

1. Human limitations in understanding languages
2. CEF eTranslation as an assistant
3. How it all functions from inside

What is a hyperpolyglot?

It's a person who speaks more than 11 languages*.

- Italian cardinal [Giuseppe Mezzofanti](#) (1774-1849):
30 languages + studied another 42
- A teacher Derick Herning from the Shetlands:
22 languages
- EC translator Ioannis Ikonomou:
32 languages



* Source: Michael Erard, author of [Babel No More](#)

What is a hyperpolyglot?

It's a language professional of some kind:

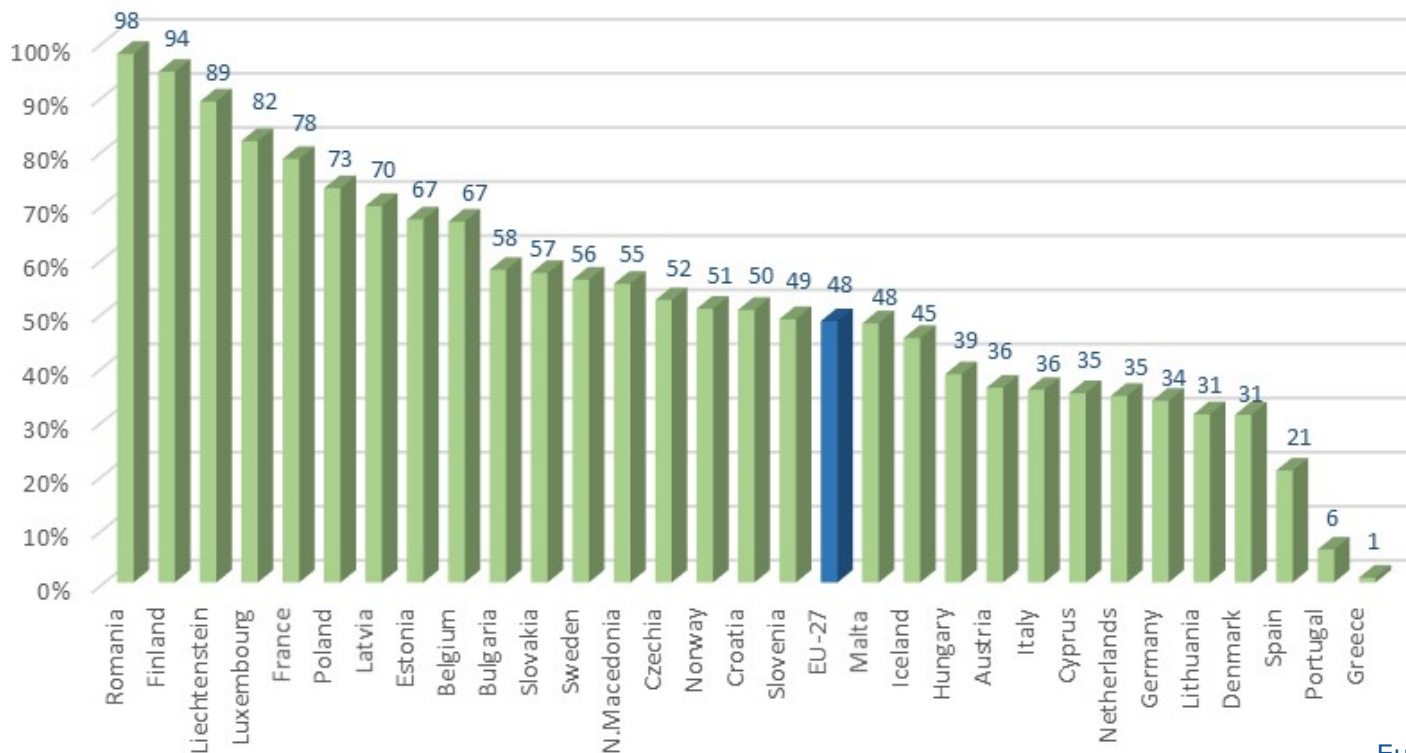
- Translator / interpreter
- Linguist
- Journalist
- Spiritual leader
- Actor



What is a hyperpolyglot?

