

# Platforma eTranslation a MIE @ lucru

**Dan Tufiş**

Institutul de Cercetare pentru Inteligență Artificială

”Mihai Drăgănescu”

Academia Română

**Alexandru Ceauşu**

DG CNECT

Directorate-General for Communications

Networks, Content and Technology

*Traducere și adaptare după un scenariu de*

*Markus Foti*

*eTranslation Program Manager, DGT R.3*

# Argumente pentru TA

- Acces rapid și ieftin la informația și serviciile în limbi străine.
- Înțelegerea informației primite într-o limbă străină care altminteri nu ar putea fi folosită sau ar necesita timp și bani pentru a o traduce
- Crearea și/sau exploatarea siturilor web multilinguale
- Facilitarea căutării și analizei cros-linguale a informației
- Poate fi utilizată direct de oameni (interfață web) și indirect, mediate de masini (servicii SOAP, RESTful)...

CE, cel mai mare consumator de traduceri și alte servicii multilingve, a investit și continuă să investească în acest domeniu



# Platforma eTranslation (breviar)

## MT@EC

- Stil juridic
- MT Statistică

## eTranslation

- Motoare adaptate la domeniu
- Interfață web îmbunătățită
- Integrare în servicii online publice
- Motoare neurale
- Bazat pe Cloud

## CEF AT

- Instrumente NLP suplimentare precum recunoașterea entității menționate, transliterație și multe altele
- Servicii generice

# Interfața MT@EC



MT@EC – Traducere automată

Comisia Europeană

Despre MT@EC | Sondaj | Ajutor



Traducere fișiere

Traducere fragmente

Cererile mele

Setările mele

română



Deconectare

Selecțați cel puțin un fișier-sursă:

Alegeți fișierele...

SAU

Glisați și fixați fișierele aici.

Traducere din...

Traducere în...

Format traducere:

Identic cu fișierul sursă

TMX

XLIFF

Trimite-mi traducerea prin e-mail.

Șterge după descărcare.

Tradu fișierul

Fișiere selectate

# Interfața eTranslation

Drop files to upload (or click)

Supported formats: 

From\*

E ▼

To\*

BG

CS

DA

DE

EL

ES

ET

FI

FR

GA

HR

HU

IS

IT

LT

LV

MT

NB

NL

PL

PT

RO

SK

SL

SV

Domain

Cutting edge (default) ▼

Output format

Same as source

TMX

XLIFF

Translate document

Se încarcă  
documentul...



EUROPEAN COMMISSION

MEMO

Brussels, 6 May 2013

**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.

Se obține  
traducerea sa prin  
e-mail.



COMMISSION EUROPÉENNE

NOTE

Bruxelles, 6 mai 2013

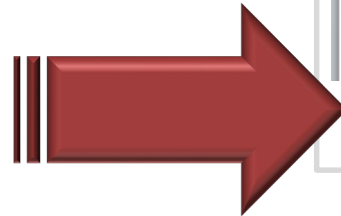
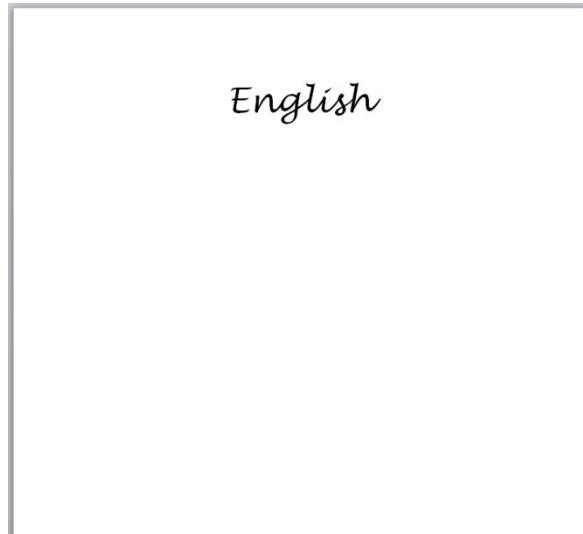
**Règles plus intelligentes pour des denrées  
plus sûres»: Paquet décisif C  
moderniser, de simplifier  
agroalimentaire en Europe**

**1) Ce qui est «de la fo**

L'UE «de la fourche à la fo  
santé en faisant en sorte  
que finale sur notre  
législatif est un syst  
maillons de la chaî  
nuisibles, la prote  
partir de ces ma  
tous les danger  
comprend égale  
de manière ha  
consommateur

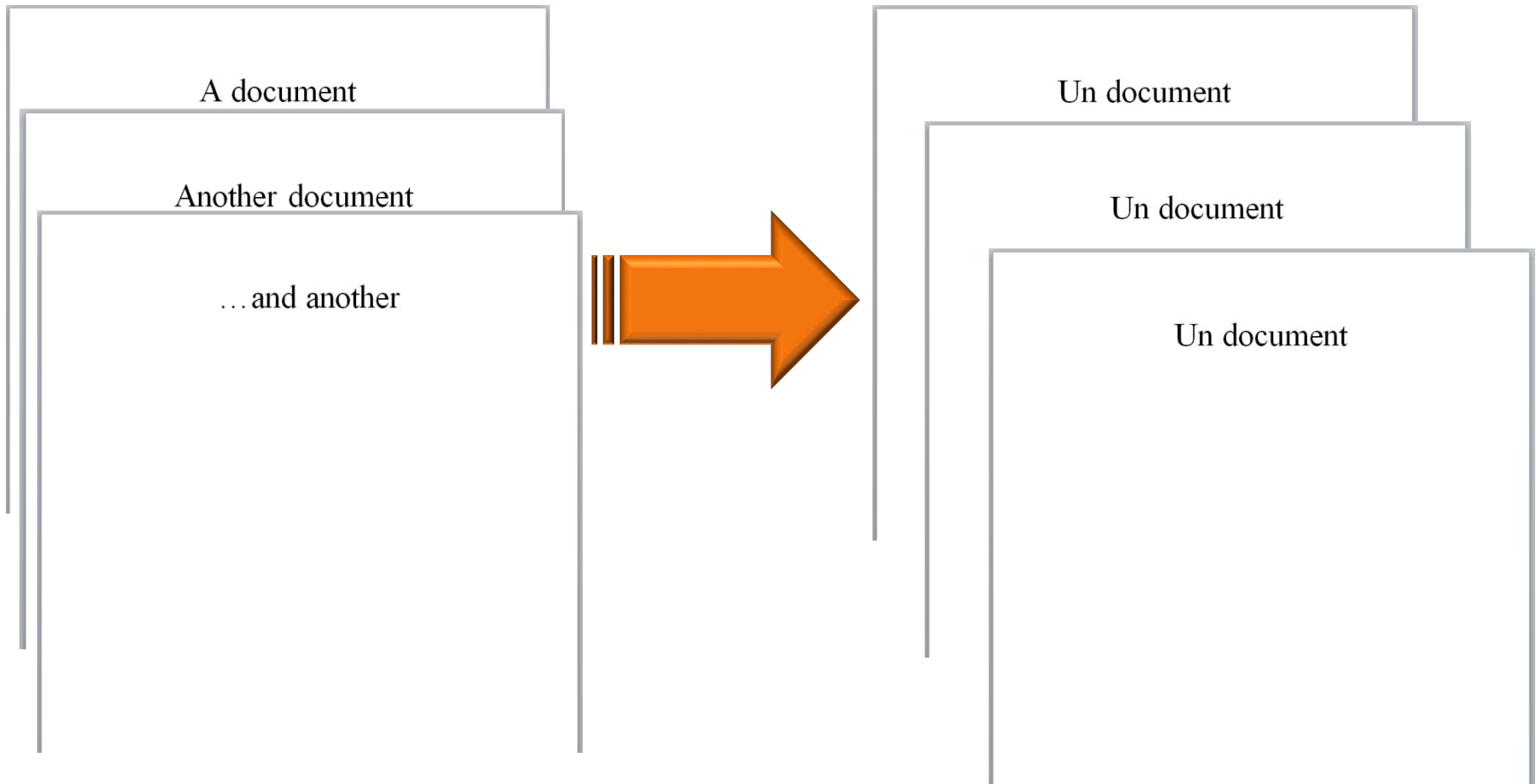
**Păstrează  
formatarea  
originală!**

Un original,  
mai multe traduceri...





# Mai multe documente traduse odată...



# Platforma eTranslation

- Îmbunătățește și dezvoltă performanțele MT@EC
    - Sporirea vitezei și a productivității traducerilor, reducerea costurilor
    - Lansat în iulie 2017 (webservice pt. snippets)
    - Nov. 2017 (web page pt. documente)
    - Soluție bazată pe ”cloud”
    - **Motoare de traducere neurală**
  - Disponibilă pentru:
    - Persoane din administrațiile publice (transmit documente printr-o pagină web)
    - Comunicare între mașini (cereri SOAP, interfețe RESTFUL)
- Exemple tipice de comunicare multilinguală mediată de eTranslation:
1. interacțiunea cu infrastructurile de Servicii Digitale (EESSI, ODR, Open Data Portal, Europeana, etc.)
  2. Arhive de date multilinguale (EURLex, N-Lex, Internal Market Information system...)

