eTranslation

Machine Translation at the European Commission

Claus Thornby Larsen
Language Technology Coordinator
Directorate-General for Translation

København Universitet
16. November 2021

DGT in figures

- Directorate-General for Translation (DGT) is the Commission's internal LSP
- 24 Language Departments. 1300 translators (Danish language: 55 translators and 8 assistants (Luxembourg/Brussels))
- Translation of legislative documents (50 %), reports, communication and web content into mother tongue
- Translates between 2.3 and 3 mio. pages/year
- Outsourcing (approx. 38 %)

Equality of languages and technology?

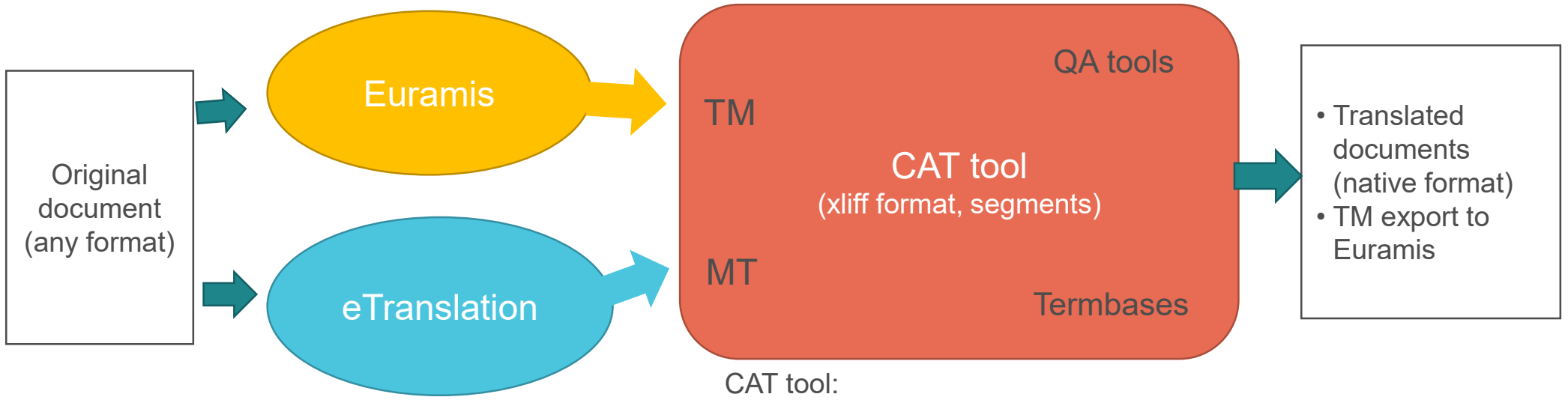
- Euramis (multilingual database with all translated text)
- IATE (Interactive Terminology for Europe)
- CAT Tool (Computer-Assisted Translation Tool)
- **eTranslation (internally developed)**
- Quality Assurance tools (internally developed)

Machine translation at the EU – past and present

- Systran (rule based). Low quality and language coverage
- MT@EC (2012-2018): Statistical machine translation
- eTranslation (2018 >): Neural machine translation
- AI (new): Quality estimation/language recognition/summarisation/anonymisation/classification (Eurovoc)