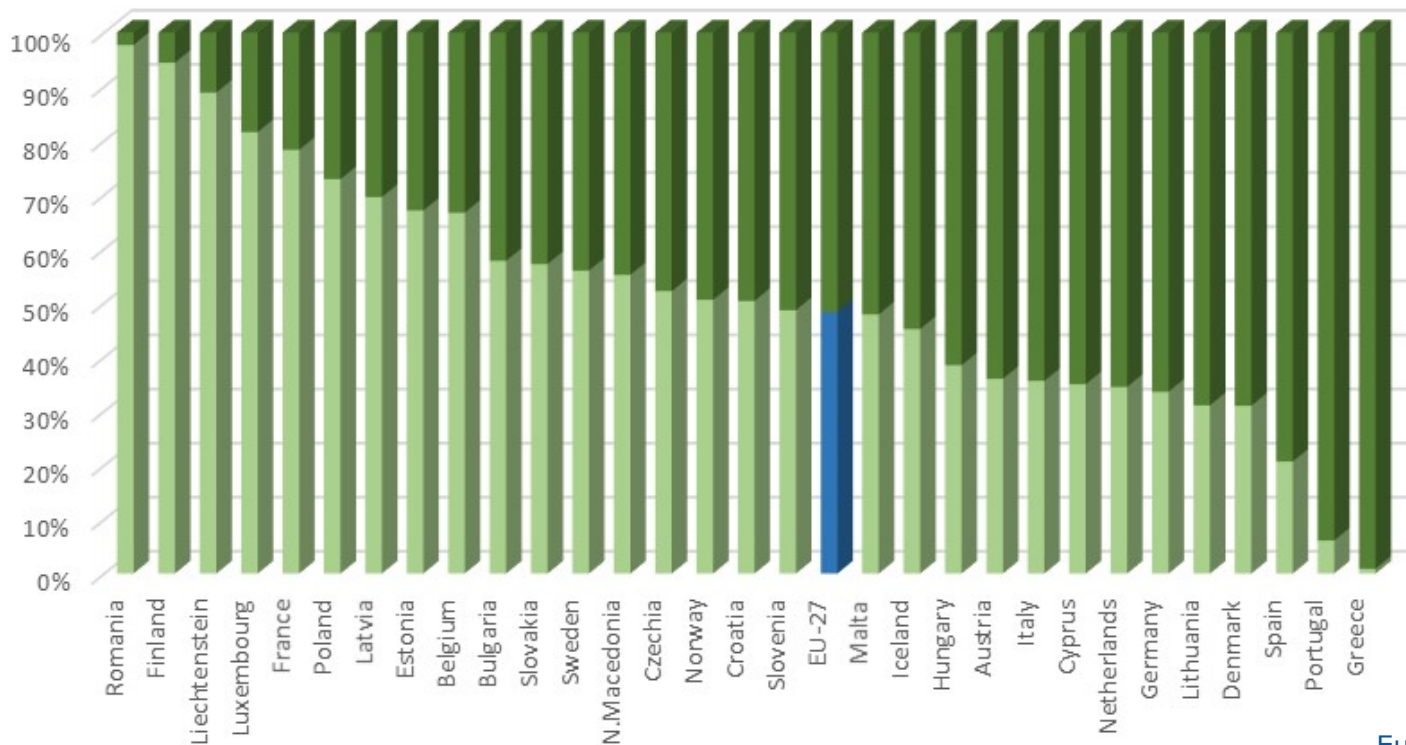


% upper secondary education pupils who learn two or more foreign languages at school, 2019



Eurostat

% upper secondary education pupils who learn two or more foreign languages at school, 2019



Eurostat

Machine translation solutions



- **eTranslation** (29 languages)
<https://webgate.ec.europa.eu/etranslation/>
- bing.com/translator (73 languages)
- translate.google.com (109 languages)
- translate.yandex.com (99 languages)
- <https://fanyi.sogou.com> (61 languages)

The Road to MT at the Commission



1975 -1998
**(28 language pairs,
rule-based)**

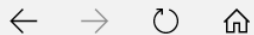


Launched 26 June 2013
**(All EU languages,
statistical)**



Launched 15 Nov. 2017
(All EU languages +IS, NB, neural MT)





Translate text

Translate documents

My translation requests

My settings

English

Logout

There is a bogymán in the courtyard.
There is a boogyman in the courtyard.
There is a boogyman in the courtyard.

Pihalla on mörkö.
Pihalla on mörkö.
Pihalla on mörkö.

113 / 2500



From

English



To

Finnish

— Advanced options

Domain

General Text



E-mail me my translation

Translate text



There is a bogymán in the courtyard.
There is a boogyman in the courtyard.
There is a boogyman in the courtyard.

Pihalla on mörkö.
Pihalla on mörkö.
Pihalla on mörkö.

113 / 2500



From

English



To

Finnish

Advanced options

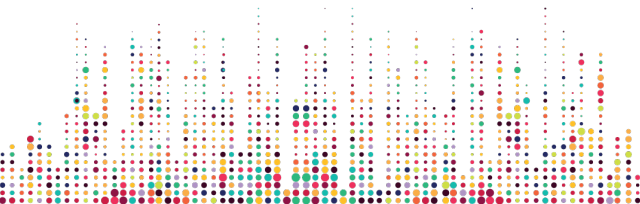
Domain



- EU Formal Language
- General Text
- Court of Justice Case Law
- Cultural
- Deutsche Bundesbank neural
- IP Case Law
- Ministère des Finances (France)
- Public Health
- Valtioneuvoston Kanslia

eTranslation: Two thematic engines

- **EU Formal:** built using translations of EU documents – best for formal documents and texts related to EU policies
- **General Text:** built using a wider scope of documents – best for informal documents and non-EU related texts





Käännä tekstiä

Käännä asiakirjoja

Omat käännöspyyntöt

Omat asetukset

suomi

Kirjautu ulos

Napsauta tästä asiakirjan lataamiseksi

Tuetut tiedostomuodot:

Lähdekieli *

suomi

Kohdekieli *

Valittu 27 kieltä 28:stä ✕

+ Lisäasetukset

- | | | | |
|--|--|---|---|
| <input checked="" type="checkbox"/> bulgaria | <input checked="" type="checkbox"/> englanti | <input checked="" type="checkbox"/> espanja | <input checked="" type="checkbox"/> hollanti |
| <input checked="" type="checkbox"/> iiri | <input type="checkbox"/> islanti | <input checked="" type="checkbox"/> italia | <input checked="" type="checkbox"/> kiina (yksinkert... |
| <input checked="" type="checkbox"/> kreikka | <input checked="" type="checkbox"/> kroaatti | <input checked="" type="checkbox"/> latvia | <input checked="" type="checkbox"/> liettua |
| <input checked="" type="checkbox"/> malta | <input checked="" type="checkbox"/> norja | <input checked="" type="checkbox"/> portugali | <input checked="" type="checkbox"/> puola |
| <input checked="" type="checkbox"/> ranska | <input checked="" type="checkbox"/> romanian | <input checked="" type="checkbox"/> ruotsi | <input checked="" type="checkbox"/> saksa |
| <input checked="" type="checkbox"/> slovakki | <input checked="" type="checkbox"/> sloveeni | <input checked="" type="checkbox"/> tanska | <input checked="" type="checkbox"/> tšekki |
| <input checked="" type="checkbox"/> turkki | <input checked="" type="checkbox"/> unkari | <input checked="" type="checkbox"/> venäjä | <input checked="" type="checkbox"/> viro |

