The CEF eTranslation platform @ work

Markus Foti
eTranslation Programme Manager
DGT R.3

ELRC Workshop, Bulgaria, 24 October 2018
eTranslation platform at a glance

**eTranslation**
- MT@EC
  - Launched June 2013
  - Legalese
  - Statistical (Moses)

**CEF.AT**
- Launched July 2017 (webservice for snippets), Nov. 2017 (web page for documents)
- Cloud based
- Neural engines
- More NLPs (transliteration, named entity recognition...)
- Generic services
- Projects (ELRC, market research)
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
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.
Get your translation back by e-mail.

Règles plus intelligentes pour des denrées plus sûres»: Paquet décisif pour moderniser, de simplifier et de renforcer l’agroalimentaire en Europe.

1) Ce qui est «de la fourche à la fourchette»

L’UE «de la fourche à la fourchette» : santé en faisant en sorte que la conduite finale sur notre alimentation législatif est un système infaillible, la protection par pesantez des maillons de la chaîne de production nuisibles, la protection dePaths de ces maladies et tous les dangers contenus de manière harmonieuse pour consommateurs.
One original, many translations...

English

Français
Deutsch
Italiano
Português
Polski
Lietuvių kalba
Ελληνικά
Български език
Svenska

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!
### Who is using eTranslation: DSIs

<table>
<thead>
<tr>
<th>DSI\BB</th>
<th>eTranslation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRIS</td>
<td>*</td>
</tr>
<tr>
<td>Cybersecurity</td>
<td></td>
</tr>
<tr>
<td>EESSI</td>
<td>*</td>
</tr>
<tr>
<td>Europeana</td>
<td>*</td>
</tr>
<tr>
<td>ODR</td>
<td>*</td>
</tr>
<tr>
<td>Open Data</td>
<td>*</td>
</tr>
<tr>
<td>Safer Internet</td>
<td></td>
</tr>
<tr>
<td>e-Justice Portal</td>
<td>*</td>
</tr>
<tr>
<td>eDelivery</td>
<td></td>
</tr>
</tbody>
</table>
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...

Dane Publiczne | April 2, 2017
PDF

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...}

Dane Publiczne | April 23, 2017
PDF

Geostatystyczny portal
Geostatystyczny portal is a modern tool for spatial presentation of the information resource in Poland, allowing the publication of aggregated statistical data in the form of...

Dane Publiczne | December 22, 2015

Foreign migration testing methodology in the national population and hou...
A population and housing census carried out in Poland in 2011 was put into practice new organisational
Machine-to-machine service
eTranslation on the European Data Portal

Commuting in Poland — Results of the NSP 2011

The title and description of this dataset are machine translated. Original language: Polish

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 2011). The methodology for this test is developed on the basis of previous work linked to the implementation of the "Migration population of employees in 2006." This time existing source of data for this subject (tax) was supplemented by resources other research. The information delivered by these sources was made through the performance of work by sex and age and, in particular, allow a more accurate identification of territorial cohesion, jobs, which in combination with the actual place of residence of those persons to steer flows.

Distributions

Commuting — NSP 2011 (PDF)

Tags

NSP 2011, dojazdy do pracy, spis powszechny

Additional Info
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!
Behind the scenes: how it works

- **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**
Neural Machine Translation (NMT)

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 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>
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>

24 June 2018
Starting in 1980, the European Union gradually adopted legislation putting an end to the diverging schedules of the national clock changes. In 2018 however, the purpose of clock changes has become much less relevant, with studies suggesting that energy savings are now marginal and citizens increasingly complaining about negative health impacts.

**MT@EC (SMT):**
От 1980 г. насам Европейският съюз постепенно прието законодателство, водещо до различни графици на национално часовите промени. През 2018 г. обаче целта на часовите промени има все по-малко уместно, с проучвания, които показват, че икономиите на енергия са незначителни и гражданите все по-често се оплакват от отрицателно въздействие върху здравето им.

**eTranslation (NMT):**
От 1980 г. насам Европейският съюз постепенно прие законодателни актове, с които се слага край на различните списъци с промените в националния часовник. През 2018 г. обаче целта на часовниците получи много по-малко значение, като проучванията сочат, че икономиите на енергия сега са незначителни и гражданите все повече се оплакват от отрицателните последици за здравето.
Visitor surge helps Bulgaria’s festivals to flourish
Bulgaria, with its diverse natural landscapes and historical sites going back to the Thracians, is starting to sell itself as a place for culture and scenery, shedding a reputation as a venue for Black Sea pub crawls.

MT@EC (SMT):
Нарастването на посетителите на България помага фестивалите процъфтяват
България, с неговите разнообразни природни и исторически обекти чрез връщане към траките, започва да продава само по себе си като място за култура и пейзажи, хвърля репутация като място за Черно море експониране на барове.

eTranslation (NMT):
Скокът на посетителите помага на българските празници да процъфтяват
България, с разнообразния си природен ландшафт и исторически обекти, които се връщат на Thracians, започва да се продава като място за култура и декори, което свързва репутацията като място на виадукти на Черно море.
If you ever venture on a Bulgarian holiday, remember a few things: Bulgarians love to catch up with family, welcome visitors, feed them, and talk to them about their lives and culture. A big part of this bonding happens over meals. Bulgarian food, traditional drinks and Bulgarian cheese all compliment the splendid hospitality and surroundings.

**MT@EC (SMT):**
Ако някога сте предприяли на български празник, не забравяйте няколко неща: Българите обичат да настигат семейства, посрещат посетители ще ги и си поговори с тях относно своя живот и култура. Голяма част от това свързване се осъществява във храна. Българска, традиционните напитки и българските сирена всички допълват чудесния гостоприемство и околностите.

**eTranslation (NMT):**
Ако някога сте предприяли на български празник, си спомняме за няколко неща: Българите обичат да настигат семействата и посрещат посетители, ще ги и си поговори с тях относно своя живот и култура. Голяма част от това свързване се осъществява във храната. Българска, традиционните напитки и българските сирена всички допълват чудесния гостоприемство и околностите.
Original Text (https://www.thebrokebackpacker.com/bulgarian-food-15-dishes-bucket-list/): Bulgarian food, traditional drinks and Bulgarian cheese all complement the splendid hospitality and surroundings.

But:

Bulgarian food, traditional drinks and Bulgarian cheese all complement the splendid hospitality and surroundings.

Becomes:

Българската храна, традиционните напитки и българското сирене допълват разнасяне на храна, гостоприемство и околна среда

And today:

В България храни, традиционни напитки и българско сирене са допълнение към гостоприемството и околностите.
The adapted vs. non-adapted scenario, i.e.

a) General domain vs. EU legislation domain

For as many CEF languages as possible we would need to come up with

- a text in the general domain, e.g. a news article, and
- an in-domain text, e.g. legislation

Show the MT@EC-generated English equivalent for each text

b) General domain vs. DSI-specific domain: for which languages are such engines available?

(if there is no domain adapted engine for the <country’s language-EN> pair, you/we may opt to show adapted and non-adapted engines in another language pair, leading to a call for data for the country’s language)
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!
Благодаря ти!