derived from EU translations
  (*Euramis translation memory system: 800 M segments in 24 languages and annual growth rate > 20%*)
- designed for EU relevant collaboration
- team of *computational linguists* working with *translators and linguists* in DGT
- work to improve MT for *all* EU languages

Security
Customer support
MT@EC: what next

**CEF (Connecting Europe Facility)**
- A funding programme for building and deploying infrastructures.
- Includes deploying *mature* technologies to build, enable and operate pan-European **Digital Services**.
- Includes an Automated Translation (AT) platform as one of its core building blocks for digital services.
- A key component of the AT platform is MT@EC.
- Emphasis is placed on **secure, quality, customisable** machine translation.

**MT Conference for Public Administrations**
- on **5 December 2014**, in Brussels
- organised by the European Commission (DG Translation)
The automated translation platform

- To facilitate cross-border information exchange and enable cross-border access to online content and services provided by the digital service infrastructures of the CEF.
- To offer MT services to EU institutions and public administrations in the Member States.
- To build on the existing Commission Machine Translation service (MT@EC)
- Emphasis is placed on secure, quality, customisable machine translation.

Watch this space: