The CEF eTranslation platform @ work

Markus Foti
eTranslation Programme Manager
DGT R.3
eTranslation platform at a glance

**eTranslation**
- MT@EC
  - Launched June 2013
  - Legalese
  - Statistical (Moses)
- CEF.AT
  - More NLPs (transliteration, named entity recognition...)
  - Generic services
  - Projects (ELRC, market research)

**CEF.AT**
- Launched July 2017 (webservice for snippets), Nov. 2017 (web page for documents)
- Cloud based
- Neural engines
eTranslation platform at a glance

• Available for:
  — individuals (submit documents through a web page)
  — machine-to-machine use

• Users:
  — Digital Service Infrastructures (EESSI, ODR, Open Data Portal, Europeana, etc.)
  — System suppliers (EURLex, N-Lex, Internal Market Information system...)
  — Individuals in public administrations

• Benefits:
  — Increase speed and productivity
  — Reduce costs
  — Facilitate information exchange
MT@EC - Machine Translation

Select at least one source file: Choose files...

Translate from...

Translate into...

Domain: ALL

Output format: Same as source, TMX, XLIFF

E-mail me my translation.
Delete after download.

Translate document
Smarter rules for safer food: Commission proposes landmark package to modernise, simplify and strengthen the agri-food chain in Europe

1) What is "from farm-to-fork" policy?

The EU "from farm-to-fork" policy covers legislation to protect health by ensuring that the food we produce and import that ends up on our plates is safe to eat. This body of legislation is a complex, integrated system of rules which takes care of all the links in the food chain from: the protection of plants from pests, the protection of animals and humans from animal diseases (and from such diseases which are transmissible to humans) and the protection of human health from all the hazards that might arise along the agri-food chain. It also includes all those rules which allow the Single Market to function smoothly in this area and provides rules for the protection of consumers.

Official controls are carried out at various stages in the food production chain by EU Member States to ensure that the legislation is being correctly applied.
Règles plus intelligentes pour des données plus sûres»: Paquet décisif Com.
moderniser, de simplifier et agroalimentaire en Europe

1) Ce qui est «de la fourche à la fourchette»
L'UE «de la fourche à la fourchette» est un système de maillons de la chaîne alimentaire, la protection des consommateurs à partir de ces maladies et tous les dangers qui en découlent. Il comprend également des mesures de manière harmonisée pour les consommateurs.
One original, many translations...

All 24 EU languages
Many documents at once...

A document

Another document

…and another

Un document

Un autre

… et encore un
eTranslation protects your privacy

- All documents deleted after 24 hours or after delivery (on demand)
- IPR rights transferred (Your translation belongs to YOU!)
No one looking at your translations!
Machine-to-machine services
eTranslation on the N-Lex portal

N-Lex offers a single entry point to national law databases in individual EU countries.

Choose one of the countries below to search for legal acts or find out more about its legal system.
Machine-to-machine service
eTranslation on the N-Lex portal
Machine-to-machine service
eTranslation on the European Data Portal
Machine-to-machine service eTranslation on the European Data Portal

11 datasets found for "απογραφή πολωνία"

Methodology for migration abroad in the national population and housing ...
A population and housing census carried out in Poland in 2011 was applied to the new organisational and methodological approach, it was decided to move away from the traditional...

Commuting in Poland — Results of the NSP 2011
Study commuting in 2011 for total employment was carried out on the basis of the data contained in the registers of the official statistics, the National Census in 2011 (NSP...
Machine-to-machine service eTranslation on the European Data Portal
How to Connect to eTranslation

• Open to DSIs and public administrations

  ― Contact CEF-AT@ec.europa.eu with your request and use case
  ― We will provide the technical documentation on how to connect
    • SOAP request or RESTful interface
  ― Contact us for credentials
  ― Adapt your service to multilingual use!
• Statistical Machine Translation

  — MT@EC built only on "EU translations" (Euramis database)
  — Covers all 24 languages
  — Euramis (and MT@EC) cover EU policies, subjects and language but limited everyday language
  — Moses-based engines
  — Performs best when trained on large volumes of text pairs (source-translated) in specific domains
What is NMT?

– Machine learning: artificial neural networks trained on existing translations
– The computer devises its own rules on how to translate
– Radical departure from the phrase-based SMT approaches

Why is it important?