Lataa asiakirja



Napsauta tästä asiakirjan lataamiseksi

Tuetut tiedostomuodot:

Lähdekieli *

suomi

Kohdekieli *

Valittu 27 kieltä 28:stä ✕

— [Lisäasetukset](#)

Aihealue

EU Formal Language

Tuotoksen
tiedostomuoto:

Sama kuin lähdetiedoston

TMX

XLIFF

QE (Beta)

- Lähetä käännös sähköpostiini
- Poista lataamisen jälkeen.

Käännä asiakirja



Translate text

Translate documents

My translation requests

My settings

English

Logout

Document in
Bulgarian.docx

14 KB



Another document
in Bulgarian.docx

14 KB



Supported formats:

From *

Bulgarian

To *

English ✕

Domain

EU Formal Language

Output format

Same as source

TMX

XLIFF

E-mail me my translation

Delete after download.

Translate document



Translate text

Translate documents

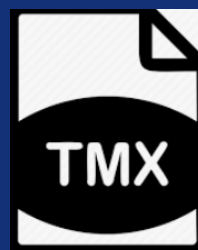
My translation requests

My settings

English

Logout

Supported formats:





[Translate text](#)

[Translate documents](#)

My translation requests

[My settings](#)

English ▾

[Logout](#)



Documents are available for download for 24 hours

<input type="checkbox"/>	Translated file	Status	Message	Document format	Target format	Source language	Target language
<input type="checkbox"/>	Rapport de mission_FR.docx	Ongoing		docx	docx	English	French
<input type="checkbox"/>	Another document in Bulgarian_FR.d...	Ongoing		docx	docx	English	French
<input type="checkbox"/>	Document in Bulgarian_FR.docx	Ongoing		docx	docx	English	French
<input type="checkbox"/>	PMO-2020-00018-00-00-PT-ORI-00_FR....	Translated		pdf	docx	Portuguese	French
<input checked="" type="checkbox"/>	PMO-2020-00018-00-00-PT-PRT-00_FR....	Translated		rtf	rtf	Portuguese	French

(1 - 5 / 5)

[Download](#)

CEF eTranslation webservice – “Machine to machine”



```
<text-to-translate>Hello World!</text-to-translate>  
<source-language>EN</source-language>  
<target-language>FI</target-language>  
<domain>GEN</domain>  
<email-destination>firstname.lastname@vnk.fi</email-destination>
```



**Client
application
in Helsinki**

```
<target-language>FI</target-language>  
<translated-text>Hei maailma!</translated-text>
```

**CEF
eTranslation
webservice**



It all starts here:

Kaikki alkaa täältä:

**1 570 000 pages
in 9 months 2020**

EURAMIS



Bilingual data: 24 000 000 Finnish-English sentences

Showing 30 hits / 30 (more results available)

TM: Budget Doc. No.: CE Req. Serv.: PB Year: 2020 SL: EN (Direct) Trans: unknown



EN: Commitment appropriations for cohesion policy (heading 1b) are set to rise by , and for agricultural, fisheries and **environmental policy** by .

FI: Koheesiopolitiikan maksusitoumusmäärärahoja (otsake 1 b) on lisätty ja maatalous-, kalastus- ja ympäristöpolitiikan määrärahoja .

Show the 3 remaining hits

TM: Budget Doc. No.: CE Req. Serv.: BD Year: 2020 SL: EN (Direct) Trans: unknown



EN: Commitment appropriations for cohesion policy (heading 1b) are set to rise by , and for agricultural, fisheries and **environmental policy** by .

FI: Koheesiopolitiikan maksusitoumusmäärärahoja (otsake 1 b) on lisätty ja maatalous-, kalastus- ja ympäristöpolitiikan määrärahoja .

Show the 3 remaining hits

TM: EP-Committees Doc. No.: 1196183 Req. Serv.: GREF-PRES Year: 2020 SL: EN (Direct) Trans: inh Doc. Type: RE Obs.: 2019/2824(RSP)



EN: having regard to Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine **environmental policy** (Marine Strategy Framework Directive),

FI: ottaa huomioon 17. kesäkuuta 2008 annetun Euroopan parlamentin ja neuvoston direktiivin 2008/56/EY yhteisön meriympäristöpolitiikan puitteista (meristrategiapuitteidirektiivi),

eTranslation reuses existing official translations

The price of failure is too high.
Energy is the life blood of our society.

Un échec serait trop coûteux.

L'énergie est le fluide vital de notre société.

Official translation

INTRODUCTION

The price of failure is too high.

Energy is the life blood of our society. The well-being of our people, industry and economy depends on safe, secure, sustainable and affordable energy. At the same time, energy related emissions account for almost 80% of the EU's total greenhouse gas emissions. The energy challenge is thus one of the

Introduction

Un échec serait trop coûteux.

L'énergie est le fluide vital de notre société. Le bien-être de nos populations, la prospérité de nos entreprises et de l'économie sont tributaires d'une énergie sûre, durable, disponible à un prix abordable et dont l'approvisionnement est garanti. D'autre part, les émissions de gaz à effet de serre liées à

[Go to unilingual display](#)

English (en)

French (fr)

Please choose

Display

[Display information about this document](#)

52010DC0639

/* COM/2010/0639 final */ COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Energy 2020 A strategy for competitive, sustainable and secure energy

52010DC0639

COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ DES RÉGIONS Énergie 2020 Stratégie pour une énergie compétitive, durable et sûre /* COM/2010/0639 final */

Bilingual data: sources



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH



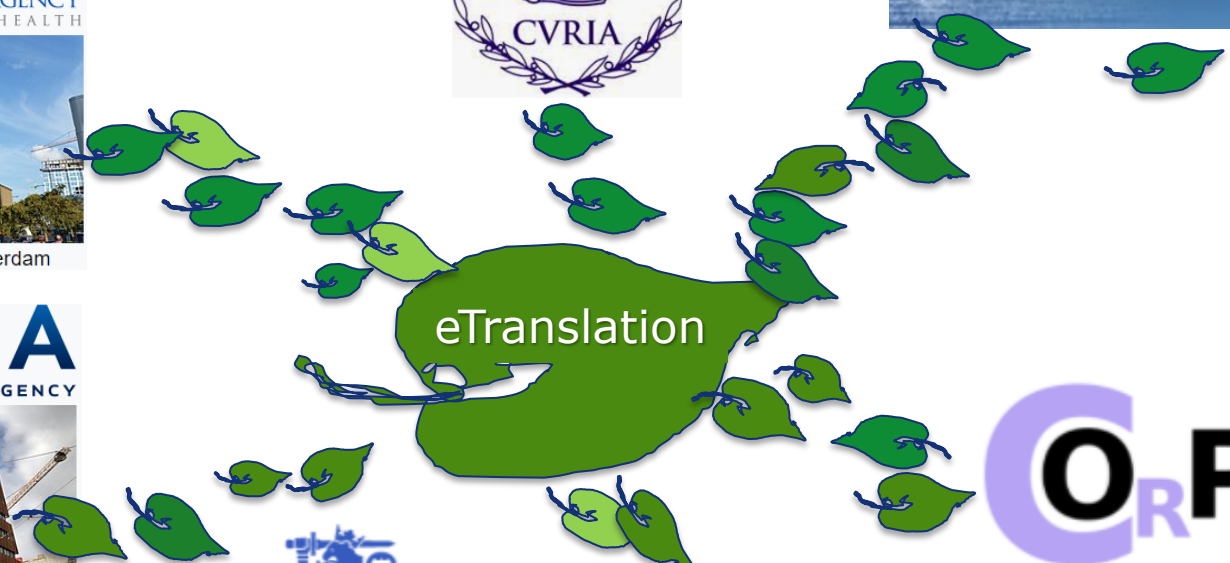
EMA headquarters in Amsterdam



EUROPEAN CHEMICALS AGENCY



Headquarters in Helsinki



PRIME MINISTER'S
OFFICE
VALTIONEUVOSTON
KANSLIA



jorg.tiedemann@helsinki.fi



European
Commission

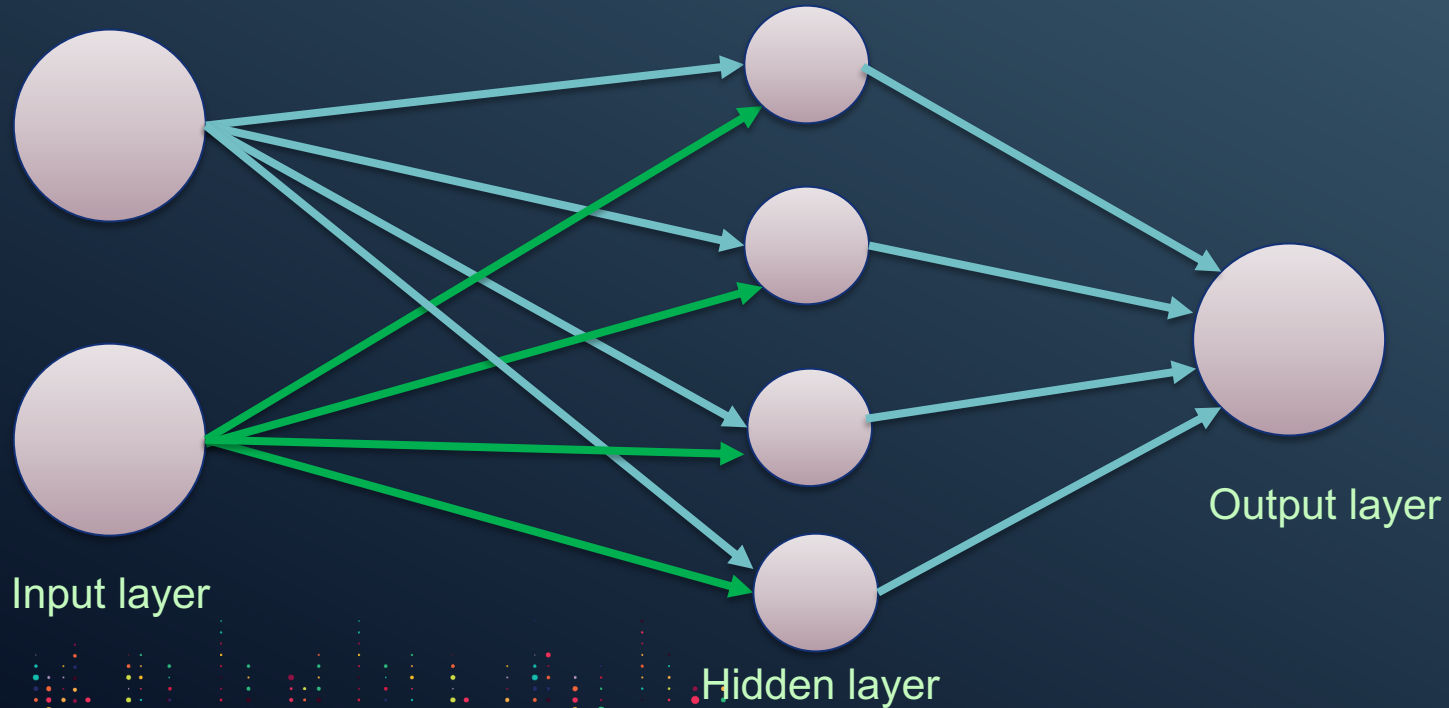
The neural translation engine

- It's comparable to:
 - a complex calculator
 - a functional black box
 - a complex mathematical function $y = f(x)$:

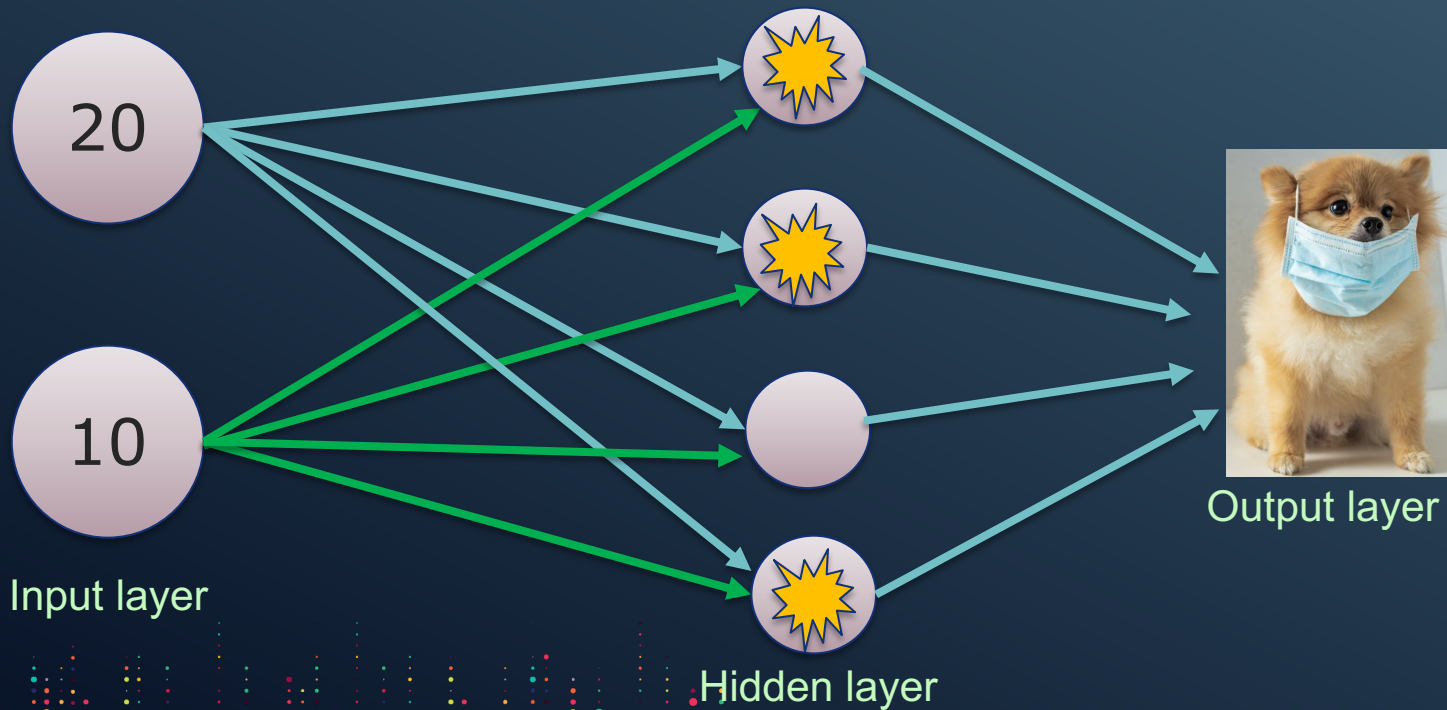
translated text = f (source text)



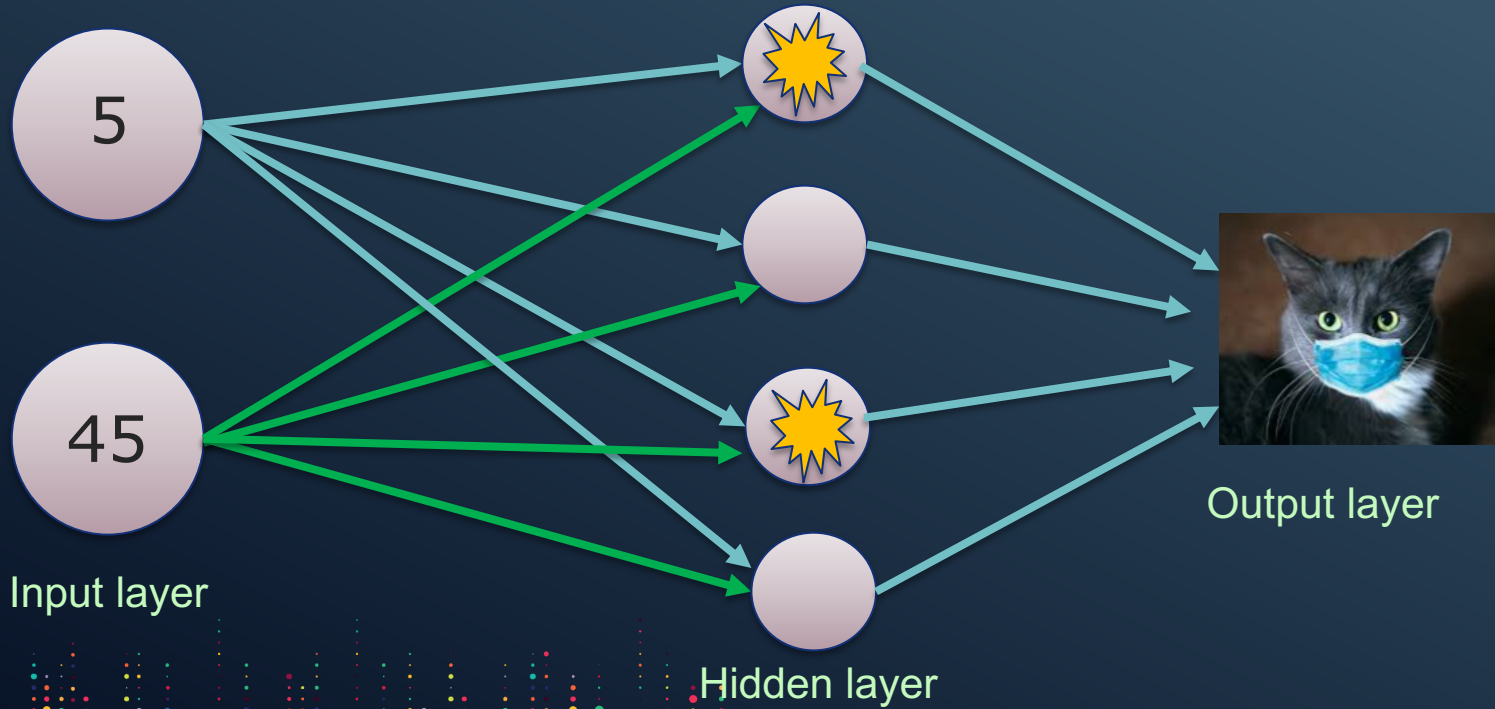
The neural network



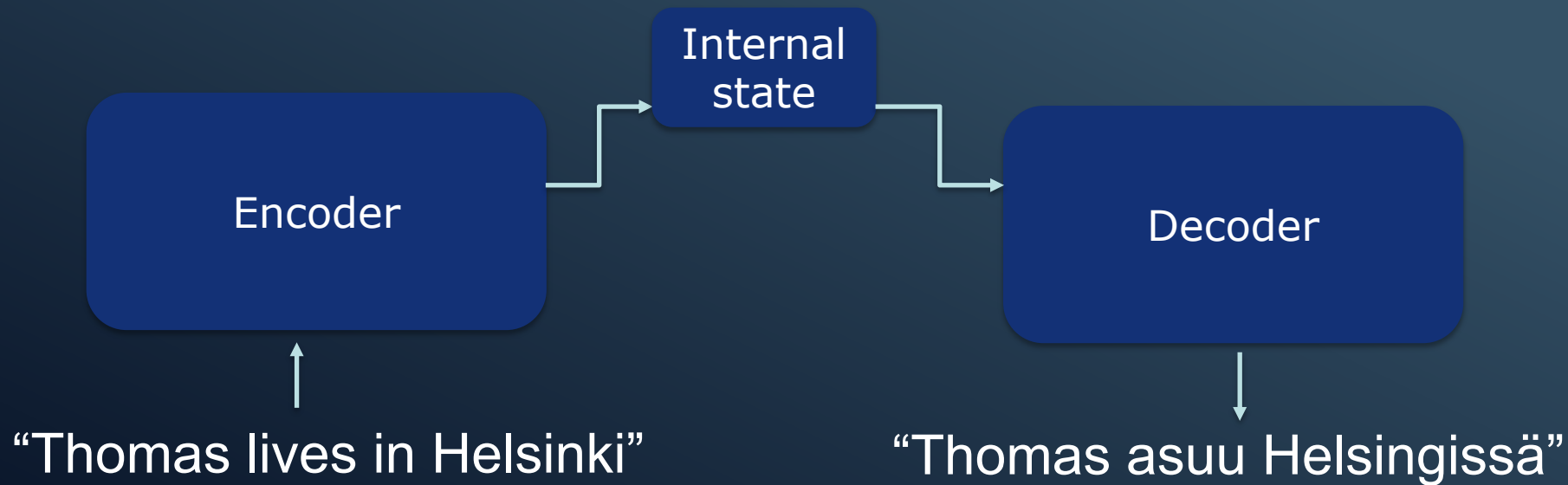
The neural network



The neural network



Simplest neural translation engine



The neural translation engine

- It needs to be produced first:
 - training lasts several days
- Two separate translation engines for each language pair are trained:
 - EN -> FI
 - FI -> EN



Words as vectors

Tomáš Mikolov

Czech computer scientist

2013

invented

“Word to vector”
method



Words as vectors

Each word in the vocabulary is represented as a long line ("vector") of numbers:

1. **A:** (-0,007, 0,0004, 0,0062, -0,0026,...) ← 300 "attributes" for each word
2. **ABANDON:** (-0,0078, 0,0023, 0,0049, 0,0015,...)
3. **ABILITIES:** (0,0067, 0,009, -0,005, -0,0006,...)
4. **ABILITY:** (-0,0094, 0,0001, -0,0087, 0,0092,...)
5. **ABORTION:** (0,0021, 0,0079, -0,0073, -0,0085,...)
6. **ABOUT:** (0,0095, 0,0029, 0,0007, -0,0016,...)
7. **ABOVE:** (0,0091, -0,0074, 0,0054, 0,0034,...)
- <...>
9996. **YOURS:** (0,0041, 0,0006, 0,0019, 0,0041,...)
9997. **YOURSELF:** (0,0034, -0,0007, 0,004, 0,0034,...)
9998. **YOUTH:** (0,0046, 0,0002, -0,0039, 0,003,...)
9999. **ZONE:** (0,0039, -0,007, -0,0067, 0,0074,...)
10000. **ZULU:** (-0,0067, 0,0097, -0,019, -0,054,...)



Words as vectors

- It allows the numerical comparison of the meaning of words. “similar vector” means “similar meaning”, e.g.:

Similarity estimate:



Sentence 1: “My wife and I don’t often program together”

0.926 Sentence 2: “My spouse doesn’t work with me, we program separately”

0.715 Sentence 3: “Move onto the next slide quickly!”

0.698 Sentence 4: “Mortgages are expensive”

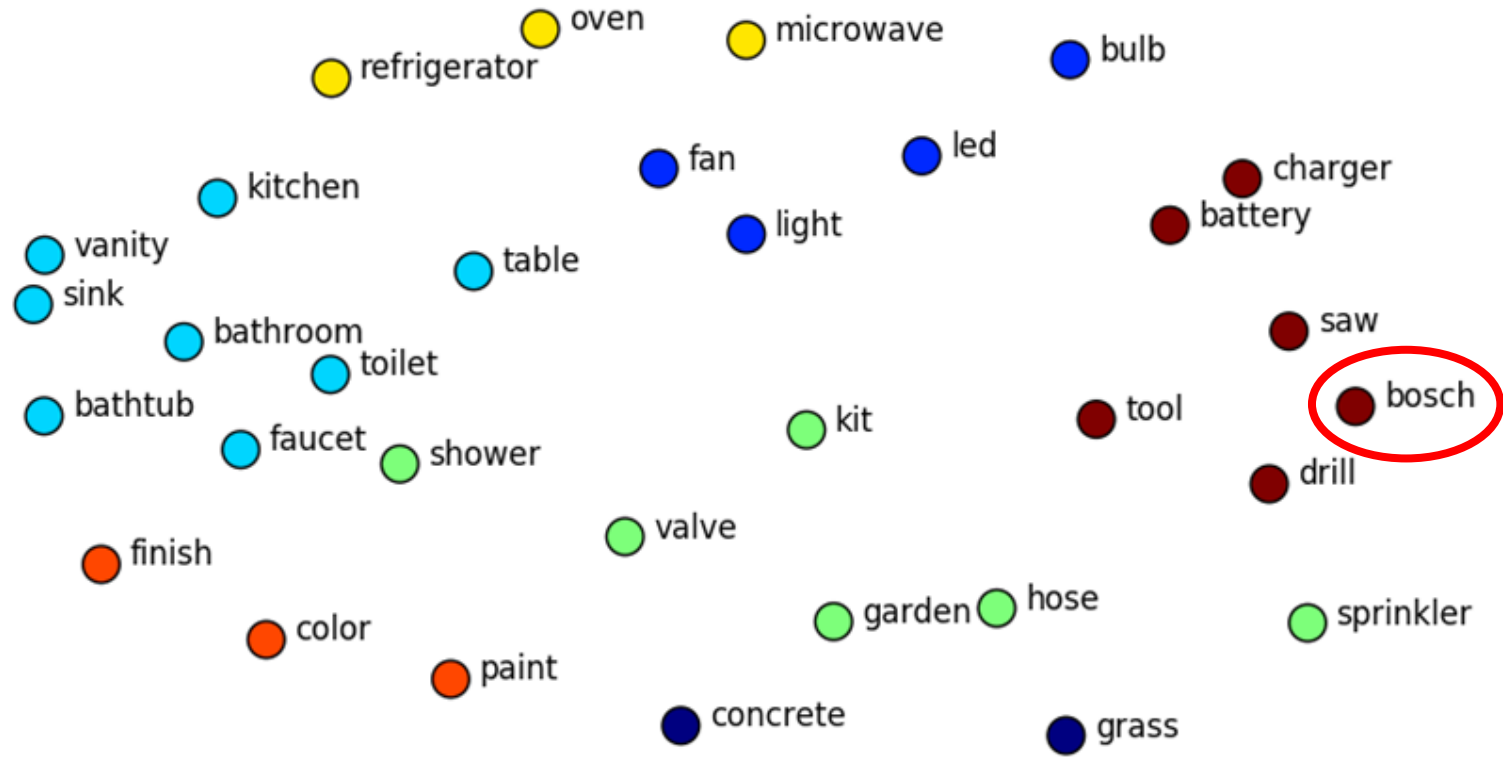
- We can even perform mathematical operations with words:

“Obama” – “USA” + “France” = “Sarkozy”

“Dublin” – “Ireland” + “France” = “Paris”



Word embedding in a “coordinate system”



