



eTranslation: Machine Translation for EU Translators and more

Markus Foti

Head of Sector for Machine Translation
European Commission

16 January 2020



ROY
SCHEIDER

ROBERT
SHAW

RICHARD
DREYFUSS

JAWS



Co-starring LORRAINE GARY · MURRAY HAMILTON · A ZANUCK/BROWN PRODUCTION
Screenplay by PETER BENCHLEY and CARL GOTTlieb · Based on the novel by PETER BENCHLEY · Music by JOHN WILLIAMS
Directed by STEVEN SPIELBERG · Produced by RICHARD D. ZANUCK and DAVID BROWN · A UNIVERSAL PICTURE ·
TECHNICOLOR[®] PANAVISION[®]   ORIGINAL SOUNDTRACK AVAILABLE ON MICRORECORDS & TAPES
...MAY BE TOO INTENSE FOR YOUNGER CHILDREN



Traduction

Français

Utilisateur: Alain REICHLING

[Suggestions](#) → [Fermeture de session](#) → [Aide](#)
[Traduire un fichier](#) | [Traduire un texte](#) | [Traduire un texte en ligne](#) | [Traduire une page Web en ligne](#)
1. Langue source:

Espagnol
 Danois
 Allemand
 Grec
 Anglais
 Français
 Italien
 Néerlandais
 Portugais
 Suédois

2. Langue(s) cible:

Espagnol
 Allemand
 Grec
 Anglais
 Italien
 Néerlandais
 Portugais

3. Domaine(s):**4. Autres options:**
Codes utilisateurs:

Compte rendu **Lettre** **Manuel d'instructions**
 CELEX
 Confidentiel

5. Introduire le texte :**6. Action:**

Annuler

Sauvegarder les options

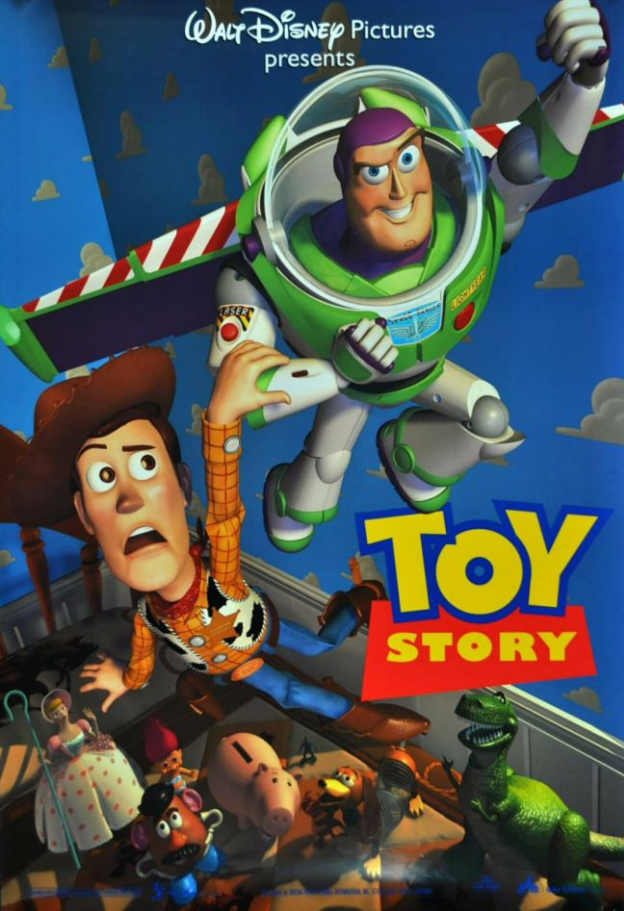
Charger les options

Traduire



ECMT

2.5 M pp/2009



1995



1999



2010



File Home Review Advanced View Add-Ins Help

Project Settings Copy Paste Batch Tasks File Actions

Cut Copy Paste Clear Formatting QuickInsert

Concordance Search Translation Memory

Select Previous Match Select Next Match Apply Translation

Show Translations Add New Term Quick Add New Term Terminology

Confirm Copy Source to Target Copy All Source to Target Clear Target Segment Segment Actions

Go To Navig...

Editor

Translation Results - GROW-2019-00104-00-01-00-ES-EN-00_ES-EN-MAIN,GROW-2019-00104-00-01-ES-PRT-00_ES-EN-RET,GROW-2019-00104-00-01-ES-PRT-00_CelexES-EN-AL,...

Project Settings

La·Comisión·Europea·en·su·Dictamen·Motivado·-·Infracción·2018/2306·dice: Displays results from lookups in translation memories and draft translations from

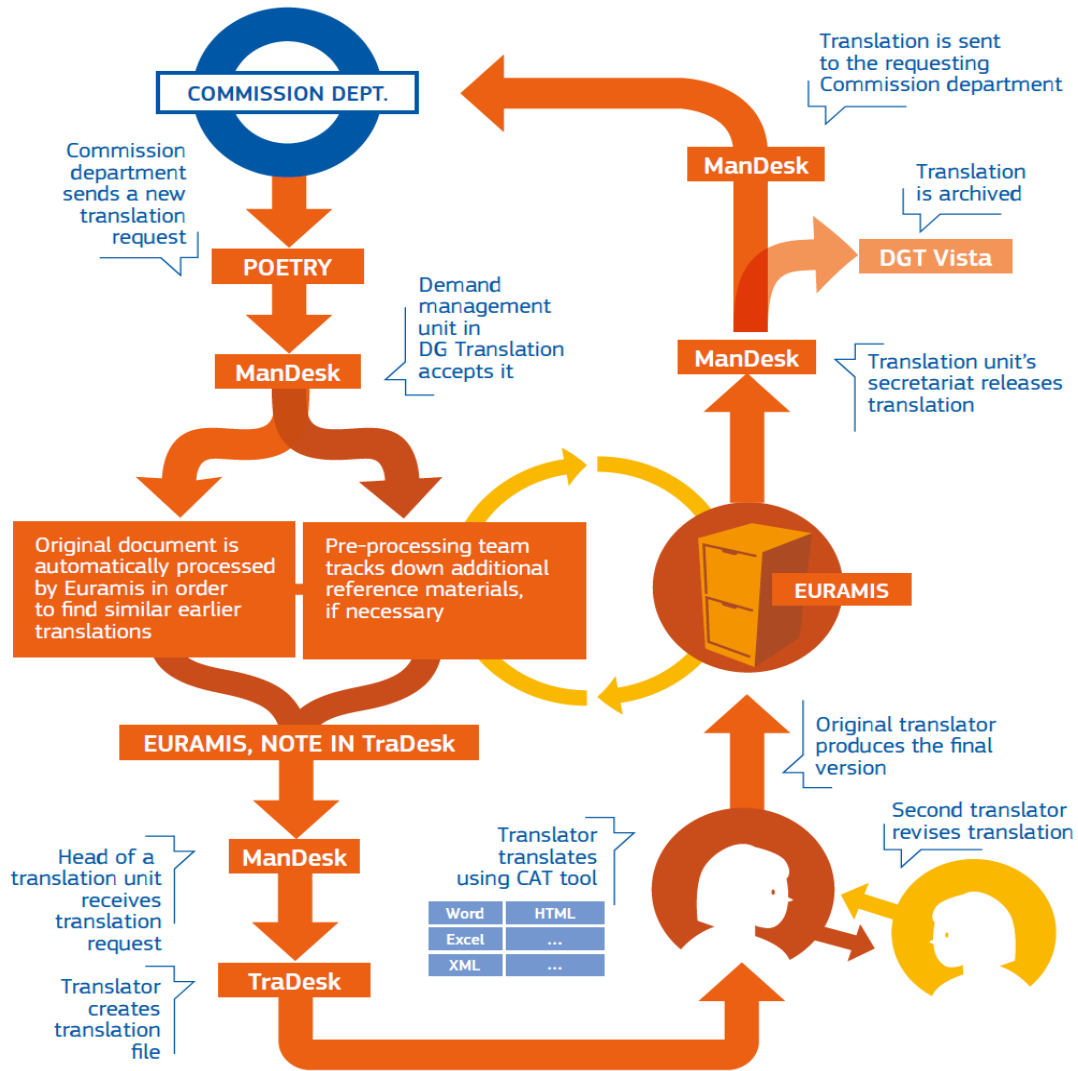
La·Comisión·Europea·en·su·Dictamen·Motivado·-·Infracción·2018/2306·dice:	CM	In·its·Reasoned·Opinion·-·Infringement·No°2018/2306,·the·European·Commission·stated·the·following:	Translator ALLIOCA Doc. No. 001040001
--	----	--	--

GROW-2019-00104-00-01-00-ES-EN-00_ES-EN-MAIN 25-09-2019 15:53:01 NET1\allicoa

Translation Results - GROW-2019-00104-00-01-00-ES-EN-0... Fragment Matches - GROW-2019-00104-00-01-00-ES-EN-0... Concordance Search Comments TQAs (0) Messages

GROW-2019-00104-00-01-00-EN-TRA-00.docx.sdlxliff [Translation]				
3	MINISTERIO	100%	MINISTRY	TB
4	DE·CIENCIA,·INNOVACIÓN·Y·UNIVERSIDADES	85%	OF·SCIENCE,·INNOVATION·AND·UNIVERSITIES	TB
5	La·Comisión·Europea·en·su·Dictamen·Motivado·-·Infracción·2018/2306·dice:		In·its·Reasoned·Opinion·-·Infringement·No°2018/2306,·the·European·Commission·stated·the·following:	P
6	3.		3.	P
7	<i>SOBRE·LA·CONFORMIDAD·DEL·MARCO·JURÍDICO·ESPAÑOL·CON·LA·DIRECTIVA·2005/36/CE</i>		<i>CONCERNING·THE·COMPLIANCE·OF·THE·SPANISH·LEGAL·FRAMEWORK·WITH·DIRECTIVE·2005/36/EC</i>	
8	<i>Obligaciones·de·transparencia</i>	100%	<i>Transparency·obligations</i>	LI

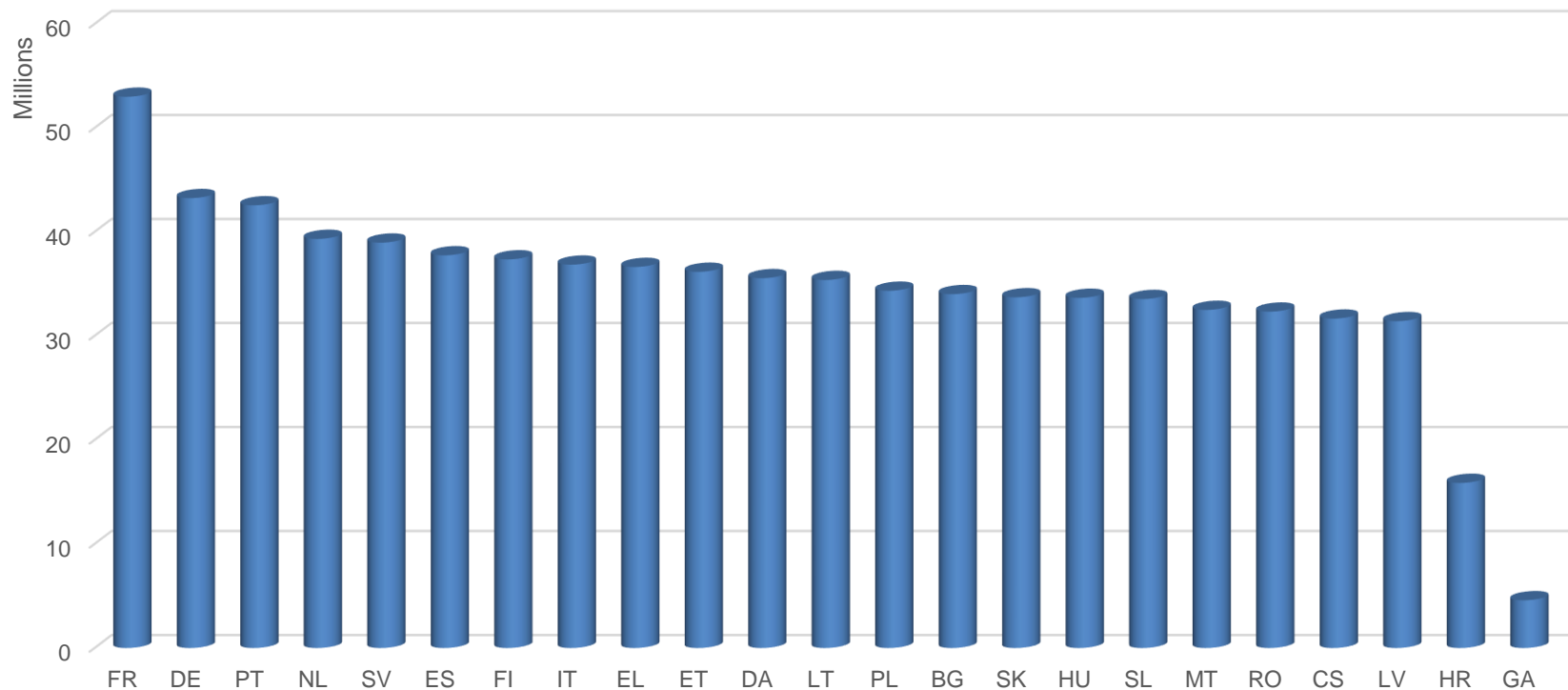
Translation workflow at the Directorate-General for Translation



A Treasure Trove of Eurospeak!



No of segments for MT training EN↔XX (millions)





European
Commission



MT@EC - Machine Translation

Markus FOTI
European Commission

[About](#) | [Feedback](#) | [Help](#)



[Translate documents](#) [Translate text](#) [My translation requests](#)

[My settings](#) English [Logout](#)

Select at least one source file :

OR

Translate from...



Translate into...



Output format:

Same as source TMX XLIFF

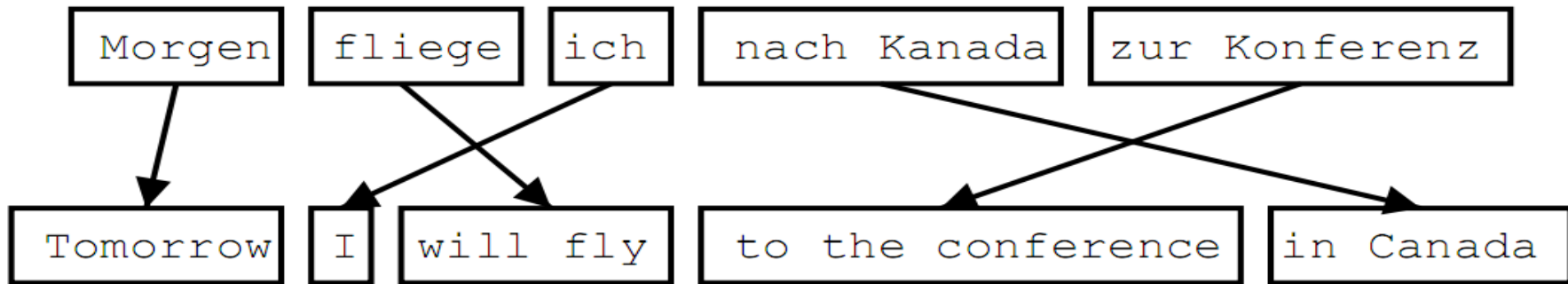
Enable translation Delete after download

Selected files

MT@EC

- Launched 26 June 2013
- **Statistical** machine translation in the Commission's data centre
- **Moses** based
- **State-of-the-art... at the time**

Building sentences in statistical MT





MT@EC

ECMT
2.5 M pp/2009

18 M pp/2017

The AI Revolution: Neural Machine Translation



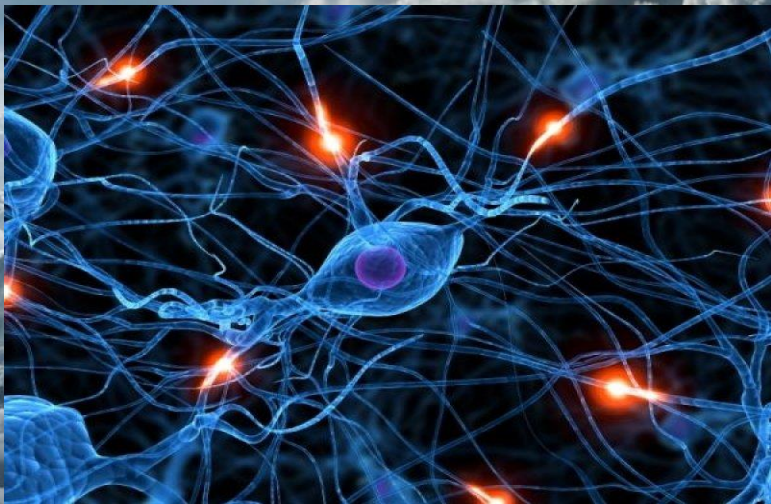
Building sentences in neural MT

Economic growth has slowed down in recent years .

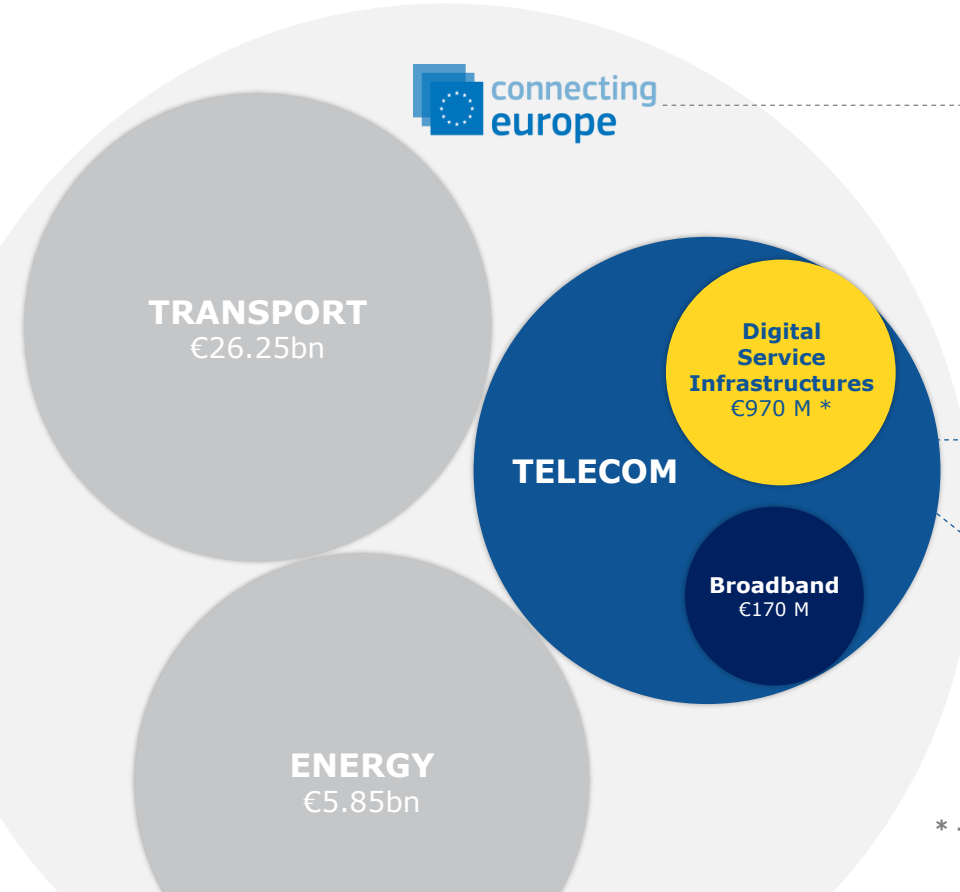
Das Wirtschaftswachstum hat sich in den letzten Jahren verlangsamt .

Economic growth has slowed down in recent years .

La croissance économique s' est ralentie ces dernières années .



The Connecting Europe Facility – funding programme 2014 -2020



CEF Regulation

The Connecting Europe Facility (CEF) is a regulation that defines how the Commission can finance support for the establishment of trans-European networks to reinforce an interconnected Europe.

CEF Telecom Guidelines

The CEF Telecom guidelines cover the specific objectives and priorities as well as eligibility criteria for funding of broadband networks and Digital Service Infrastructures (DSIs).

CEF Work Programmes

Translates the CEF Telecom Guidelines in general objectives and actions planned on a yearly basis.

* - 100 M Juncker Package



15 November 2017

**A Connecting Europe Facility
Building Block**

DG Translation and DG Connect

<https://webgate.ec.europa.eu/etranslation>



DGT

EURAMIS

eTranslation

Qualifications translated

Download

Data cleaning

- differences in segment length (source vs target language)
- very short or very long segments
- misalignments (alignment “shifted” by one or more segments)
- deduplication
- segments containing text in a language different from its label
 - **e.g.: segments containing long quotations in a different language than language of the segment**
- Sentence Managers very beneficial
- CLONE of Euramis – not Euramis
- About 10% removed

You need a Toolkit...



- currently our preferred toolkit
- NOT a turn-key solution
- Open source

NMT

Marian NMT

@marian_nmt

keye



**1090 followers
on Twitter!**

What we did

- Architecture – Recurrent Neural Network (shallow)
- Separate vocabulary per language
- Vocabulary size: 80K words
- Byte-pair encoding
 - **unknown/rare words rebuilt from common letter sequences**
 - **e.g. con-tract, servi-ces, author-ities**
- Millions of iterations, checked every 30K

Examples of pre/post-processing

- Date formats: 20 June 1967 or June 20, 1967?
- Quotation marks: “, «, „ or ” (the Danish and Italian cases)
- Legislative acts: *Commission Delegated Regulation (EU) 2019/945 of 12 March 2019 on...*
- Extra spaces
- Morphological adaptations of Slavic placeholders
- Many called into question by NMT



Engine Factory

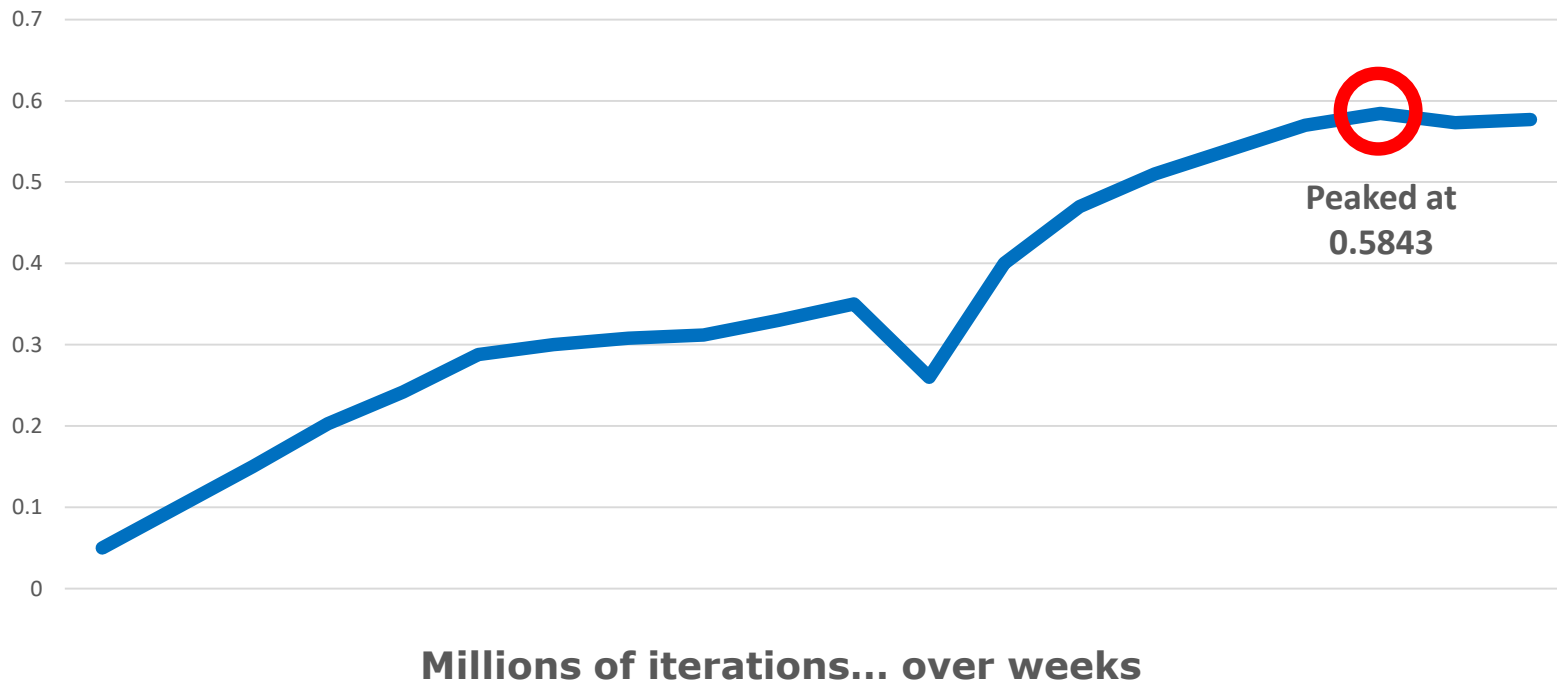
- Orchestrates machines
- Keeps track of builds
- Enables restarts
- Reports on progress
- Applies pre/post-processing

When is an engine done?

"To save time and resources we stopped the trainings if sentence-wise normalized cross-entropy on the validation set did not improve in 5 consecutive validation steps."

*Csaba Oravecz et al., eTranslation
WMT system description, 2019*

Progress in BLEU score during NMT training (EN→DA)



After WMT

"We focused [...] on finding the most appropriate combination of standard techniques and tools, which can thus directly be ported to production systems."

*Csaba Oravecz et al., eTranslation
WMT system description, 2019*

After WMT

Shallow recurrent network (2 layers)

→ **Deep Transformer (6 layers)**

Byte-pair encoding

→ **Sentence piece management**

e.g Hello_World. → Hello-_Wor-ld-.

Separate vocabulary

→ **Joint vocabulary**



Solve your language problems!

E-mail me my translation

0 / 2500

From

To

[+ Advanced options](#)

Translate text

Risolvere i problemi linguistici!

Lösen Sie Ihre Sprachprobleme!

Résoudre vos problèmes linguistiques!

Translate text

Translate documents

My translation requests

My settings

English

Logout

Click here to upload

Supported formats:        

From *

English

To *

French ✕ German ✕ Icelandic ✕

— Advanced options

Domain ⓘ

EU Formal Language

Output format

Same as source

TMX

XLIFF

QE (Beta)

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

Translate document

<https://webgate.ec.europa.eu/etranslation>

What works best

Texts related to EU policies – because the corpus behind it is built using translated EU documents

Legal (or legalistic) text

ES: Es cierto que todavía no se ha reunido la Comisión Interministerial **que crea** el art. 81 del R.D. 581/2017 **de transposición** al derecho español de la Directiva Europea de Reconocimiento de Cualificaciones Profesionales (en adelante PQD).

eTranslation:

It is true that the Interministerial Committee **established by** Article 81 of Royal Decree 581/2017 **for the transposition** into Spanish law of the European Directive on the Recognition of Professional Qualifications has not yet been convened.

Google Translate:

It is true that the Interministerial Commission **that creates** art. 81 of R.D. 581/2017 **of transposition** into Spanish law of the European Directive of Recognition of Professional Qualifications (hereinafter PQD) [**que todavía no se ha reunido** is not even translated]

What's not so good

- Unknown words
- “Out-of-domain” text
- Creative text

Unknown words

John is going to Ireland.

Caoimhe is going to Ireland.

“Out-of-domain” text

The **chair** was broken for the duration of the meeting.

Creative text

*'Come gather round people wherever you
roam, and admit that the waters around you
have grown'*

The Times They Are A-Changin', Bob Dylan

Typical comments from translators about MT

Pros:

- + aid for typing
- + time saving
- + can suggest creative solutions
- + guides terminological research
- + psychologically helpful:
something to work with instead of
a blank page (weaker languages)

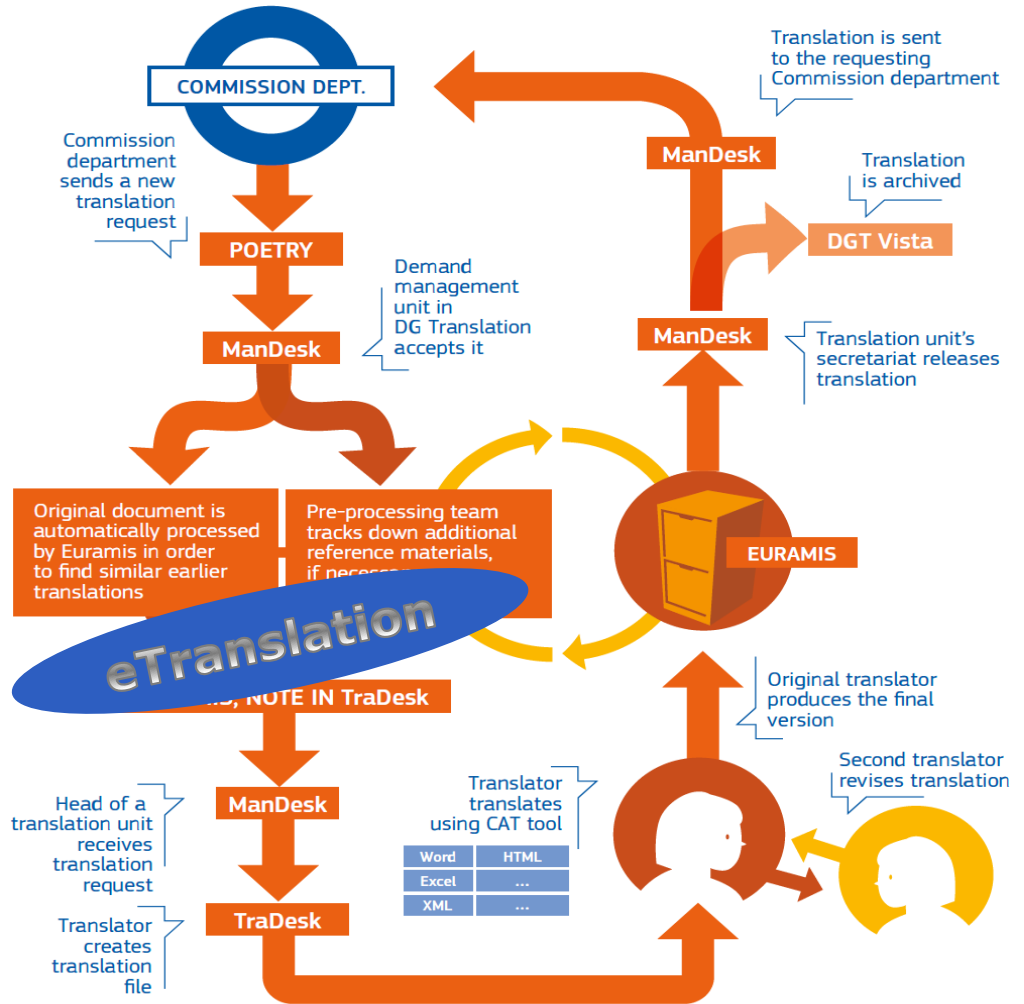
Cons:

- word order, can be overly literal
- can sound fluent but be off the mark
- omissions and additions
- incorrect terminology, which could
lead you astray
- less experienced translators tend to
make more/different mistakes than
when translating from scratch

When using neural engines, look out for...

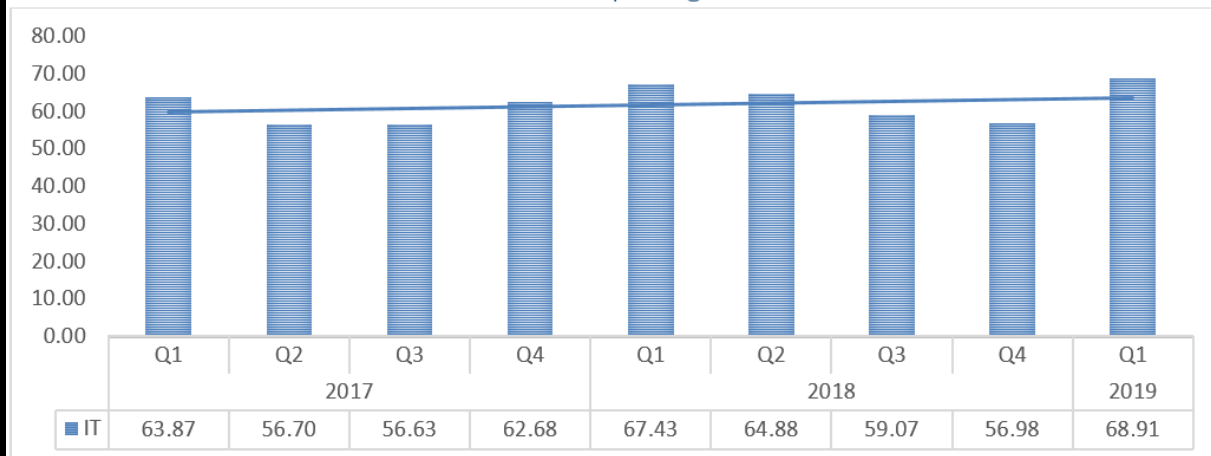
- **Incorrect terminology that may be very convincing** – never assume the machine knows better!
- **Repetitions and omissions** – mistakes you would not make yourself when translating from scratch!
- **Mistranslations, big or small** – especially when working from languages or on subject areas you are less confident in
- **Inconsistencies** – MT is sentence-based whereas you need to be able to ensure consistency over long documents

