Engaging Content
Engaging People

A World Leading SFI Research Centre

Science Foundation Ireland sfi For what's next

MT Developments in the EU

Andy Way

School of Computing, Dublin City University, Dublin, Ireland

HOST INSTITUTION
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

HOST INSTITUTION
DCU Ollscoil Chathair Bhaile Átha Cliath Dublin City University

PARTNER INSTITUTIONS
TU Dublin
UCD Dublin
CIT Cork Institute of Technology
Institiúid Teicneolaíochta Chorcaí
AIT Connect & Discover
Maynooth University National University of Ireland Maynooth

FUNDED BY
Ireland's European Structural and Investment Funds Programmes 2014-2020
Co-funded by the Irish Government and the European Union
European Union
European Regional Development Fund

Overview of Talk

- EU and the Language Situation in Europe
- Language Technology Resources & eTranslation MT System
- EU Funding for MT and related projects
- Irish MT
- Upcoming Conferences of Interest (CEF eTranslation, EAMT)

The European Union today

- The EU now has three alphabets and 24 official languages.
- Some 60 other languages are currently spoken in particular regions or by specific groups.
- Immigration has also brought numerous additional languages to the EU.

The Official Languages of Europe

The EU has 24 official languages.

These are:

- Bulgarian
- Croatian
- Czech
- Danish
- Dutch
- English
- Estonian
- Finnish
- French
- German
- Greek
- Hungarian
- Irish
- Italian
- Latvian
- Lithuanian
- Maltese
- Polish
- Portuguese
- Romanian
- Slovak
- Slovenian
- Spanish
- Swedish

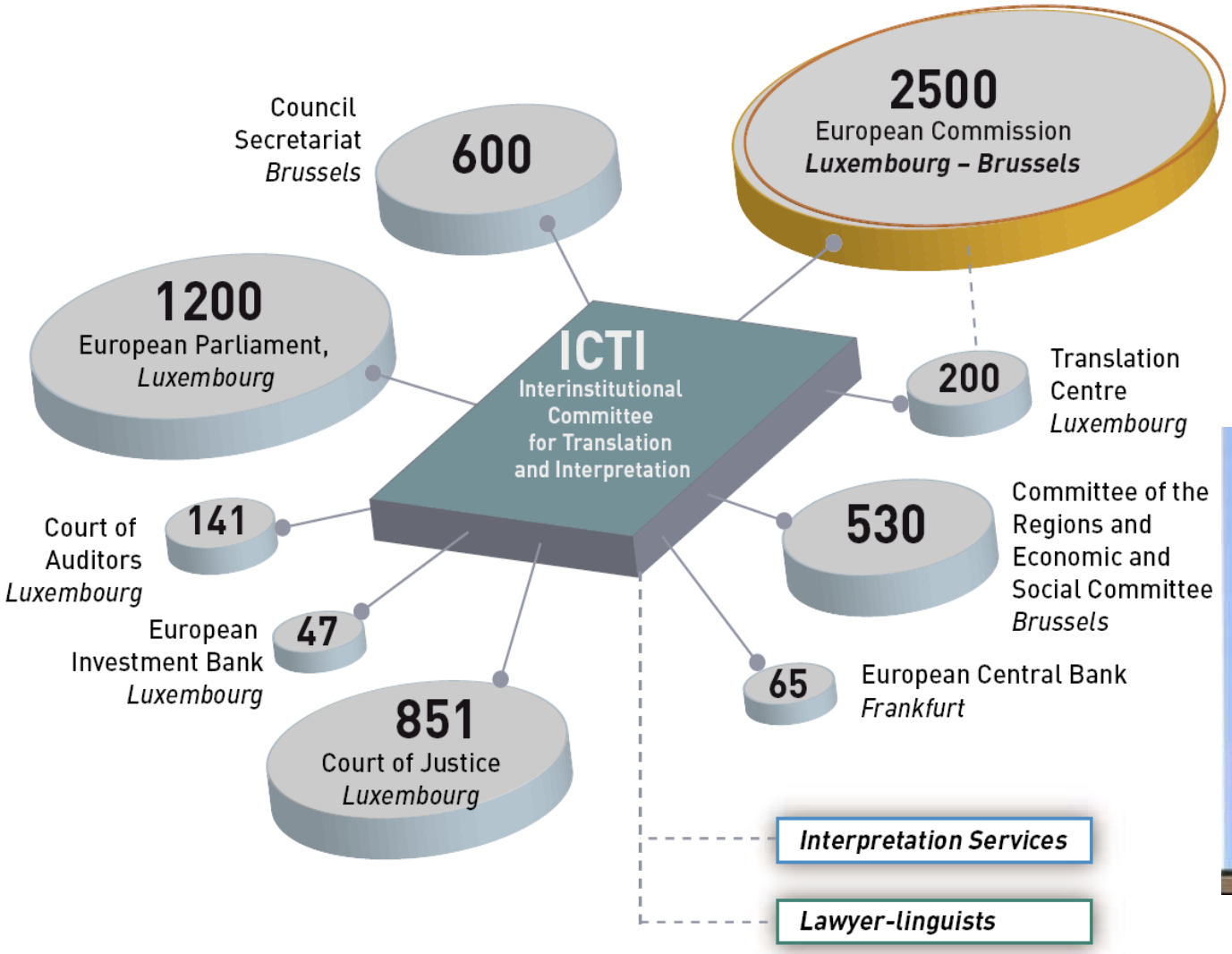
We'll come back to this later on, but Irish is an official EU language, but not (yet!) a working language of the EU …

EU Indigenous, Regional or Minority Languages

https://op.europa.eu/s/oklx

Languages covered in this guide
Albanian, Armenian, Azeri, Basque, Belarusian, Bosnian, Breton, Bulgarian, **Castilian Spanish, Catalan, Croatian, Czech,** Danish, Dutch, English, Estonian, Faroese, Finnish, French, **Galician, Georgian, German, Greek,** Hungarian, Icelandic, Irish, Italian, Latvian, Luxembourgish, Lithuanian, **Macedonian, Maltese, Montenegrin, Norwegian, Polish, Portuguese,** Romani, Romanian, Russian, **Sámi, Scots Gaelic, Serbian, Slovak, Slovenian, Swedish,** Turkish, Ukrainian, **Welsh,** Yiddish

EU Translation Services



(slide 9)

Interinstitutional cooperation on translation and interpretation

Euramis DGT Translation Memory

Engaging Content
Engaging People

Number of Translation Units (TUs)

	DGT-TM 2007	DGT-TM 2011	DGT-TM 2012	DGT-TM 2013	DGT-TM 2014	DGT-TM 2015	DGT-TM 2016	DGT-TM 2017	DGT-TM 2018	DGT-TM 2019	Total Updates	Total incl. 2007
BG	708.658	450.943	272.595	378.416	480.896	203.616	480.328	612.186	292.293	436.945	3.608.218	4.316.876
CS	890.025	1.981.222	283.826	478.709	499.949	203.860	476.681	603.587	286.220	427.650	5.241.704	6.131.729
DA	433.871	1.993.718	279.746	472.024	499.107	201.723	457.623	601.809	293.424	425.084	5.224.258	5.658.129
DE	532.668	1.918.637	284.072	472.081	499.844	202.634	472.385	586.631	300.062	424.610	5.160.956	5.693.624
EL	371.039	1.897.576	285.483	462.304	509.080	203.674	479.260	589.759	285.862	431.764	5.144.762	5.515.801
EN	**2.187.504**	**2.282.281**	**322.377**	**538.949**	**571.548**	**234.085**	**561.582**	**723.950**	**345.245**	**504.611**	6.084.628	**8.272.132**
ES	509.054	1.903.748	284.977	477.829	501.814	203.214	479.208	611.286	295.844	429.876	5.187.796	5.696.850
ET	1.047.503	1.863.860	280.549	461.051	504.702	203.988	476.735	616.121	294.630	433.764	5.135.400	6.182.903
FI	514.868	1.877.643	283.213	459.927	505.202	200.655	476.263	607.936	294.493	433.476	5.138.808	5.653.676
FR	1.106.442	1.849.953	273.961	462.431	485.652	200.361	454.271	603.848	279.587	419.805	5.029.869	6.136.311
GA	0	0	2.848	0	41.461	7.673	9.127	37.181	30.014	53.652	181.956	181.956
HR	0	0	0	0	276.502	197.626	473.611	609.422	300.838	430.147	2.288.146	2.288.146
HU	1.159.975	1.865.343	284.282	480.050	499.624	203.608	475.701	611.267	283.723	428.881	5.132.479	6.292.454
IT	542.873	1.922.649	281.503	474.035	494.972	202.048	461.164	604.173	293.687	425.466	5.159.697	5.702.570
LT	1.126.255	1.863.259	286.018	461.375	478.942	204.190	478.441	613.818	287.074	433.546	5.106.663	6.232.918
LV	1.120.835	1.855.851	284.641	461.190	499.579	204.715	478.837	605.802	292.894	434.344	5.117.853	6.238.688
MT	1.021.855	457.946	263.804	386.677	462.515	203.420	461.298	597.847	298.275	413.131	3.544.913	4.566.768
NL	502.557	1.910.694	281.683	469.990	489.002	202.243	468.909	583.313	281.903	419.242	5.106.979	5.609.536
PL	1.052.136	1.875.561	282.551	454.864	496.168	203.474	482.659	617.630	294.173	434.101	5.141.181	6.193.317
PT	945.203	1.918.683	284.310	471.810	478.851	203.263	475.659	608.208	293.161	429.301	5.163.246	6.108.449
RO	650.735	466.410	270.763	393.398	469.845	202.707	469.971	606.717	284.331	438.032	3.602.174	4.252.909
SK	1.065.399	1.890.726	285.422	480.073	494.713	203.978	478.940	615.277	285.995	436.152	5.171.276	6.236.675
SL	1.026.668	1.899.515	284.642	479.147	494.689	204.342	477.041	610.746	293.833	435.384	5.179.339	6.206.007
SV	555.362	1.931.052	283.589	478.204	496.390	202.700	475.563	608.798	300.625	426.882	5.203.803	5.759.165
Total	19.071.485	37.877.270	6.226.855	10.154.534	11.231.047	4.703.797	10.981.257	14.087.312	6.788.186	10.005.846	112.056.104	131.127.589

https://wt-public.emm4u.eu/Resources/DGT-TM_Statistics.pdf (30th March, 2019)

Other EU Language Technology Resources

- JRC-Acquis: a multilingual parallel corpus
- DGT-Acquis: multilingual parallel corpora from the EU Official Journal
- DCEP: Digital Corpus of the European Parliament
- DGT Translation Memory

All downloadable from:

https://ec.europa.eu/jrc/en/language-technologies

EUROPEAN COMMISSION (EC) &
EUROPEAN PARLIAMENT (EP) &
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL (ECDC)

Manuscript. The final publication is available at
http://link.springer.com/article/10.1007%2Fs10579-014-9277-0

An overview of the European Union's highly multilingual parallel corpora

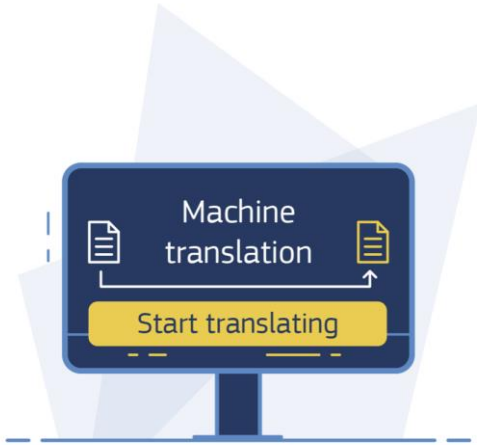
Corresponding author: Ralf Steinberger
Ralf.Steinberger@jrc.ec.europa.eu
Tel: +39 0332 786271
Fax: +39 0332 785154

Ralf Steinberger, Mohamed Ebrahim, Alexandros Poulis, Manuel Carrasco-Benitez, Patrick Schlüter, Marek Przybyszewski & Signe Gilbro

2/11/2014

Ralf Steinberger, European Commission – Joint Research Centre (JRC), Ispra, Italy, Ralf.Steinberger@jrc.ec.europa.eu
Mohamed Ebrahim, Cognizant-SetCon GmbH, Munich, Germany, Mohamed@alqamus.de (previously European Commission – Joint Research Centre)
Alexandros Poulis, Lionbridge, Finland, Alexandros.Poulis@lionbridge.com (previously European Parliament – Directorate General for Translation)
Manuel Carrasco-Benitez, European Commission – Directorate General for Translation (DGT), Luxembourg, Manuel.Carrasco-Benitez@ec.europa.eu
Patrick Schlüter, European Commission – Directorate General for Translation (DGT), Luxembourg, Patrick.Schlüter@ec.europa.eu
Marek Przbyszewski, European Commission – Directorate General Education And Culture (EAC), Brussels, Belgium, Marek.PRZYBYSZEWSKI@ec.europa.eu
Signe Gilbro, European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden, info@ecdc.europa.eu

eTranslation MT System



Translate to and from 27 languages!

Bulgarian (BG)	French (FR)	Italian (IT)	Romanian (RO)
Croatian (HR)	Finnish (FI)	Latvian (LV)	Russian (RU)
Czech (CS)	German (DE)	Lithuanian (LT)	Spanish (ES)
Danish (DA)	Greek (EL)	Maltese (MT)	Slovak (SK)
Dutch (NL)	Hungarian (HU)	Norwegian Bokmål (NB)	Slovenian (SL)
English (EN)	Icelandic (IS)	Polish (PL)	Swedish (SV)
Estonian (ET)	Irish (GA)	Portuguese (PT)	

Secure and confidential

eTranslation follows security best practices and assures the confidentiality and security of your data. It is used by EU and national administrations.

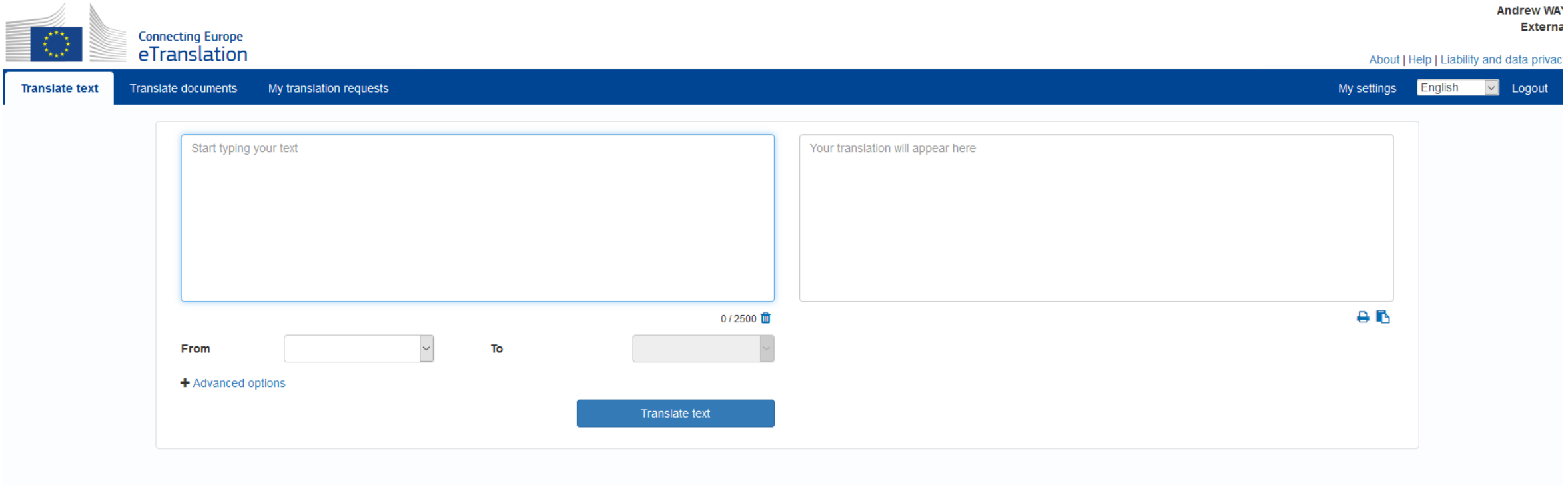
Free and easy to use

eTranslation is free of charge. Save time when translating your documents between any two official EU languages and more!

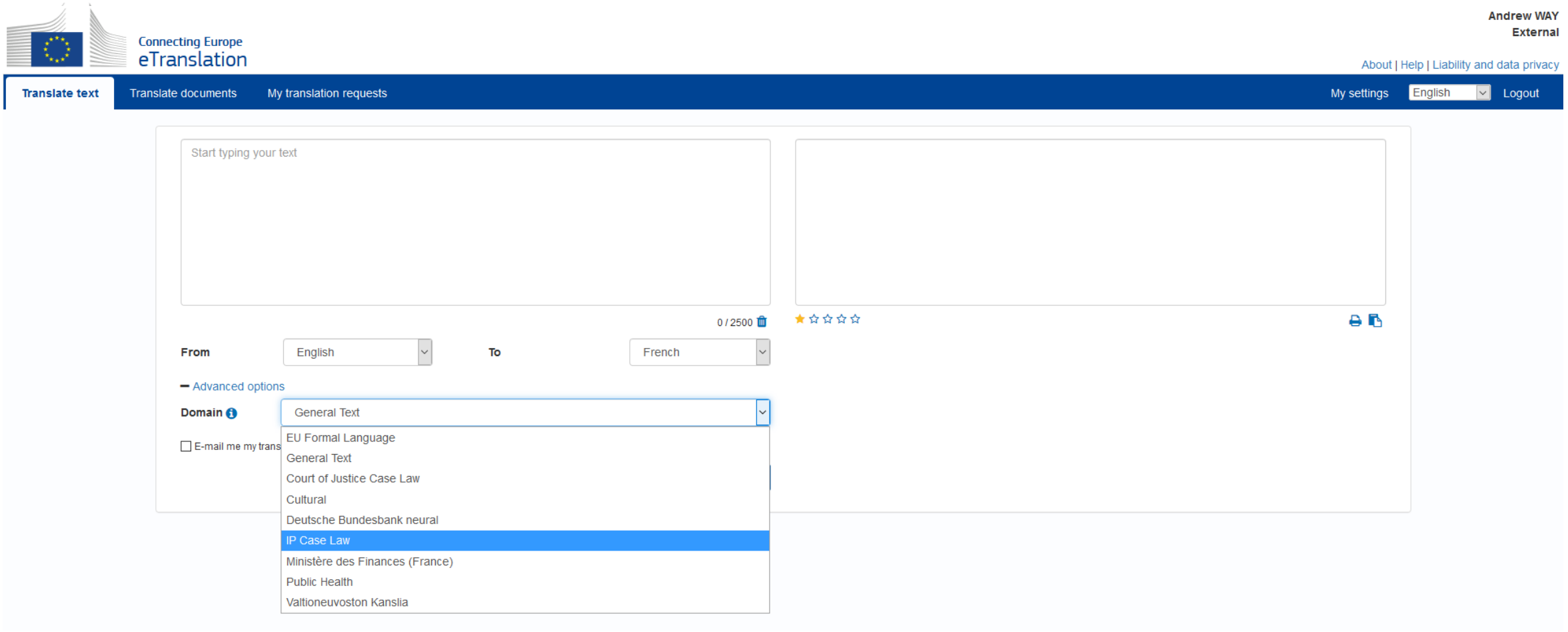
High quality

The eTranslation service uses state-of-the-art neural machine translation to translate formatted documents and plain text between any of the 24 official European Union languages, as well as Icelandic, Norwegian and Russian.

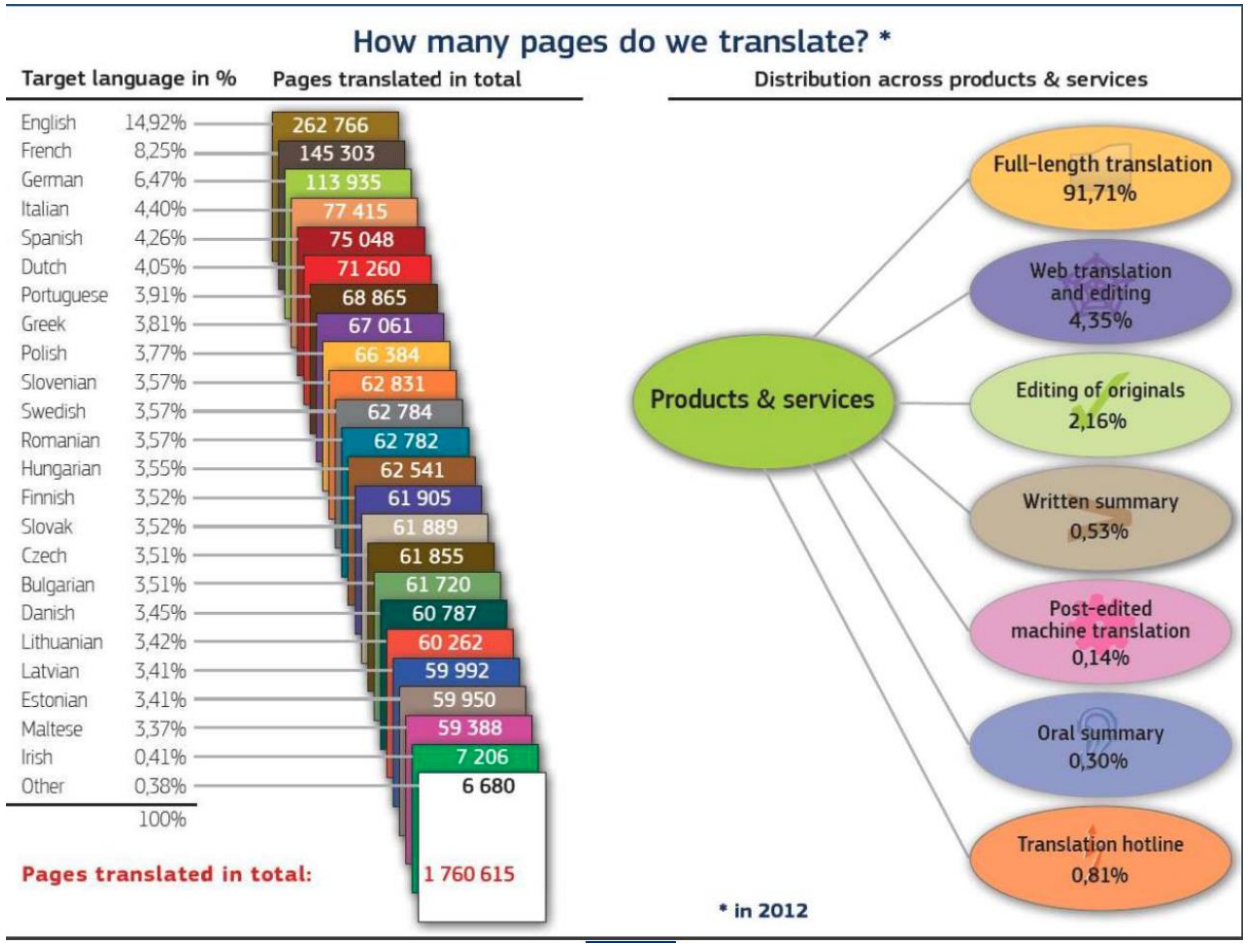
eTranslation Interface



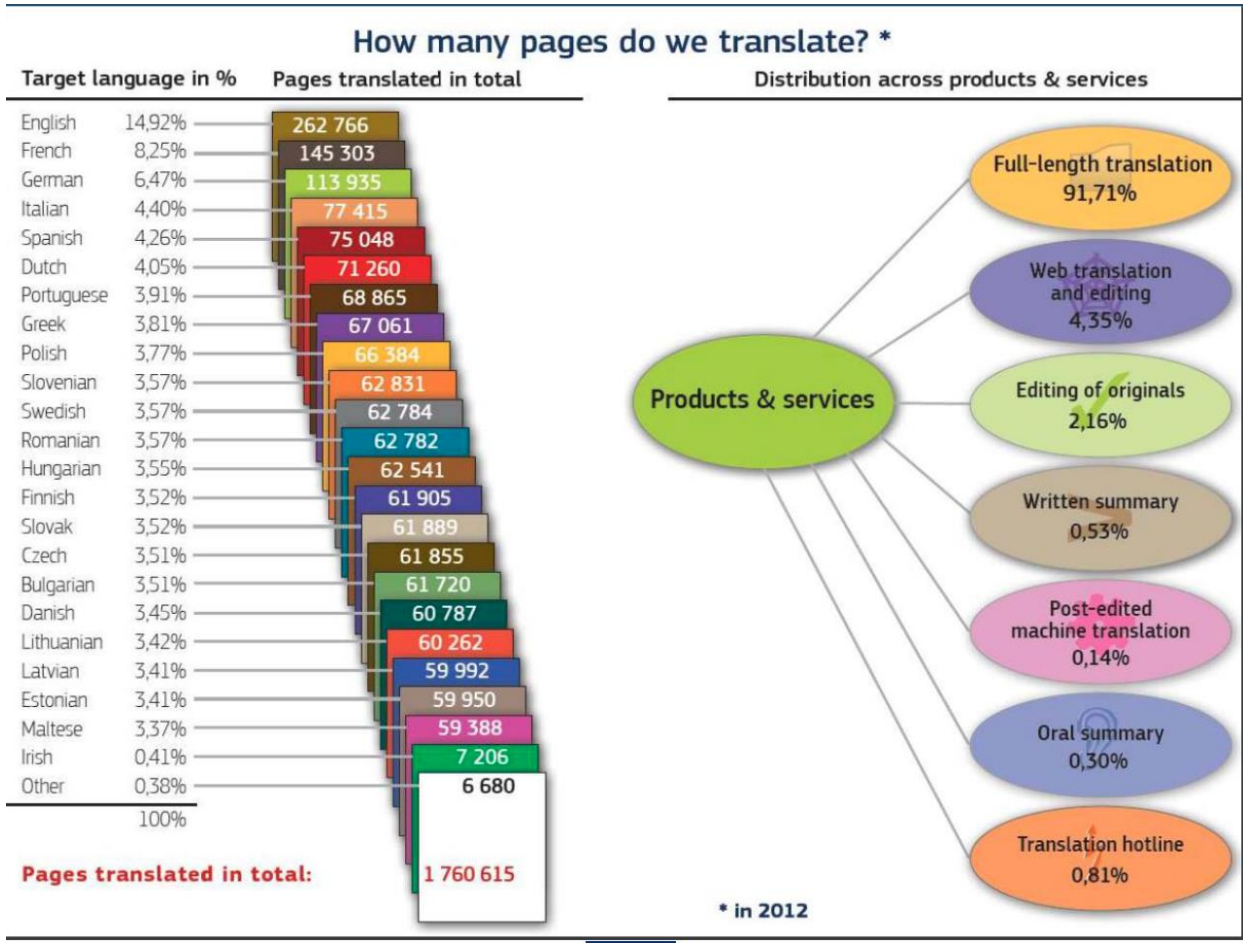
eTranslation Interface



eTranslation Volume and Use-Cases



eTranslation Volume and Use-Cases



(Some) Non-CEF-funded H2020 EU-funded Projects

EUROPEAN LANGUAGE GRID

European Language Grid

Support the Multilingual Digital Single Market.

- **Supporting the Multilingual Digital Single Market**
- **Growing a vibrant European LT community**
- **Helping coordinate all European LT activities**
- **Creating a powerful, scalable LT platform using a grid architecture**

https://www.european-language-grid.eu/

(Some) Non-CEF-funded H2020 EU-funded Projects

Engaging Content
Engaging People

Bergamot

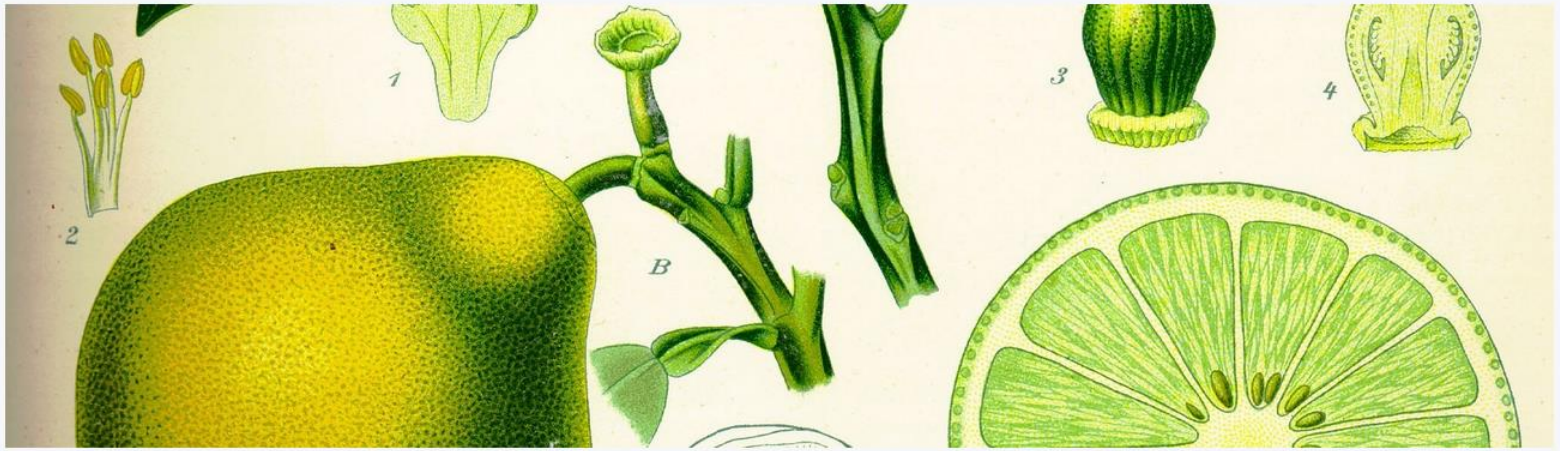
The Bergamot project will add and improve client-side machine translation in a web browser.

Consortium Partners
News & Events
Publications
Deliverables
Software
Data
Contact
Related Projects

© 2020 Bergamot
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825303

Bergamot

https://browser.mt/



The Bergamot Project

The Bergamot project will add and improve client-side machine translation in a web browser.

Unlike current cloud-based options, running directly on users' machines empowers citizens to preserve their privacy and increases the uptake of language technologies in Europe in various sectors that require confidentiality. Free software integrated with an open-source web browser, such as Mozilla Firefox, will enable bottom-up adoption by non-experts, resulting in cost savings for private and public sector users who would otherwise procure translation or operate monolingually.

Bergamot is a consortium coordinated by the University of Edinburgh with partners Charles University in Prague, the University of Sheffield, University of Tartu, and Mozilla.

(Some) Non-CEF-funded H2020 EU-funded Projects

elitr.eu

Homepage | The Project | Team | Results

The Project

What is ELITR

The project "European Live Translator" is supported by the EU under the Horizon 2020 programme, and coordinated by Ondřej Bojar, Charles University, Prague.

It runs from January 1st, **2019** to December 31st, **2021**.

The partners:

- Charles University, Prague
- University of Edinburgh, Scotland
- Karlsruhe Institute of Technology, Germany
- PerVoice, Italy
- alfatraining, Germany
- the affiliated party in this project is Supreme Audit Office, the Czech Republic

The goal of the ELITR project is to remove the language barrier in communication within and among European citizens, companies, institutes and organizations at large assemblies like conferences, in smaller live discussions like workshops and in discussions held over long distance like formal or informal online meetings. Within the project, we will create automatic subtitling system of live meetings and conference presentations and provide the system of spoken language translation (interpreting). Furthermore, the project will design and implement automatic minuting, i.e. create structured summaries from automatic transcript of a discussion. We also aim to advance the state of the art to deliver machine translation of high-quality at the level of whole documents and to scale up to multilingual settings with dozens of source and/or target languages. The technologies will be tested by the Supreme Audit Office of the Czech Republic and alfaview®, a German online conferencing system.

Keywords:

- Machine translation
- Multilingualism, language diversity
- Automated translation
- Natural language processing
- Service oriented architectures

Objectives

ASR: Automatic Speech Recognition

SLT: Spoken Language Translation

MT: Machine Translation

AM: Automatic Minuting

Funding:

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825460.

ERC-funded H2020 Projects

Multi-modal Context Modelling for Machine Translation

Fact Sheet | Results

Objective

Automatically translating human language has been a long sought-after goal in the field of Natural Language Processing (NLP). Machine Translation (MT) can significantly lower communication barriers, with enormous potential for positive social and economic impact. The dominant paradigm is Statistical Machine Translation (SMT), which learns to translate from human-translated examples.

Human translators have access to a number of contextual cues beyond the actual segment to translate when performing translation, for example images associated with the text and related documents. SMT systems, however, completely disregard any form of non-textual context and make little or no reference to wider surrounding textual content. This results in translations that miss relevant information or convey incorrect meaning. Such issues drastically affect reading comprehension and may make translations useless. This is especially critical for user-generated content such as social media posts -- which are often short and contain non-standard language -- but applies to a wide range of text types.

The novel and ambitious idea in this proposal is to devise methods and algorithms to exploit global multi-modal information for context modelling in SMT. This will require a significantly disruptive approach with new ways to acquire multilingual multi-modal representations, and new machine learning and inference algorithms that can process rich context models. The focus will be on three context types: global textual content from the document and related texts, visual cues from images and metadata including topic, date, author, source. As test beds, two challenging user-generated datasets will be used: Twitter posts and product reviews.

This highly interdisciplinary research proposal draws expertise from NLP, Computer Vision and Machine Learning and claims that appropriate modelling of multi-modal context is key to achieve a new breakthrough in SMT, regardless of language pair and text type.

Project Information

MultiMT
Grant agreement ID: 678017
Status
Ongoing project
Start date: 1 July 2016
End date: 30 June 2021

Funded under
H2020-EU.1.1.

Overall budget
€ 1 493 771

EU contribution
€ 1 493 771

Hosted by
IMPERIAL COLLEGE OF SCIENCE TECHNOLOGY AND MEDICINE
United Kingdom

ERC-funded H2020 Projects

DeepSPIN: Deep Structured Prediction in NLP

About | News | Publications | Team

About

DeepSPIN ("Deep Structured Prediction in Natural Language Processing") is a research project funded by the European Research Council (ERC) and coordinated by André Martins. It runs 2018-2023 within Instituto de Telecomunicações at Instituto Superior Técnico, and Unbabel, in Lisbon, Portugal.

Deep learning is revolutionizing the field of **Natural Language Processing** (NLP), with breakthroughs in machine translation, speech recognition, and question answering. New language interfaces (digital assistants, messenger apps, customer service bots) are emerging as the next technologies for seamless, multilingual communication among humans and machines.

From a machine learning perspective, many problems in NLP can be characterized as **structured prediction**: they involve predicting structurally rich and interdependent output variables. In spite of this, current neural NLP systems ignore the structural complexity of human language, relying on simplistic and error-prone greedy search procedures. This leads to serious mistakes in machine translation, such as words being dropped or named entities mistranslated. More broadly, neural networks are missing the key structural mechanisms for solving complex real-world tasks requiring deep reasoning.

This project attacks these fundamental problems by bringing together deep learning and structured prediction, with a highly disruptive and cross-disciplinary approach. First, we will endow neural networks with a **planning mechanism** to guide structural search, letting decoders learn the optimal order by which they should operate. This makes a bridge with reinforcement learning and combinatorial optimization. Second, we will develop new ways of automatically **inducing latent structure** inside the network, making it more expressive, scalable and interpretable. Synergies with probabilistic inference and sparse modeling techniques will be exploited. To complement these two innovations, we will investigate new ways of incorporating **weak supervision** to reduce the need for labeled data.

Three highly challenging applications will serve as testbeds: **machine translation**, **quality estimation**, and **dependency parsing**. To maximize technological impact, this is done with collaboration with Unbabel, a start-up company in the crowd-sourcing translation industry.

ERC-funded H2020 Projects

Marta R. Costa-jussà

Home | Experience | Publications | Teaching | Projects | Media | Blog | Events

Home

Marta R. Costa-jussà is a Ramon y Cajal Researcher at the Universitat Politècnica de Catalunya (UPC, Barcelona). She received her PhD from the UPC in 2008. Her research experience is mainly in Machine Translation. She has worked at LIMSI-CNRS (Paris), Barcelona Media Innovation Center, Universidade de São Paulo, Institute for Infocomm Research (Singapore), Instituto Politécnico Nacional (Mexico) and the University of Edinburgh. She has participated in 18 European (including a MarieCurie Action) and Spanish national projects. She has organised 10 workshops in top venues and published more than 100 papers. She has been part of the Editorial Board of the Computer Speech and Language journal. She regularly cooperates with companies as a scientific consultant. Currently, she is leading the Spanish Project of AMALEU and AdaVoice. Recently, she has received an ERC Starting Grant 2020 and two Google Faculty Research Awards (2018 and 2019).

https://orcid.org/0000-0002-5703-520X

Google Scholar

LUNAR: Lifelong UNiversal lAnguage Representation

(Some) Marie Curie-funded H2020 Projects

HORIZON 2020

Machine Translation Impact on Language Learning

Fact Sheet

Project description

DE EN ES FR IT PL

Machines break language barriers

An easy and fast way to translate a text is to have a machine do the job online - instantly. Machine translation (MT) systems use machine-learning technologies to translate hundreds of billions of words online daily. The EU-funded MTrill project will examine the impact of MT systems on the learning and processing of English as a second language. As regards language acquisition, the research will focus on language binding – the ability to combine single words properly in a sentence. The project will conduct a syntactic priming experiment. Here, participants will be tested on whether they produce the same syntactic structures that had been previously seen during an MT task.

Show the project objective

Field of science

/social sciences/psychology/psycholinguistics
/humanities/languages and literature/languages - general

Project Information

MTrill
Grant agreement ID: 843455
Status
Ongoing project
Start date 25 April 2019
End date 24 April 2021

Funded under
H2020-EU.1.3.2.

Overall budget
€ 184 590,72

EU contribution
€ 184 590,72

Coordinated by
DUBLIN CITY UNIVERSITY
Ireland

(Some) Marie Curie-funded H2020 Projects

HORIZON 2020

Semantics-Based Machine Translation

Fact Sheet

Project description

DE EN ES FR IT PL

Machine translation based on semantics

Machine translation systems could be a cost-effective alternative to human translators in a variety of situations but there are serious flaws in existing systems. Some take semantics partially into account, others tend to focus on the fluency of the translated text rather that the adequacy of the read. The EU-funded SEBAMAT project aims to provide state of the art translation that considers word senses rather than words only. The project also strives to use role labelling for identifying the semantic roles of the words in a sentence.

Show the project objective

Field of science

/humanities/languages and literature/languages - general

Programme(s)

H2020-EU.1.3.2. - Nurturing excellence by means of cross-border and cross-sector mobility

Project Information

SEBAMAT

Grant agreement ID: 844951

Status
Ongoing project

Start date
1 April 2020

End date
31 March 2022

Funded under
H2020-EU.1.3.2.

Overall budget
€ 165 085,44

EU contribution
€ 165 085,44

Coordinated by
ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS

Greece

EU CEF Funding and MT-Oriented Projects

EU CEF Funding and MT-Oriented Projects

eTranslation

2017-EU-IA-0178
Broader Web-Scale Provision of Parallel Corpora for European Languages

2018-EU-IA-0049
CEF Data Marketplace

2018-EU-IA-0050
PRINCIPLE

2018-EU-IA-0051
NTEU (Neural Translation for the European Union)

2018-EU-IA-0052
OCCAM

2018-EU-IA-0061
EuroPat: Unleashing European Patent Translations

2018-EU-IA-0063
Continued Web-Scale Provision of Parallel Corpora for European Languages

2018-EU-IA-0079
Translation Automation Services for EU Council Presidency

2019-EU-IA-0013
Multilingual Anonymisation toolkit for Public Administrations

2019-EU-IA-0015
CEFAT4Cities

2019-EU-IA-0031
Unsupervised MT for Low-resourced language pairs (MT4All)

2019-EU-IA-0034
Curated Multilingual Language Resources for CEF AT

2019-EU-IA-0045
User-focused Marian

2019-EU-IA-0049
Federated eTranslation TermBank Network

EU CEF Funding and MT-Oriented Projects

CEF programme

CONNECTING EUROPE FACILITY

- CEF ENERGY
- CEF TELECOM
  - CALLS
  - HOW TO APPLY
  - BENEFICIARIES' INFO POINT
  - CEF TELECOM PROJECTS
  - SIGN UP AS AN EXPERT
  - WIFI4EU
- CEF TRANSPORT

OCCAM

Programme:
CEF Telecom

Call year:
2018

Location of the Action:
Belgium, Czech Republic

Implementation schedule:
October 2019 to September 2021

Maximum EU contribution:
€973,894

Total eligible costs:
€1,298,525

Percentage of EU support:
75%

Coordinator:
Vysoke uceni technicke v Brne (Brno University of Technology) (Czech Republic)
https://www.vutbr.cz/en/

Status:
Ongoing

DSI:
eTranslation

Additional information:
Digital Single Market (DSM) strategy
http://ec.europa.eu/priorities/digital-single-market

2018-EU-IA-0052

The objective of the Action (OCCAM - OCR, ClassificAtion & Machine Translation) is to focus on the integration and extension of CEF Automated Translation DSI into multilingual digital cross-border services.

The Action aims at the integration of image classification, Translation Memories (TMs),Optical Character Recognition (OCR), and Machine Translation (MT) to support the Automated Translation of scanned documents – a document type that currently cannot be processed by the CEF eTranslation service. Scanned forms typically contain template text, for which translations are available in TMs. The Action aims at recognising such text from scanned images, and translating it automatically. By filling this gap, OCCAM will directly contribute to lowering the language barrier.

The languages to be covered by OCCAM will develop two use cases: (a) Business Registers Interconnection System (BRIS) use case focusing on Belgian and Czech Business Registers and (b) Digital Humanities use case. For use case (a), OCCAM will develop an eDelivery-compliant Reference Implementation for the BRIS DSI. For use case (b), OCCAM will provide extensive guidance for introducing the combination of OCR and MT into the day-to-day workflow of Digital Humanities researchers. Both use cases will support the OCCAM Action in becoming sustainable in the long term.

EU CEF Funding and MT-Oriented Projects

CEF programme

CONNECTING EUROPE FACILITY

- CEF ENERGY
- CEF TELECOM
  - CALLS
  - HOW TO APPLY
  - BENEFICIARIES' INFO POINT
  - CEF TELECOM PROJECTS
  - SIGN UP AS AN EXPERT
  - WIFI4EU
- CEF TRANSPORT

Curated Multilingual Language Resources for CEF AT

Programme:
CEF Telecom

Call year:
2019

Location of the Action:
Bulgaria, Croatia, Hungary, Poland, Romania, Slovakia, Slovenia

Implementation schedule:
June 2020 to May 2022

Maximum EU contribution:
€720,000

Total eligible costs:
€960,000

Percentage of EU support:
75%

Coordinator:
Research Institute for Linguistics (Hungary)

Status:
Ongoing

DSI:
eTranslation

Additional information:
Digital Single Market (DSM) strategy
http://ec.europa.eu/priorities/digital-single-market

2019-EU-IA-0034

The Action aims to compile curated datasets in seven languages (Bulgarian, Croatian, Hungarian, Polish, Romanian, Slovak and Slovenian) in domains of relevance to the European Digital Service Infrastructures (DSIs) with a view to enhance the Automated Translation.

The prime source of data is stemming from national corpora of the above-mentioned languages and will cover domains relevant for CEF DSIs, such as eHealth and Europeana as well as for eGovernment. The Action will deliver at least 14 Million sentences (estimated to contain at least 140 Million words) from domains including culture, education, health and science. Moreover, the Action will contribute to address the gap in machine translation technology, which crucially depends on the provision of domain specific quality language resources for the under-resourced languages.

By delivering seven large size monolingual datasets, which themselves will facilitate the improvement of the CEF Automated Translation core service platform, the Action will enable any interested user to access information about the relevant EU Member States, including information about local companies and investment opportunities. Thus, the Action will also support the economic growth in Europe, by supporting the CEF AT core service platform for exchanging information in multiple languages.

EU CEF Funding and MT-Oriented Projects

CEF programme

CONNECTING EUROPE FACILITY

- CEF ENERGY
- CEF TELECOM
  - CALLS
  - HOW TO APPLY
  - BENEFICIARIES' INFO POINT
  - CEF TELECOM PROJECTS
  - SIGN UP AS AN EXPERT
  - WIFI4EU
- CEF TRANSPORT

European Language Resource Infrastructure (ELRI)

Programme:
CEF Telecom

Call year:
2016

Location of the Action:
France, Ireland, Portugal, Spain

Implementation schedule:
October 2017 to September 2019

Maximum EU contribution:
€1,173,610

Coordinator:
Fundación centro de tecnologías de interacción visual y comunicaciones (Spain)
http://www.vicomtech.org/

Status:
Closed

DSI:
eTranslation

Additional information:
Digital Single Market (DSM) strategy
http://ec.europa.eu/priorities/digital-single-market

2016-EU-IA-0111

The main objective of ELRI (European Language Resource Infrastructure) Action to facilitate the provision of an infrastructure to help collect, prepare and share language resources that can in turn improve translation services has been achieved. In particular, resources shared with the European Commission's Directorate-General for Translation (DGT) contribute to improve the EU automated translation services that are freely available to all public institutions.

The initiative notably addressed current issues related to sharing resources directly at the European level or beyond, by providing national public relay stations where resources remain under member states' laws and regulations until further clearance is negotiated and granted.

ELRI targeted resources that are relevant to Digital Service Infrastructures and involved public institutions and public translation centres in France, Ireland, Portugal and Spain as a first step, with a future extension to additional member states as a key objective.

Collecting & Curating Irish Data for MT Training

Engaging Content
Engaging People

PRINCIPLE

1. Language Resources (LRs) are collected from data holders and 'early adopters' from each of the 4 countries involved (in the specific domains of eJustice and eProcurement)
2. Machine Translation (MT) systems are produced from these LRs and evaluated to ensure high quality output
3. The MT systems are provided to 'early adopters' for free for the duration of the project
4. 'Early adopters' use the MT systems and provide feedback
5. Based on MT output evaluation and early adopter feedback, high quality LRs are identified
6. Parallel LRs of high quality are uploaded to the ELRC-SHARE portal in order to improve the automated translation system eTranslation.

DCU Ollscoil Chathair Bhaile Átha Cliath Dublin City University

LANGUAGE BANK National Library of Norway

UNIVERSITY OF ICELAND SCHOOL OF HUMANITIES

Iconic Part of the RWS Group

Filozofski fakultet Sveučilišta u Zagrebu

Consortium Members: *Dublin City University (Project Coordinator), University of Iceland, Faculty of Humanities and Social Sciences, University of Zagreb, National Library of Norway, Iconic Translation Machines Ltd.*

The Situation of the Irish Language in the EU

- Irish is an official language of the EU, but not a *working* language
- Ireland joined the EU in 1973, but its 'first national language', Irish, was only recognised as an official language in 2005
- This recognition was accompanied by a 5-year derogation regarding the use of the Irish language within the EU institutions.
- That derogation was extended for two subsequent five-year periods.
- The derogation expires on 31st December 2021, at which time Irish will become a full working language.
- All legislation enacted from that date will be translated into Irish as is already the case for the EU's other 23 official languages.

Can we solve the problem using human translation?

European Commission seeking 50 Irish language translators

Competition announced ahead of Irish achieving full official working status in 2022.

Thu, Feb 13, 2020, 17:54

The deadline for applications to the post of Irish translator to the EU is St Patrick's Day, 2020.

The European Commission is seeking to employ 50 translators based in Ireland and Brussels as it prepares for Irish to achieve full official working status the EU in 2022.

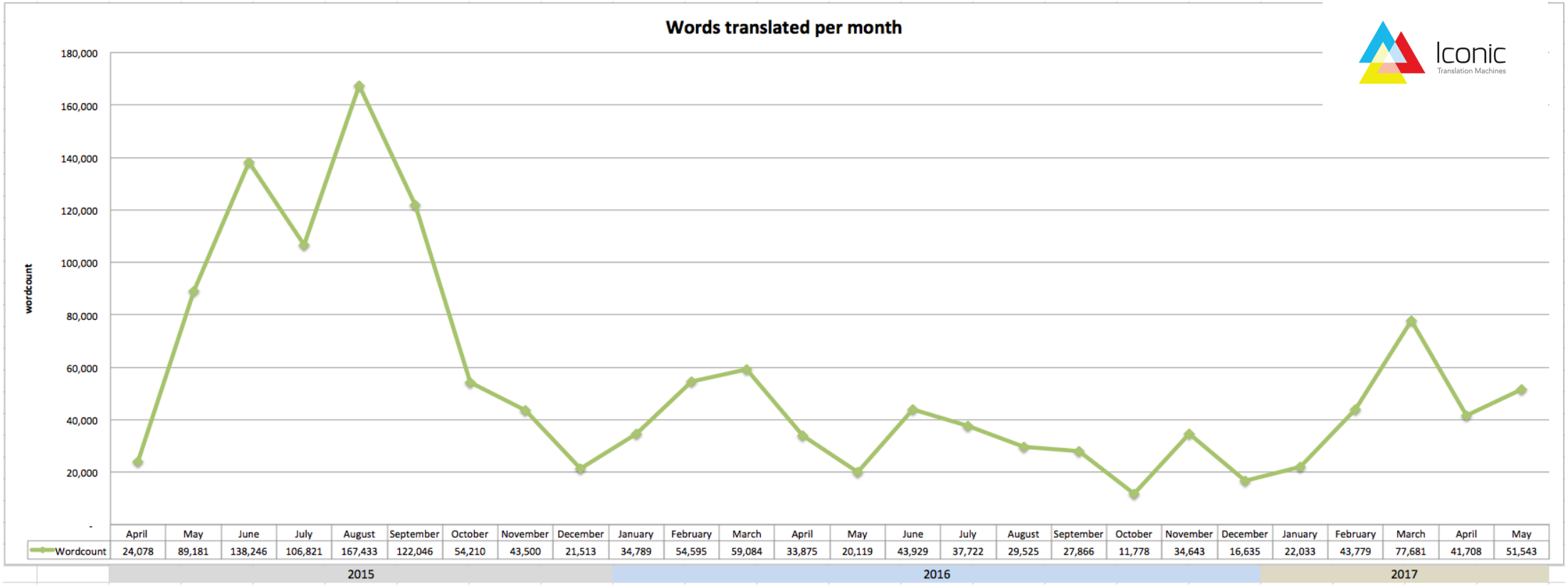
A new competition announced on Thursday aims to create a panel from which the Commission can employ translators to work in the Irish Language Unit of the European Commission's Directorate General for Translation at offices in Brussels and in Co Meath.

Applicants must be EU citizens, with "a perfect knowledge" of Irish, a thorough knowledge of English and a university degree of at least three years in any discipline.

It is the first time that translators will have the opportunity to work from European Commission offices based in Ireland.

- Pool of fluent Irish speakers is small: Ireland's 2016 census showed that only about 74,000 people speak Irish on a daily basis, just 1.7% of the total population of 4.75 million.
- Previous hiring campaigns have had limited success. In 2016, the European Personnel Selection Office announced 62 openings for Irish translators, but ultimately found only 10 suitable linguists.
- Less than a year later, in 2017, the European Parliament struggled to fill 23 of 26 open posts for Irish linguists.
- Foras na Gaelige offers a seal of accreditation for Irish translators. Their list of accredited Irish translators consists of 209 individuals who are "available on a full-time basis."

Partially automated solution? DCU's Tapadóir System



	April	May	June	July	August	September	October	November	December	January	February	March	April	May	June	July	August	September	October	November	December	January	February	March	April	May
Wordcount	24,078	89,181	138,246	106,821	167,433	122,046	54,210	43,500	21,513	34,789	54,595	59,084	33,875	20,119	43,929	37,722	29,525	27,866	11,778	34,643	16,635	22,033	43,779	77,681	41,708	51,543

- Internal trials in DAHG indicated output was useable → full roll-out
- Productivity increasing; no detrimental effect on ***any*** translators studied
- Integrated seamlessly with existing translation workflow: >**1.5 million words** so far

Collaboration with DGT and eTranslation

- *Oct 2019*: Visit from Director-General for Translation in the European Commission, Mr. Rytis Martikonis
- DCU's MA in Translation Studies and MSc in Translation Technology
- Avenues to avail of opportunities in the EU after Brexit
- Providing support to Irish language translators and interpreters in preparation for the end of the Irish language derogation in the EU from 2022.

- *Sept 2020*: discussion between eTranslation designers and DCU Irish MT team regarding data sharing and R&D collaboration

Other Significant Players

Dia daoibh! Tá Gaeilge againn!

Posted on January 23, 2020 by Microsoft Translator

Microsoft Translator adds Irish as a new language.

Our ongoing mission to break down language barriers continues with Irish: Today, we have added Irish Gaelic to Microsoft Translator. Irish Gaelic, usually referred to as the *Irish Language* or just *Irish*, and commonly known in Irish itself as Gaeilge (pronounced "gwael-guh"), is the latest addition to the Microsoft Translator family of languages. This brings Irish to all scenarios powered by Microsoft Translator, including Custom Translator, which helps customers to build translation systems for domain-specific terminology and style.

Neural machine translation technology has recently achieved impressive quality gains, characterized by highly fluent and accurate output, even for low-resourced languages such as Gaeilge. Using deep learning, we have iteratively refined our machine translation models. With today's release, our commitment to deliver high-quality machine translation for Gaeilge moves to the next stage, as we prepare to continuously improve translation quality based on feedback from our users.

Continued EU Funding for Translation

Digital Europe

Draft Orientations for the preparation of the work programme(s) 2021-2022

A) BUILDING ESSENTIAL DIGITAL CAPACITIES 6

World leading computing and data handling capacities 7

ARTIFICIAL INTELLIGENCE – Exploiting the potential of data and putting AI to work 9

CYBERSECURITY AND TRUST – Creating a cyber shield for Europe 11

WIDENING THE BEST USE OF DIGITAL TECHNOLOGIES 28

Building trust for the digital transformation 28

Language Technologies 29

Digital Transformation of Learning and Education 30

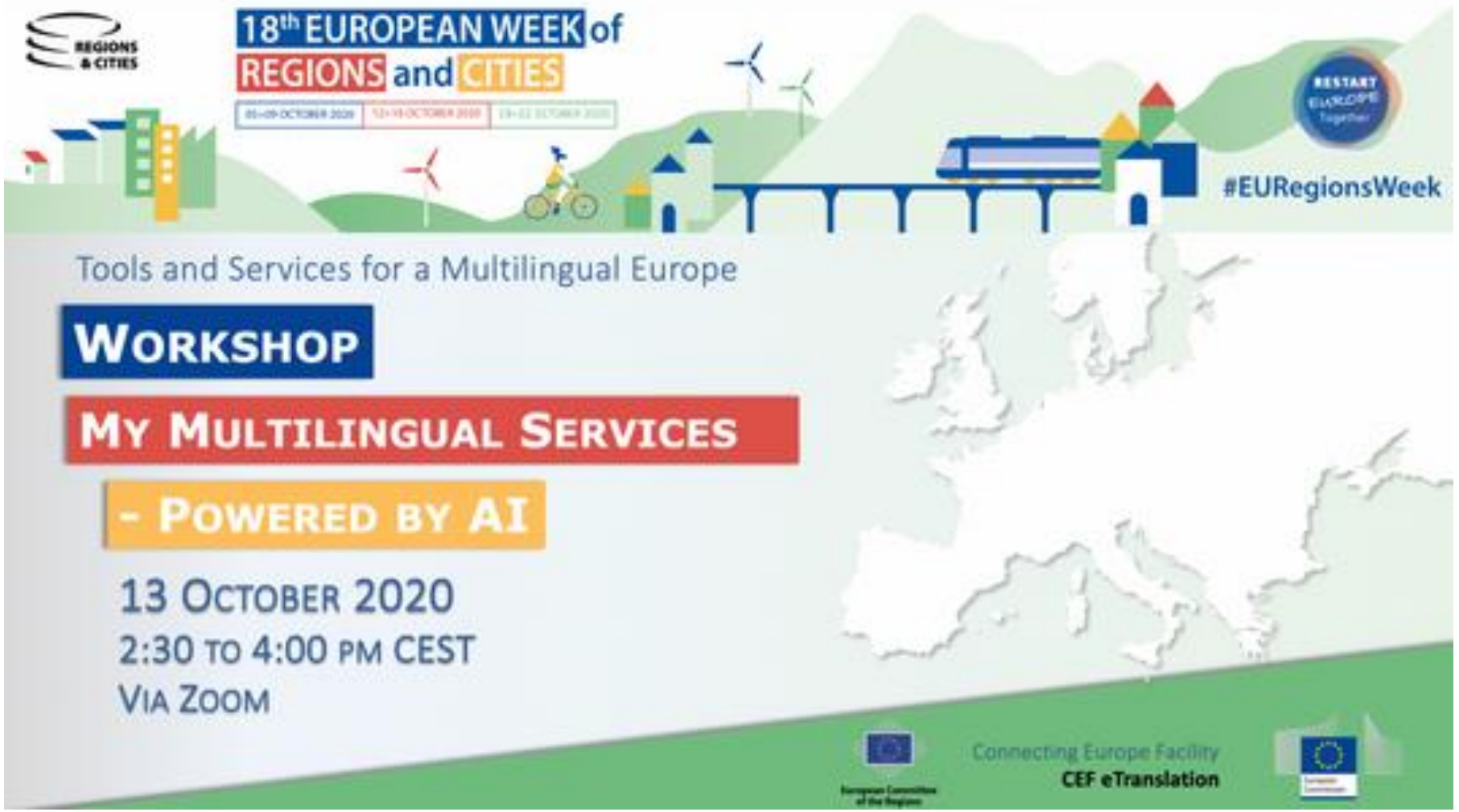
The objective of this action is to lower the language barriers for citizens and European SMEs to access the Digital Single Market and to promote the development of language processing services (e.g. voice-based services, translation, semantics and text analytics) by deploying language technologies, which will provide better access to information, products and services for European citizens and SMEs. Activities will foster the creation of an efficient ecosystem of technology providers, the localisation and uptake of language technologies, and access to the necessary data sets and computing power.

In line with this, the focus for the first two years of the programme will be on: (1) fostering the uptake of language technologies by offering support for standardisation, localisation, testing, and piloting local and cross border deployment of these technologies notably by European SMEs. More specifically, AI resources made available under Specific Objective 2 (e.g. through the data spaces and the Common European AI platform) and relevant to the language technologies will be adapted to the local language requirements. Cross-border specifications and standardisation efforts will be established in order to lower the cost of such local adaptations. Finally, the deployment and take-up of these solutions by European SMEs will be supported through dedicated grants. (2) providing automatic aggregation of the existing relevant data sets from the data space and access to HPC capabilities to support the processing, development and deployment of localised language technologies; and (3) mobilising the community through a network of local anchor points, gathering technological expertise in their language to ensure that all Member States can identify, analyse and adapt to digital trends, establish the needs and priorities of the public and private sectors, share best practice, and contribute to common specifications and standards.