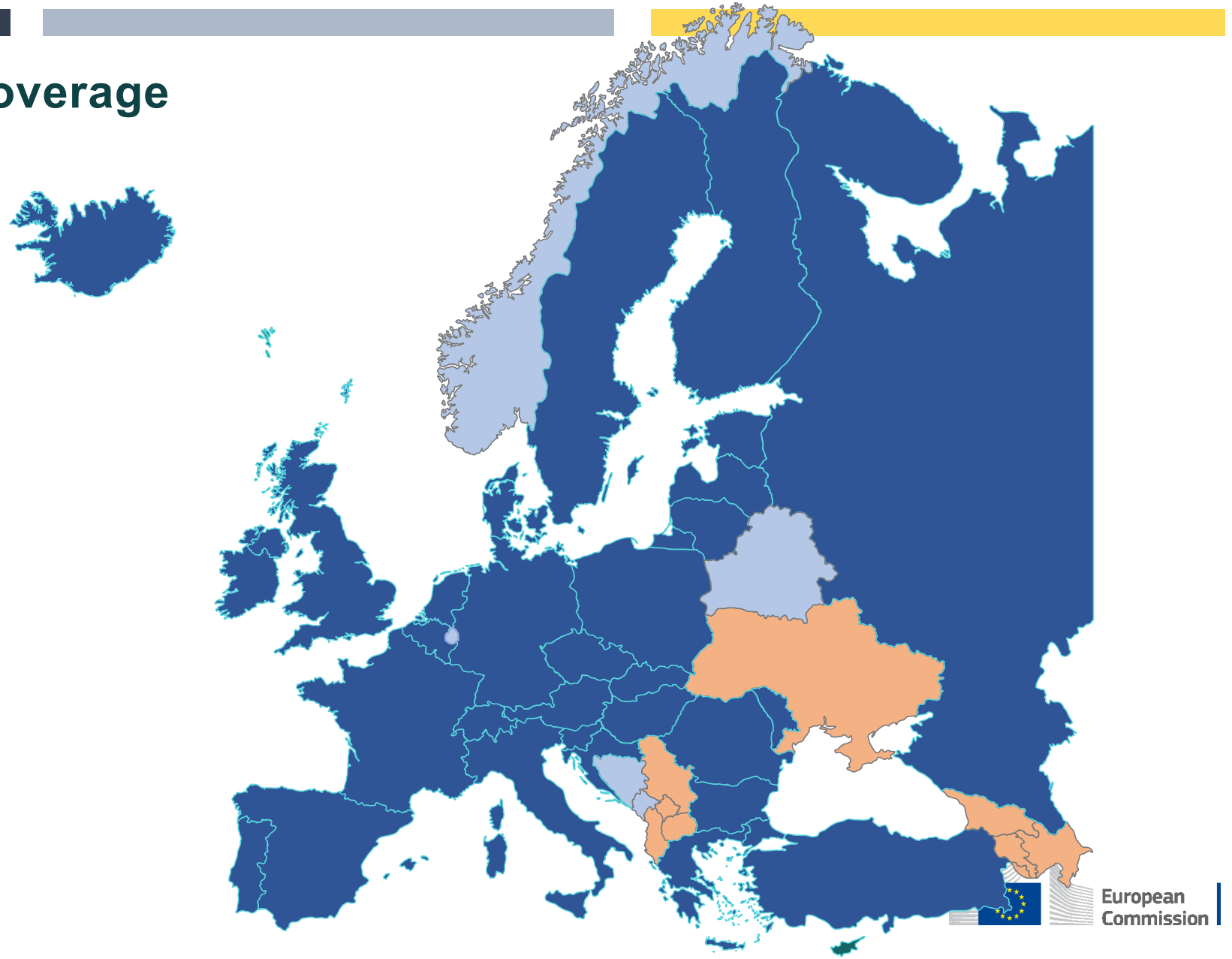
Our workflow: Language technology tools

- Translation memories (TM) and machine translation (MT) are retrieved automatically for each translation project



CAT tool:
In-house: Translation and revision
Freelance: Evaluation (revision)

eTranslation language coverage



Fully covered

Partially covered

Not covered

+ 3 non-European languages:
Mandarin Chinese,
Japanese,
Arabic

European Commission

Important facts about eTranslation

- All official EU languages (plus Norwegian, Icelandic, Russian, Chinese, Japanese, Turkish, Arabic)
- Neural machine translation trained with (mostly) own data
- > 185 million translated pages (2021)
- 250 million requests (2020)
- Developed by DGT, but available to all European public administrations, SMEs, freelance translators and universities

One system – Two audiences

- Translators and staff of the EU institutions
- Digital services of the EU institutions
- Websites of the Commission and the other European institutions

DGT eTranslation

- CEF Digital Service Infrastructures
- Pan-European digital public services
- Public administrations in the Member States
- Small and medium-sized enterprises

CEF eTranslation

Since March 2020.

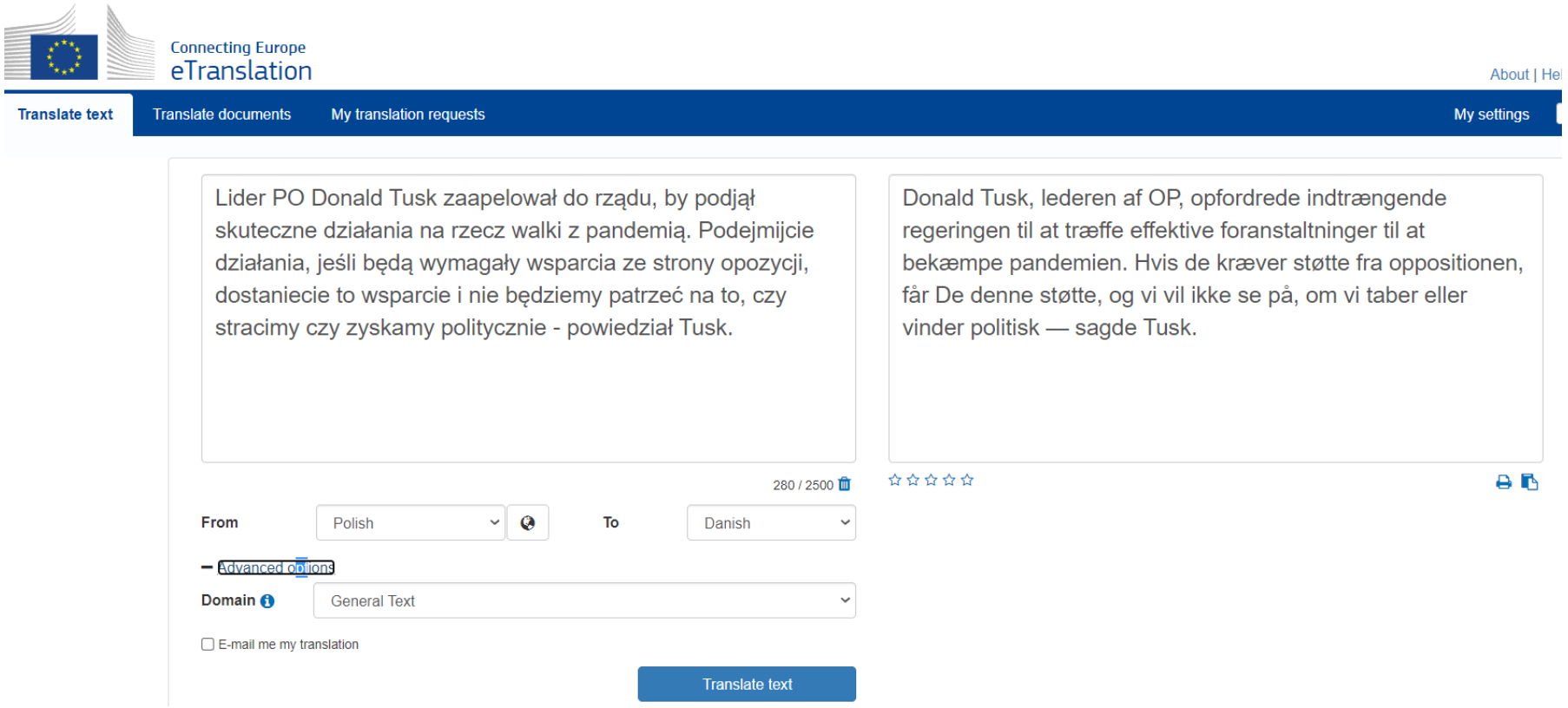
Key point for Digital Europe!

European Commission

European Commission

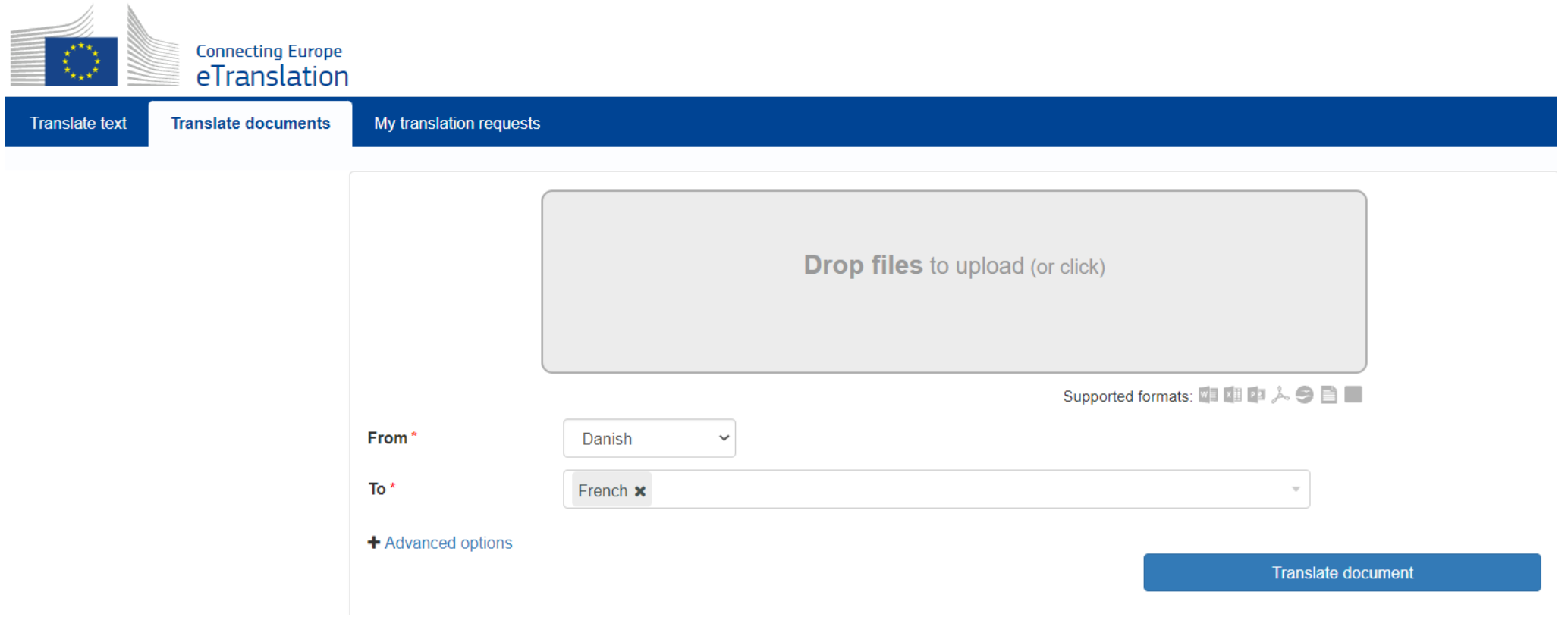
European Commission

A glance at eTranslation – translate snippets



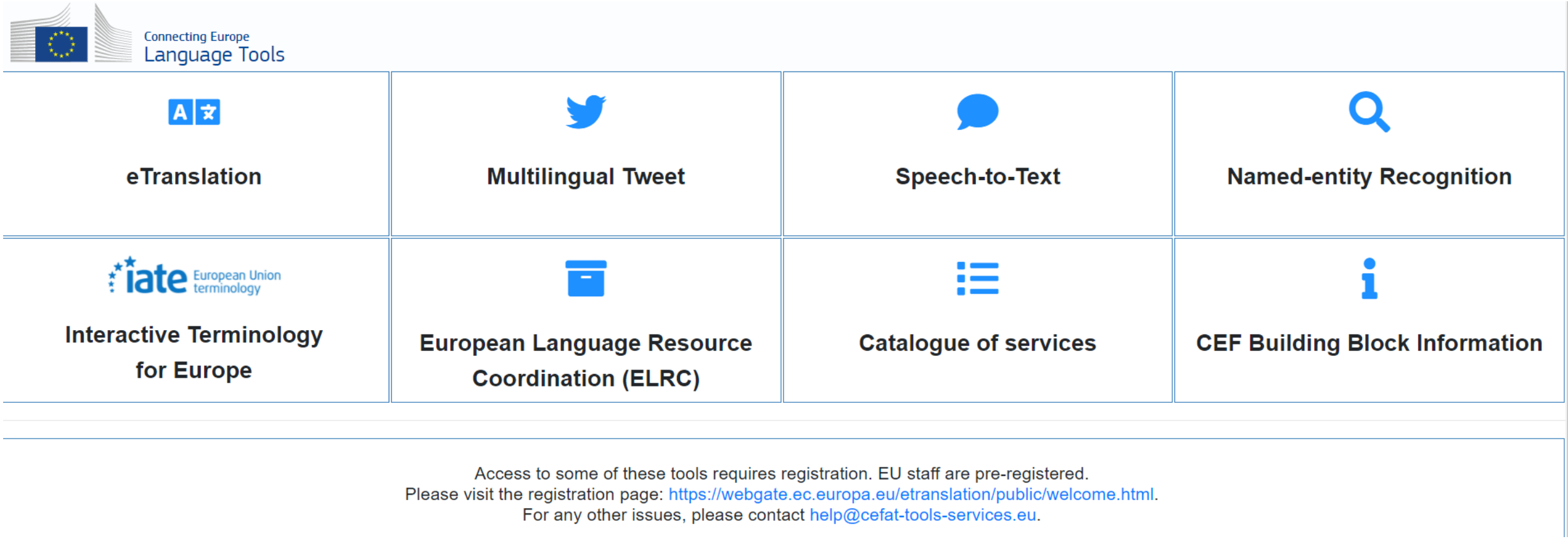
European Commission

eTranslation – upload docs for translation



European Commission

https://language-tools.ec.europa.eu/



European Commission

eTranslation in our CAT tool (TM vs. MT)

152	4.1.		4.1.	P
153	If the vehicle submitted for approval pursuant to this Regulation meets the requirements of this Regulation, approval of that vehicle type shall be granted.	98%	Hvis den køretøjstype, der ansøges om godkendelse af i henhold til dette regulativ, opfylder forskrifterne i regulativet, meddeles der godkendelse for den pågældende køretøjstype.	
154	4.1.1.		4.1.1.	P
155	The Technical Service appointed in accordance with paragraph 11. below shall check whether the required conditions have been satisfied.	100%	Den tekniske tjeneste, der er udpeget i overensstemmelse med punkt 11 nedenfor, kontrollerer, at de påkrævede betingelser er opfyldt.	
156	4.1.2.		4.1.2.	P
157	In case of doubt, account shall be taken, when verifying the conformity of the vehicle to the requirements of this Regulation, of any data or test results provided by the manufacturer which can be taken into consideration in validating the approval test carried out by the Technical Service.	100%	I tvivlstilfælde skal der ved efterprøvning af køretøjets overensstemmelse med forskrifterne i dette regulativ tages hensyn til alle oplysninger og prøvningsresultater, som forelægges af fabrikanten, og som kan tages i betragtning ved vurdering af den godkendelsesprøvning, som er foretaget af den tekniske tjeneste.	
158	4.2.		4.2.	P
159	An approval number shall be assigned to each type approved in accordance with Schedule 4 of the Agreement (E/ECE/TRANS/505/Rev.3).	AT	4.2. der tildeles godkendelsesnummer for hver godkendt type i overensstemmelse med skema 4 i aftalen (E/ECE/TRANS/505/Rev.3).	

European Commission

Metadata is important to us

In case of doubt, account shall be taken, when verifying the conformity of the vehicle to the requirements of this Regulation, of any data or test results provided by the manufacturer which can be taken into consideration in validating the approval test carried out by the Technical Service.	100%	I tvivlstilfælde skal der ved efterprøvning af køretøjets overensstemmelse med forskrifterne i dette regulativ tages hensyn til alle oplysninger og prøvningsresultater, som forelægges af fabrikanten, og som kan tages i betragtning ved vurdering af den godkendelsesprøvning, som er foretaget af den tekniske tjeneste.	Doc. No. 016960200 Stored by thomspe Translator thomspe TM Database Policies Req. Serv. GROW
~~4.1.2.~~ In case of doubt, account shall be taken, when verifying the conformity of the vehicle to the requirements of this Regulation, of any data or test results provided by the manufacturer which can be taken into consideration in validating the approval test carried out by the Technical Service.	AT	4.1.2 Hvis der er tvivl, skal der ved kontrollen af køretøjets overensstemmelse med kravene i denne forordning tages hensyn til data eller prøvningsresultater fra fabrikanten, som kan tages i betragtning ved godkendelse af den godkendelsesprøvning, som udføres af den tekniske tjeneste	

NMT pitfalls – What to keep an eye on

- At first glance, sentences translated with NMT look convincing (fluency/syntax)
- However, semantics and terminology must be checked
- Typical mistakes:
  - inconsistent terminology
  - meaning (wrong/inaccurate)
  - omissions/additions
  - negations (omitted or added)
  - made-up words

Terminological consistency

Commission declaration on the climate tracking methodology and the involvement of the European Parliament and Council	Kommissionens erklæring om klimasporingsmetoden og inddragelse af Europa-Parlamentet og Rådet
The Commission will ensure that the climate tracking methodology is accessible, transparent, and publicly available.	Kommissionen vil sikre, at klimasporingsmetoden er tilgængelig, gennemsigtig og offentligt tilgængelig.
The Commission will exchange views on the climate tracking methodology with the European Parliament and Council.	Kommissionen vil udveksle synspunkter om klimasporingsmetoden med Europa-Parlamentet og Rådet.

After next training of the MT engine!

Omission

MT

Coronavirus:	Coronavirus:
Commission Statement on consulting Member States on proposal to prolong and further expand State aid Temporary Framework	Erklæring fra Kommissionen om høring af medlemsstaterne om forslag til forlængelse og yderligere udvidelse af de midlertidige rammebestemmelser

PE

Coronavirus:	Coronavirus:
Commission Statement on consulting Member States on proposal to prolong and further expand State aid Temporary Framework	Erklæring fra Kommissionen om høring af medlemsstaterne om ~~forslag~~ ~~til~~forslaget om forlængelse og yderligere udvidelse af de midlertidige rammebestemmelser for statsstøtte

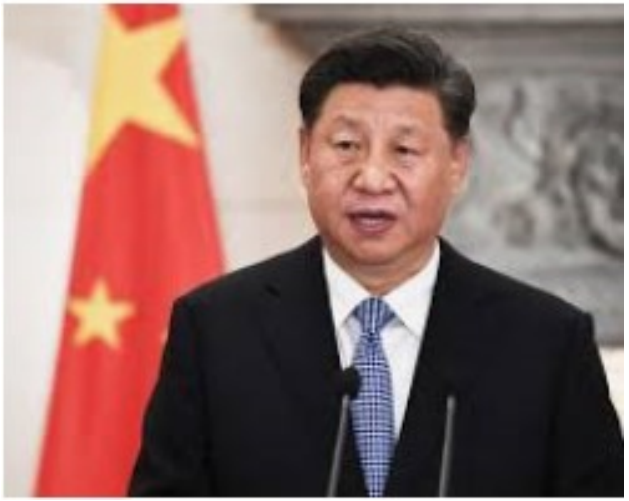
‘Made-up words’

This will enable, in particular, to increase the supply of certain types of PPE such as disposable facemasks.	Dete vil navnlig gøre det muligt at øge udbuddet af visse typer af personlige værnemidler såsom engangsmaling

Actions in response to the coronavirus pandemic:	Tiltag som reaktion på coronavis-pandemien:

Google, over 80 million coronavirus-related ads	Google, over 80 mio. oncoronrutsrelaterede annoncer

MT mistakes can be politically sensitive



- Facebook: Xi Jinping -> ‘Mr. Shithole’ (burmesisk-engelsk)
- EU-tekst: Kalkun i stedet for Tyrkiet (‘Turkey Gate’)

European Commission

Metaphors/idioms – tricky!

05.11.2021 KL. 05:30

Uvaccinerede lader sig teste mindre end dem, der har fået stikket

Flere eksperter peger på genindførelse af coronapasset som en genvej til, at flere lader sig teste.