Machine Translation for DGT Translators

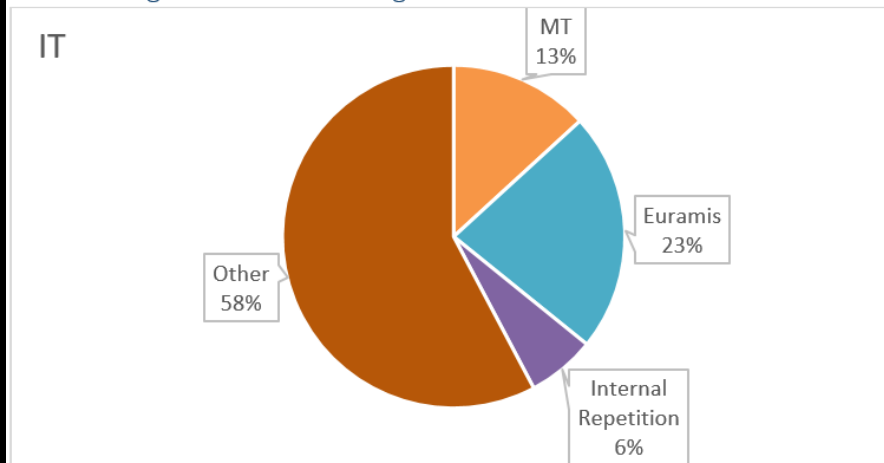


II.14 IT Language Department

II.14.1 Inclusion of machine translation in SDL package

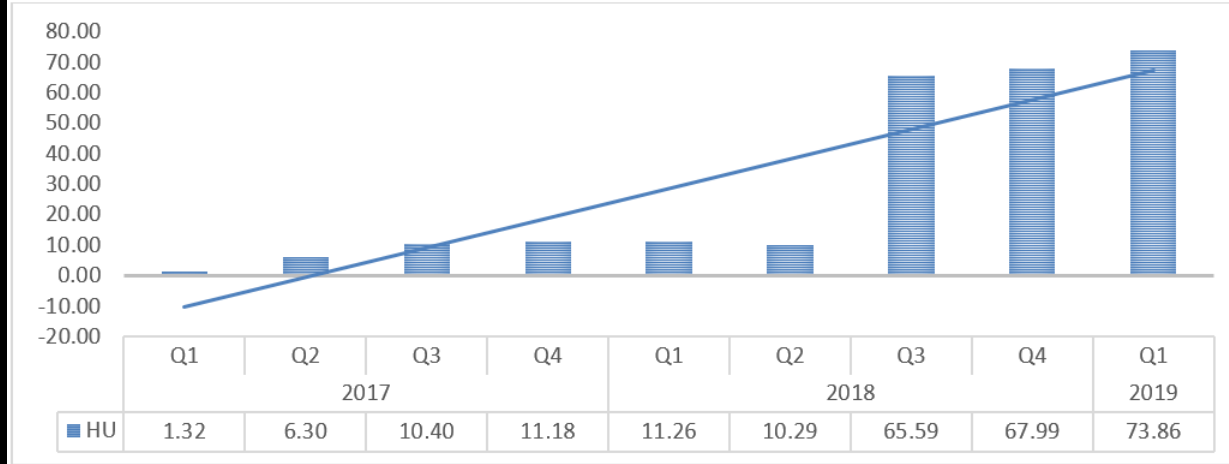


II.14.2 Origin of translated segments

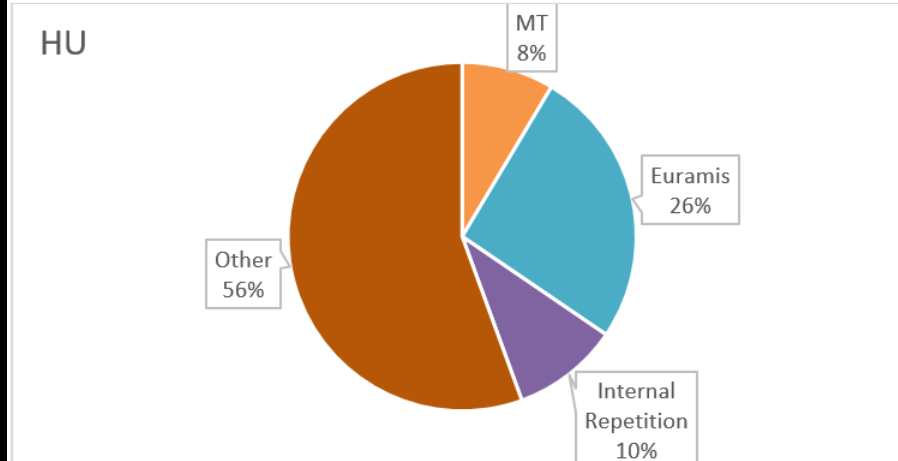


II.13 HU Language Department

II.13.1 Inclusion of machine translation in SDL package



II.13.2 Origin of translated segments



Who is using eTranslation today ?