Words with similar meanings are grouped through context



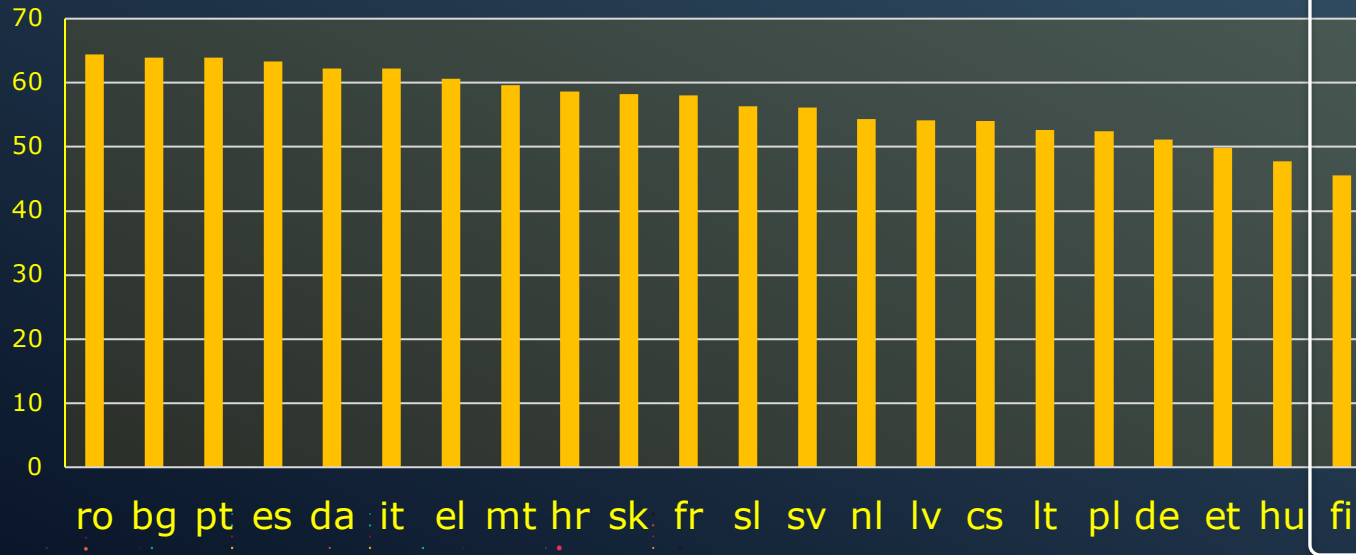
Challenges of machine translation

- Not possible to guarantee the correctness of the translation, a human reviser is needed
- Not yet possible to ensure the consistency of terminology over the entire document
- Not possible to impose the use of lists of terms
- Texts may be fluent but not accurate!



Translation quality

Translation quality score "BLEU",
translating from EN to another language

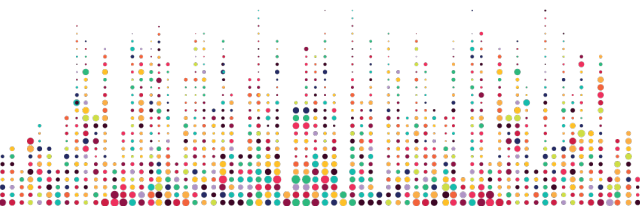


(2020 , EU formal language engines)

eTranslation security plan accepted by:

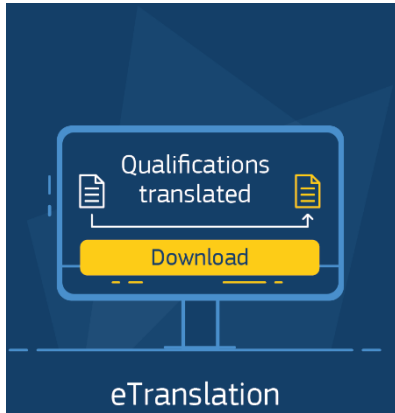
- Commission
- Parliament
- Council
- Committee of the Regions
- Economic and Social Committee
- Court of Justice
- Court of Auditors
- European Central Bank
- European Investment Bank

- Safe for “Sensitive non-classified” documents
- NO “EU Restricted/Restreint EU” or higher



How to register?



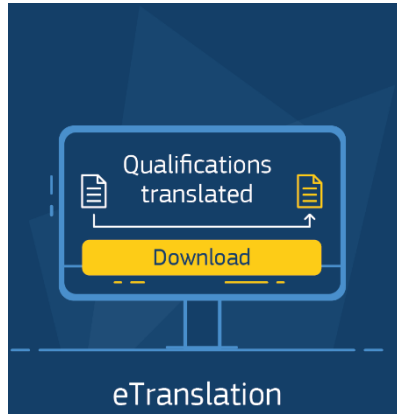


Eligible individual users

- **Self-registration via**
<https://webgate.ec.europa.eu/etranslation/public/welcome.html>

Web service (API) integration

- **Technical documentation:**
<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/How+to+submit+a+translation+request+via+the+CEF+eTranslation+webservice>
- **Contact the eTranslation Service Desk for credentials (and other support):**
help@cefat-tools-services.eu



Eligible individual users

- **Public Administrations**
- **Universities (members of the European Master's in Translation-EMT)**
- **CEF-funded projects**
- **SMEs (upon selection/validation)**

Q&A

A satellite view of Earth at night, showing city lights and the curvature of the planet against a starry background. The lights are concentrated in the lower half of the frame, with a thin blue and green atmospheric layer visible along the horizon. The background is a deep black space filled with numerous small white stars.