Jyllands-Posten

European Commission

...a challenge for all MT engines

Training neural MT engines

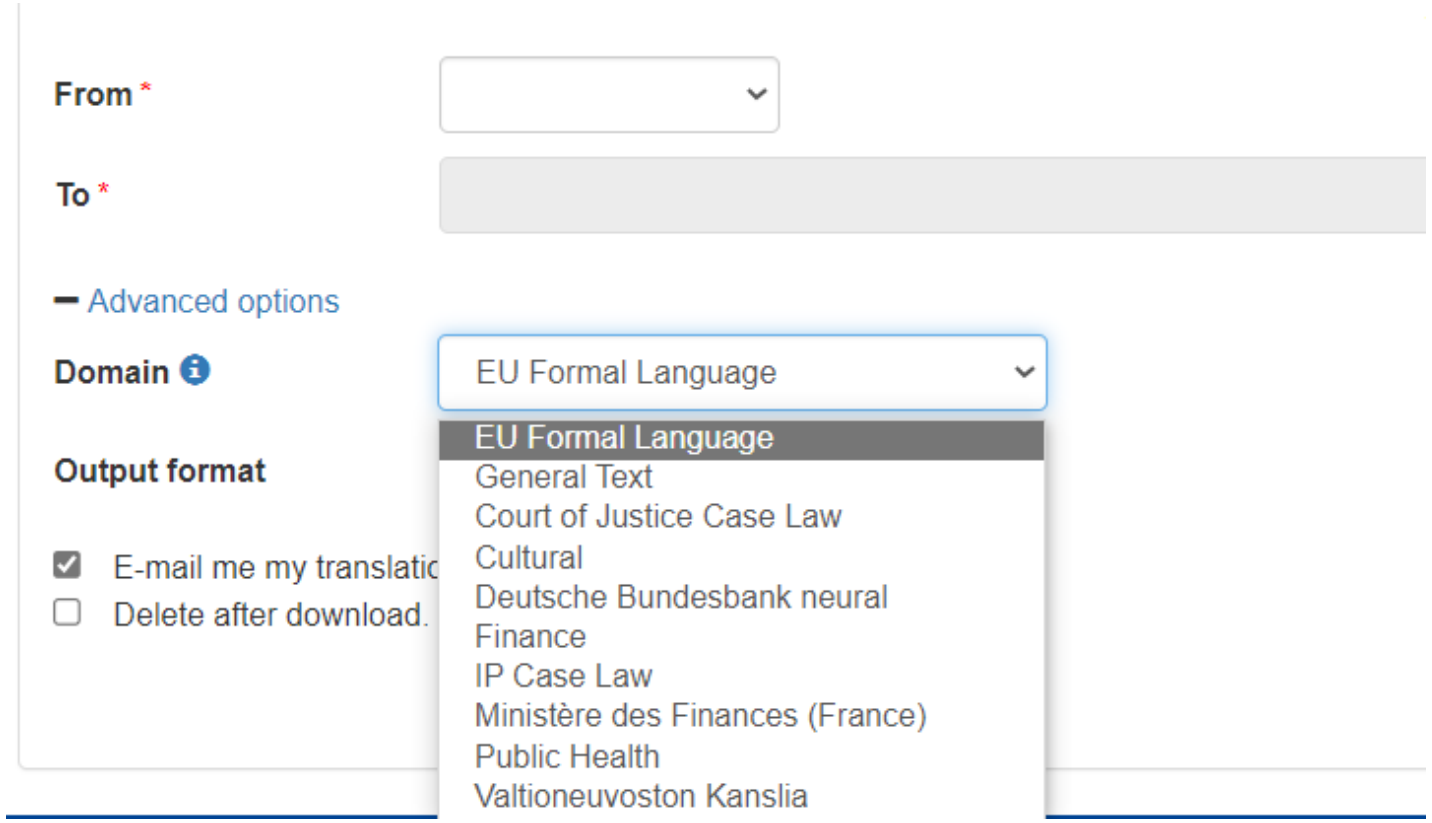
- All systems use similar software
- Quality depends on the available data and the ‘cleaning’ thereof
- Training an engine takes several days
- Two separate translation engines for each language pair are trained, e.g.:

  English -> Danish

  Danish -> English
- 254 new or upgraded engines in 2020

European Commission

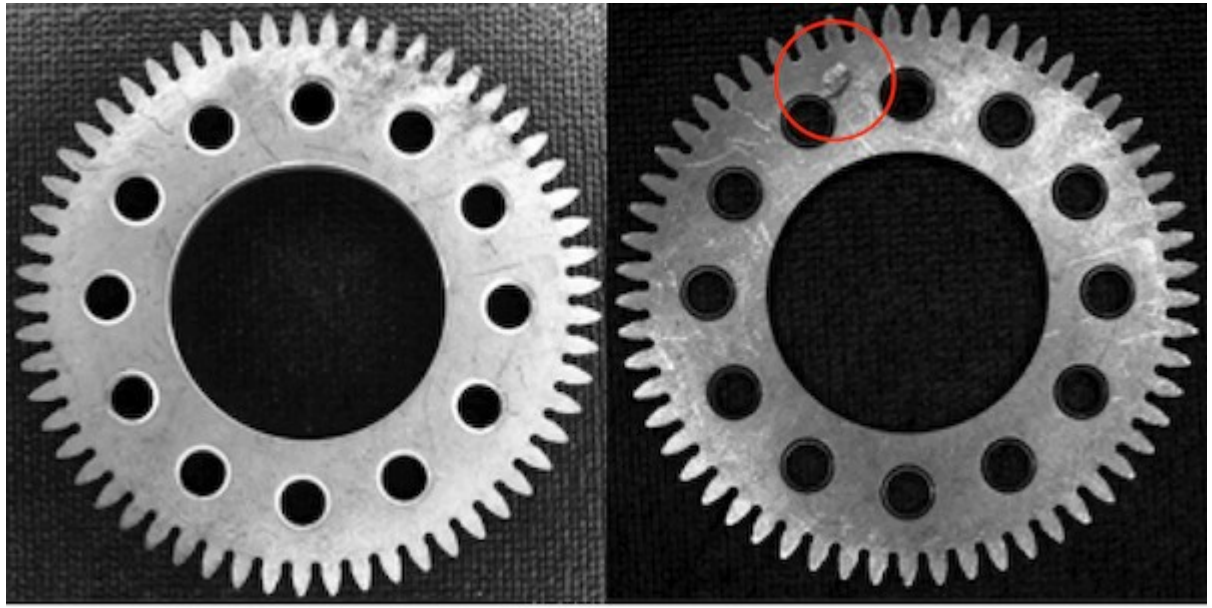
Domain-specific engines



eTranslation – toolkit and data

European Language Resource Coordination

Quality data/data curation



Photos courtesy of David Dechow and Landing AI

Whether photos or languages: Best results obtained with clear and correct data.

European Commission

Legislative vs. general text

- General language: Several linguistically correct translations

- Legal language: Only one legally correct translation

In legislative documents

- quotes must be verbatim
- defined terms must be used consistently

Moreover, a legal act is often composed of bits and pieces from other legal acts (Regulations, Directives, Decisions, EU Treaty).

European Commission

Consistent use of defined terms

Article 2

Definitions

For the purpose of this Regulation, the following definitions shall apply:

(1) 'fresh meat' means fresh meat as defined in point 1.10 of Annex I to Regulation (EC) No 853/2004;

(2) 'meat preparations' means meat preparations as defined in point 1.15 of Annex I to Regulation (EC) No 853/2004;

(3) 'meat' means meat as defined in point 1.1 of Annex I to Regulation (EC) No 853/2004;

Definitions

Artikel 2

Definitioner

I denne forordning forstås ved:

1) »fersk kød«: fersk kød som defineret i punkt 1.10 i bilag I til forordning (EF) nr. 853/2004

2) »tilberedt kød«: tilberedt kød som defineret i punkt 1.15 i bilag I til forordning (EF) nr. 853/2004

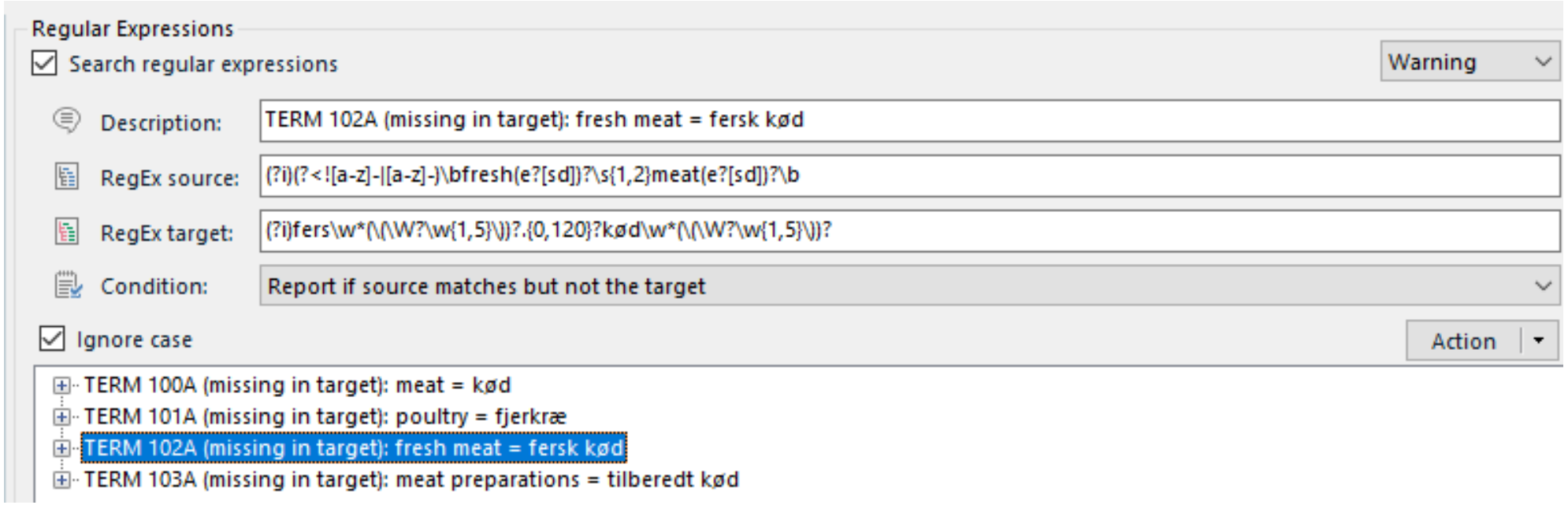
3) »kød«: kød som defineret i punkt 1.1 i bilag I til forordning (EF) nr. 853/2004