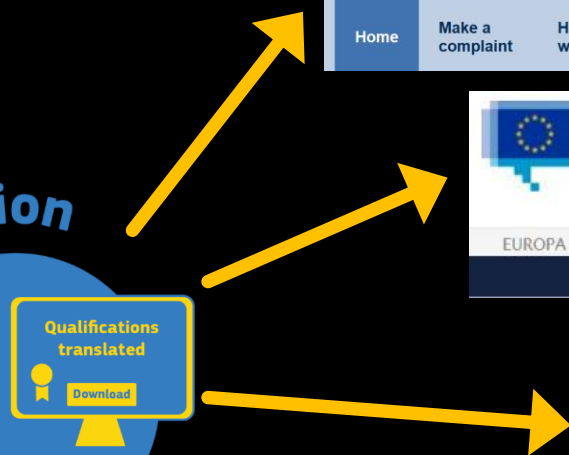
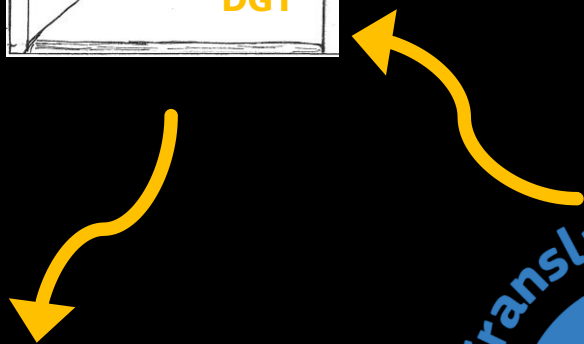
- **European institutions and bodies:**

- | | |
|--|---|
| <ul style="list-style-type: none">▪ Commission▪ Parliament▪ Council▪ Court of Justice▪ Court of Auditors | <ul style="list-style-type: none">▪ Economic and Social Committee▪ Committee of the Regions▪ European Central Bank▪ European Investment Bank,
etc. |
|--|---|

- **Online services** funded or supported by the EU/Connecting Europe Facility
- **Public administrations** in the EU countries, Iceland and Norway

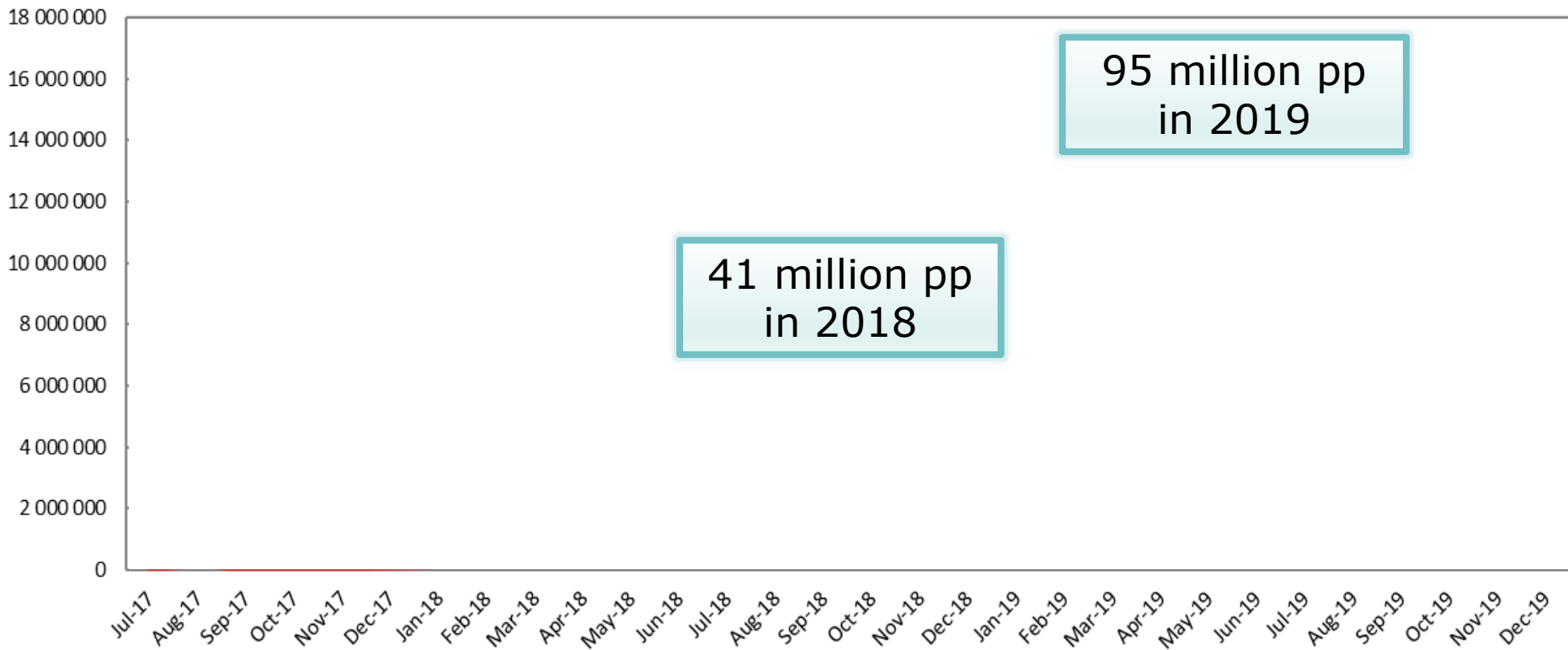


eTranslation





eTranslation Requests & Pages (2017-2019)



eTranslation

MT@EC

ECMT
2.5 M pp/2009

18 M pp/2017

95 M pp in 2019



Try it for yourself...



Machine Translation

Welcome to the registration page for [eTranslation](#), the European Commission's safe and secure machine translation system.

To use eTranslation, you need an EU Login ID. To register for EU Login, click [here](#), then return to this page.

Enter the email address associated with your EU Login account:

eTranslation is intended for European public administrations, Small and Medium-sized enterprises and University language faculties, or for [Connecting Europe Facility](#) projects.

Please indicate the type of user you are:

I certify that I belong to the category indicated above and therefore fall within eTranslation's remit.

[+ Privacy Statement](#)

[Start Translating!](#)

Register at:

<https://webgate.ec.europa.eu/etranslation/public/welcome.html>

Connecting Europe
eTranslation



Thank you!