– Translations read better: more fluent and grammatical
– Better able to fill in gaps in data used for training
– Better for highly inflected languages (e.g. German, Hungarian)

This is where the field is headed

– Will SMT become obsolete?
Neural Machine Translation Engines

- EN into:

<table>
<thead>
<tr>
<th>BG</th>
<th>CS</th>
<th>DA</th>
<th>DE</th>
<th>EL</th>
<th>ES</th>
<th>ET</th>
<th>FI</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>HR</td>
<td>HU</td>
<td>IS</td>
<td>IT</td>
<td>LT</td>
<td>LV</td>
<td>MT</td>
<td>NB</td>
</tr>
<tr>
<td>NL</td>
<td>PL</td>
<td>PT</td>
<td>RO</td>
<td>SK</td>
<td>SL</td>
<td>SV</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Neural Machine Translation Engines

- EN into:

<table>
<thead>
<tr>
<th>BG</th>
<th>CS</th>
<th>DA</th>
<th>DE</th>
<th>EL</th>
<th>ES</th>
<th>ET</th>
<th>FI</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>HR</td>
<td>HU</td>
<td>IS</td>
<td>IT</td>
<td>LT</td>
<td>LV</td>
<td>MT</td>
<td>NB</td>
</tr>
<tr>
<td>NL</td>
<td>PL</td>
<td>PT</td>
<td>RO</td>
<td>SK</td>
<td>SL</td>
<td>SV</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Right now!*
Neural Machine Translation Engines

• EN into:

<table>
<thead>
<tr>
<th>BG</th>
<th>CS</th>
<th>DA</th>
<th>DE</th>
<th>EL</th>
<th>ES</th>
<th>ET</th>
<th>FI</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>HR</td>
<td>HU</td>
<td>IS</td>
<td>IT</td>
<td>LT</td>
<td>LV</td>
<td>MT</td>
<td>NB</td>
</tr>
<tr>
<td>NL</td>
<td>PL</td>
<td>PT</td>
<td>RO</td>
<td>SK</td>
<td>SL</td>
<td>SV</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By end 2018
Original Text (EU Commission Press Release, 10 April 2018):
With an average of 49 road fatalities per one million inhabitants, European roads remained by far the safest in the world in 2017. Compared to 2016, Estonia and Slovenia reported the largest drop in fatalities with respectively -32% and -20%.

MT@EC (SMT):
V povprečju živi v EU 49 smrtnih žrtev v cestnem prometu na milijon prebivalcev, evropske ceste še vedno najvarnejše na svetu v letu 2017. V primerjavi z letom 2016, Estonija in Slovenija sta sporočili, da so največji upad števila žrtev s -32 % oziroma -20 %.

eTranslation (NMT):
V povprečju je leta 2017 število smrtnih žrtev v cestnem prometu na milijon prebivalcev še vedno najvarnejše na svetu. v primerjavi z letom 2016 sta Estonija in Slovenija zabeležili največji padec števila žrtev s -32 % oziroma -20 %.
Original Text (Reuters, 14 April 2018):
Slovenia will hold an early general election on June 3, President Borut Pahor announced on Saturday, a month after center-left Prime Minister Miro Cerar resigned over a legal obstacle to a government investment project.

MT@EC (SMT):
Slovenija bo predčasne splošne volitve junija 3, predsednikom Borutom Pahorjem, v soboto sporočili, to je en mesec po odstopu predsednika vlade Mira Cerarja levosredinska pravno oviro na vladnih naložbenih projektov.

eTranslation (NMT):
Slovenija bo opravila prve splošne volitve junija 3, predsednik Borut Pahor, objavljen v soboto, en mesec po razglasitvi predsednika vlade Mira Cerarja, da bi s pravno oviro prenesel na naložbeni projekt vlade.
Original Text (Mail Online, 29 March 2018):
'In the eight days I spent taking pictures in Slonevia, I covered the whole country and managed to visit places very few people have spent time in. I will remember Slovenia for all of my life because I feel like a part of my heart remained there.'

MT@EC (SMT):
„Pri osem dni sem Slonevia slik, pri čemer je zajela celotno ozemlje države in upravljajo za obisk krajev zelo malo ljudi so nekaj časa preživeli v. Si bom zapomnil Slovenije za vse življenje, ker se počutim kot del moje srce je tam ostala.“

eTranslation (NMT):
„V osmih dneh, ko se je spregovorilo v Slovianu, sem zajel celotno državo in uspelo obiskati, kjer je nekaj ljudi preživelo čas. Za vso svojo življenje si bom zapomnil položaj Slovenije, ker se počutim kot del mojega srca.“
THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

[...]

HAS ADOPTED THIS DECISION:

Article 1

The total sum of greenhouse gas emissions covered by Decision No 406/2009/EC for each Member State for the year 2014 arising from the corrected inventory data upon completion of the comprehensive review referred to in Article 19(1) of Regulation (EU) No 525/2013 is set out in the Annex to this Decision.

[...]
Language resources: the key to success

- More data for all languages
- Better lexical coverage for all languages
- Machine translation adapted to your domain/subject matter
Future Improvements

Requested by users such as EDP, e-Justice, Europeana, EESSI, ODR...

• Speed
• Transliteration
• More formats (PDF output, JSON)
• Named-entity recognition
• New languages (AR, RU, TK, ZH)
• Reliable and trustworthy translation for EU and National Public Administrations
• Support for languages with fewer speakers
• Opportunities for private sector, through grants and generic services
• A higher profile for language technologies, thus fostering demand
• Public availability of data collected by ELRC
Solve your language problems!
Thank you!

Hvala lepa!