Quality Assurance Tools

- Built-in and run as verification checks in CAT tool
- Ruby Checker (checking numbers/dates are correctly transferred)
- Formal QA rules for all target languages
- TatraTerm (checking that defined terms are correctly translated)

European Commission

RegEx (Regular Expressions)



QA rules (formalities/style guide)

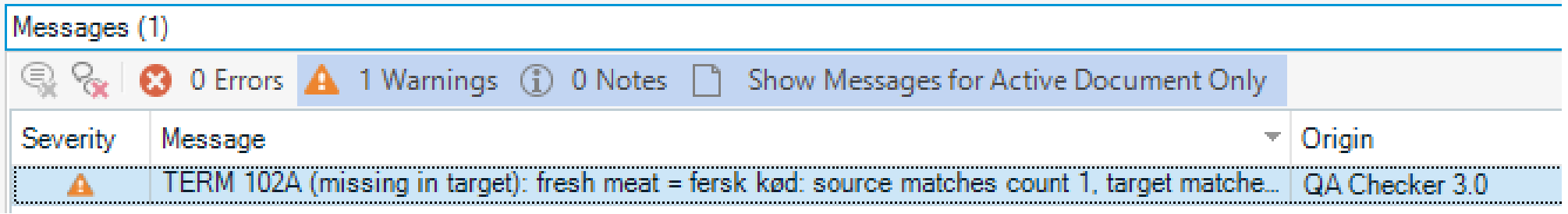
- 284n 3-letter ISO country or currency code found in ORI but not in TRA (too few/too many)
- 285n 3-letter ISO country or currency code found in TRA but not in ORI
- 290DA d.m.yyyy oversættes ikke i QE'er
- 300DA EFT i stedet for EUT?
- 301DA EUT i stedet for EFT?
- 302DA "nr." skal stå efter "(EF)" eller "(EU)"
- 303DA Forordninger fra 2015 el. senere har ikke "nr.".
- 304DA Mangler "nr." før forordningens nummer? For forordninger før 2015
- 305DA Tjek sidetal i danske EUT af 30.4.2004 (se staldtip)
- 306DA Brug lige citationstegn " eller lige apostrof '
- 307DA Brug EUR i stedet for "€"
- 308DA "EØS-Udvalg" skrives med stort U
- 309DA Forordning 1093-5/2010? "anmode om udtalelse", ikke "indhente"
- 310DA "s" mangler i "gennemførelseSmæssig"
- 311DA "supplementing" i retsaktstitler = om supplerende regler (se staldtip)
- 312DA Manglende punktum eller ekstra mellemrum før årstal i EUT-reference
- 313DA punktum mangler måske efter EUT-henvisning i fodnote
- 314DA Er der indsat forkert mellemrum mellem bogstav og tal i artikelnr., punkt o.lign.?
- 315DA Der mangler nok "r" i "opfoRdre"

European Commission

Quality Assurance Tools

- Checking that defined terms are translated as they should:

1	*Article 3*	100%	*Artikel 3*	P
2	**List of third countries or regions thereof authorised for the entry into the Union of fresh meat and meat preparations of ungulates**		**Liste over tredjelande eller regioner heri, hvorfra det er tilladt at indføre fersk kød og tilberedt kød af hovdyr til Unionen**	P
3	Consignments of fresh meat and meat preparations of ungulates intended for human consumption shall only be authorised for the entry into the Union if they come from the third countries or regions thereof authorised for the import into the Union in accordance with point (a) of Article 14 of Regulation (EU) No 206/2010.		Sendinger af frisk kød og tilberedt kød af hovdyr til konsum må kun indføres til Unionen, hvis de kommer fra tredjelande eller regioner heri, hvorfra import til Unionen er tilladt i henhold til artikel 14, litra a), i forordning (EU) nr. 206/2010.	P



European Commission

European Commission

AI is a great help for professionals...

European Commission

Thanks for the attention - Keep in touch

https://webgate.ec.europa.eu/etranslation/public/welcome.html (register as eTranslation user)

https://www.facebook.com/translatingforeurope/
(about translation in the EU institutions)

claus.larsen@ec.europa.eu

Claus Thornby Larsen

European Commission