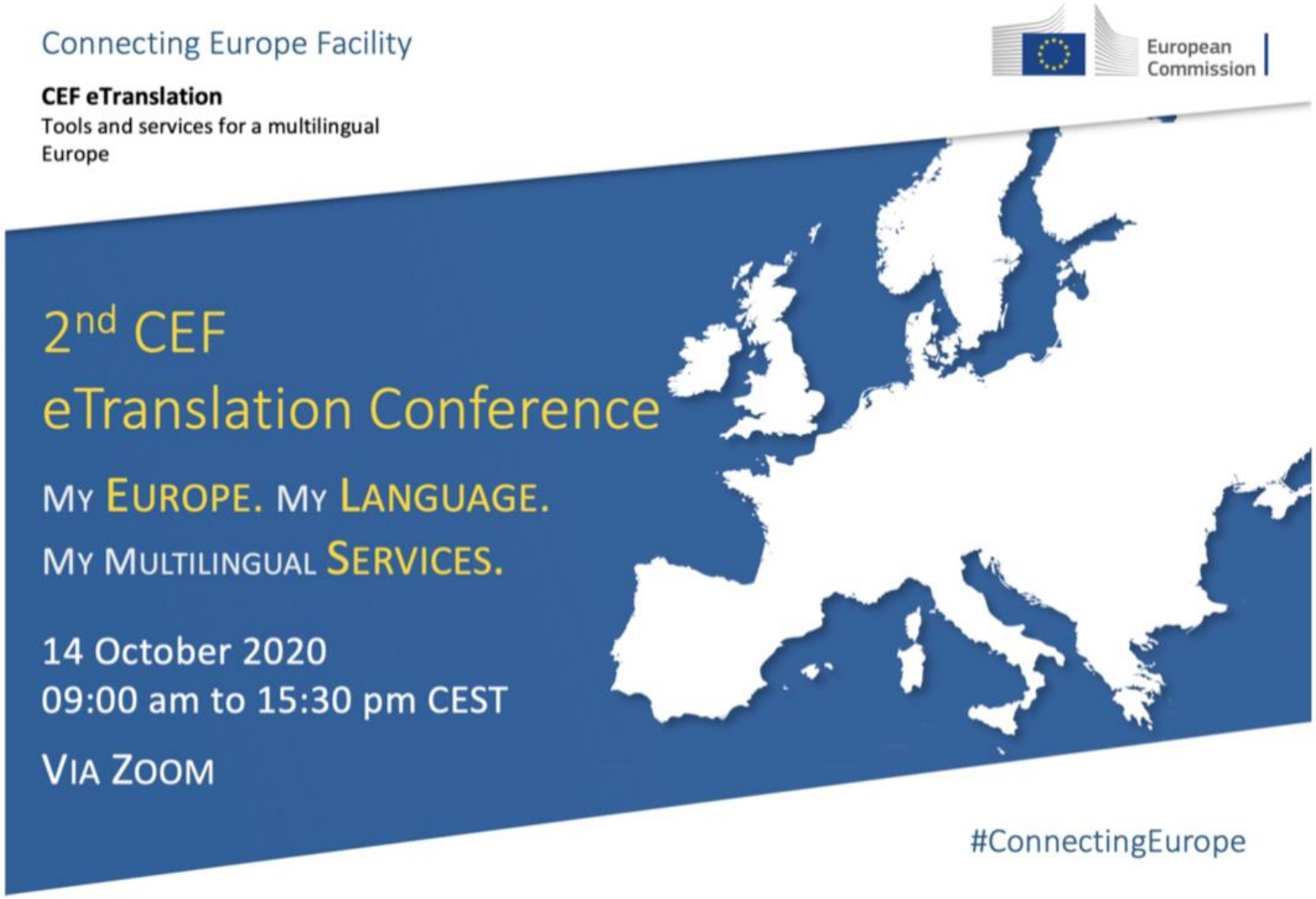
Activities on language technologies rely on, and feed into, other parts of the Digital Europe Programme: for example, the Digital Transformation Platform will provide baseline multilingual services before further localisation to dedicated application domains and services. The AI platform

http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=61102

Upcoming Conferences of Interest to AMTA Members



Upcoming Conferences of Interest to AMTA Members



Upcoming Conferences of Interest to AMTA Members

Connecting Europe Facility

CEF eTranslation
Tools and services for a multilingual Europe

European Commission

2nd CEF eTranslation Conference

MY EUROPE. MY LANGUAGE. MY MULTILINGUAL SERVICES.

14 October 2020, 09:00 am to 15:30 pm CEST, via ZOOM

CONFERENCE PROGRAMME

09:30 – 09:45	**Welcome address EC by the EC** (*Philippe Gelin, Head of Sector Multilingualism, DG CNECT, Unit G3*)
09:45 – 10:00	**Introduction: CEF AT Tools & Services update** (*Andrea Lösch, Project Manager, DFKI*)
10:00 – 11:00	**Inside CEF AT: Smart eTranslation applications and beyond**
09:30 – 09:50	Automated translation for scanned documents *(Pavel Smrz, Associate Professor, Faculty of Information Technology, Brno University of Technology)*
09:50 – 10:10	Smart Multilingual Cities (*Tonia Sapia, Marketing & Project Manager CEFAT4CITIES*)
10:10 – 10:30	Multilingual chatbot for Europeana (*Pavel Kats, Director, Jewish Heritage Network*)
11:00 – 11:30	*Virtual Coffee Break*
11:30 – 12:30	**Using eTranslation in your service – practical hints and questions**
11:30 – 11:50	How to get the eTranslation API for your service *(Khalid Choukri, CEO, ELDA)*
11:50 – 12:10	Customizing eTranslation for your services *(Luc Meertens, CEO, CrossLang)*
12:10 – 12:30	Beyond eTranslation: Your needs and requirements *(Thomas Vavra, Associate Vice President, Software, CEMA, IDC)*
12:30 – 12:45	**Live survey and results** (*Andrea Lösch, Project Manager, DFKI*)
12:45 – 14:00	*Virtual Lunch Break*
14:00 – 14:15	**Introduction Technology Watch** (*Andrea Lösch, Project Manager, DFKI*)
14:15 – 15:15	**Technology Watch: Innovative tools and services for enabling multilinguality**
14:15 – 14:30	Summarization *(Horacio Saggion, Associate Professor, Department of Information and Communication Technologies, University Pompeu Fabre - tbc)*
14:30 – 14:45	Anonymization *(Aivars Berzins, Project Manager, Tilde & Manuel Herranz, CEO, Pangeanic)*
14:45 – 15:00	Text to speech & Speech recognition *(Kris Demuynck, Professor at Faculty of Engineering and Architecture, Ghent University – tbc)*
15:15 – 15:30	**Live survey and outlook** (*Andrea Lösch, Project Manager, DFKI*)

https://ec.europa.eu/eusurvey/runner/eTranslationConference

EAMT 2020

HOME ABOUT PAPERS PROCEEDINGS SPONSORS REGISTRATION CALL FOR PAPERS
EAMT 2020 – NOVEMBER 3-5
The 22nd Annual Conference of the European Association for Machine Translation
READ MORE

EAMT 2020 – now fully online

EAMT 2020

EUROPEAN ASSOCIATION FOR MACHINE TRANSLATION

22nd ANNUAL CONFERENCE

The 22nd Annual Conference of the European Association for Machine Translation (EAMT 2020) will **take place remotely** ~~in Lisbon, Portugal~~, from November 3–5 2020 ~~on 4–6 May 2020, in the IST Congress Center at the main campus of Instituto Superior Técnico (IST) of the University of Lisbon~~.

The event is organised by Unbabel, INESC-ID and Instituto Superior Técnico.

We heartily appreciate the generous support of our sponsors and partners.

UPDATE: The EAMT 2020 **proceedings have been published online**.

General chair

- Mikel L. Forcada, EAMT President, Universitat d'Alacant (also Project/Product chair)

Track chairs

- Marco Turchi, FBK (Research track)
- Arianna Bisazza, Groningen Univ. (Research track)
- Joss Moorkens, DCU (Translators' track)
- Ana Guerberof, Univ. of Surrey (Translators' track)
- Mary Nurminen, Univ. of Tampere (User track)
- Lena Marg, Welocalize (User track)

EUROPEAN ASSOCIATION FOR MACHINE TRANSLATION

EAMT2020

Proceedings of the

22nd Annual Conference of the European Association for Machine Translation

3 – 5 November 2020

Online Conference

in place of

Instituto Superior Técnico, Lisbon, Portugal

Edited by

André Martins, Helena Moniz, Sara Fumega, Bruno Martins, Fernando Batista, Luisa Coheur, Carla Parra, Isabel Trancoso, Marco Turchi, Arianna Bisazza, Joss Moorkens, Ana Guerberof, Mary Nurminen, Lena Marg, Mikel L. Forcada

Organised by

Unbabel inesc id lisboa TÉCNICO LISBOA

https://eamt2020.inesc-id.pt/proceedings-eamt2020.pdf

FUNDED BY

Ireland's European Structural and Investment Funds Programmes 2014-2020
Co-funded by the Irish Government and the European Union

European Union
European Regional Development Fund

Science Foundation Ireland
sfi For what's next

Andy.Way@adaptcentre.ie

https://www.computing.dcu.ie/~away/