## eTranslation vă protejează intimitatea

- Toate documentele sunt șterse după 24 ore sau imediat după livrarea traducerii (la cerere)
- Drepturile de proprietate asupra traducerii vă sunt asignate doar dumneavoastră

# Traducerea automată neurală

- Învățare automată: rețele neurale sunt antrenate din traduceri existente
- O abordare radical diferită de cele bazate pe modele statistice
- Ele induc reprezentări ascunse ale textelor și exploatează generalizări surprinzătoare, ce scapă modelatorului uman.
- Traducerile sunt mai bune: mai fluente și mai corecte gramatical
- Reușesc mai bine să compenseze insuficiența datelor relevante pentru antrenare
- Prelucreează mult mai bine limbile cu morfologie dificilă (ex. Germană, Ungară)

Aceasta este direcția în care evoluează domeniul traducerii Automate

- Vor deveni traducerea automată bazată pe reguli sau traducerea automată statistică desuete? Nu neapărat...

# Motoare de traducere neurală

În-și-Din EN:

<b>BG</b>	<b>CS</b>	<b>DA</b>	<b>DE</b>	<b>EL</b>	<b>ES</b>	<b>ET</b>	<b>FI</b>	<b>FR</b>
<b>GA</b>	<b>HR</b>	<b>HU</b>	<b>IS</b>	<b>IT</b>	<b>LT</b>	<b>LV</b>	<b>MT</b>	<b>NB</b>
<b>NL</b>	<b>PL</b>	<b>PT</b>	<b>RO</b>	<b>SK</b>	<b>SL</b>	<b>SV</b>		

*24 iunie 2018*

# SMT vs NMT

Hungarian Original: **Hatalmas bevételt titkolnak el a vállalkozások**

Az online kassza önmagában nem hozhat, és nem is hozott érdemi változást az eltitkolt bevételek csökkentésében – legalábbis erre a következtetésre jutottak a lapunk által megkérdezett szakértők a Nemzeti Adó- és Vámhivatal (NAV) új akciósorozata, az úgynevezett adótraffipax sikere láttán.

*Magyar Nemzet*, 17 July (<https://mno.hu/gazdasag/hatalmas-bevetelt-titkolnak-el-a-vallalkozasok-2408682>)

MT@EC: **The huge revenue being hidden**

Online **pot may not in itself, and not the substantive change to reduce hidden revenue** at least **this was the conclusion reached** by the experts consulted by the debates of the National Tax and Customs Administration (NTCA), the so-called **akciósorozata adótraffipax** success.

eTranslation: **Huge revenue will be hidden for businesses**

**The online retention cannot, and did not, make a significant difference** in the reduction of the hidden revenue — at **least this conclusion was concluded** by the experts consulted by us and the new action of the National Tax and Customs Administration (NAV) — **the success of the so-called “tax force”**.

# eTranslation: comparație SMT/NMT

## Sursă: EN

This will require that the Greek authorities **maintain ownership of reforms and ensure their sustained implementation**, as per their commitments at the Eurogroup meeting of 22 June 2018.

In its 2013 Recommendation on 'Investing in Children', the Commission called on **Member States to "adapt the design and eligibility criteria of childcare services to increasingly diverse working patterns"**.

## Țintă: RO (SMT)

Aceasta presupune ca autoritățile elene **să mențină nivelul de asumare a reformelor și asigurarea aplicării lor susținute**, în conformitate cu angajamentele pe care și le-au asumat în cadrul reuniunii Eurogrupului din 22 iunie 2018.

În recomandarea sa din 2013 privind „Investiția în copii”, Comisia a invitat statele **membre să „adapteze” criteriilor de concepere și eligibilitate pentru serviciile de îngrijire a copiilor la muncă tot mai diversificate**”.

## Țintă: RO (NMT)

În acest scop, autoritățile elene **vor menține asumarea reformelor și vor asigura punerea în aplicare susținută a acestora**, în conformitate cu angajamentele asumate în cadrul reuniunii Eurogrupului din 22 iunie 2018.

În recomandarea sa din 2013 intitulată „Investiția în copii”, Comisia a invitat statele **membre „să adapteze” criteriile de concepere și eligibilitate ale serviciilor de îngrijire a copiilor la modele de lucru din ce în ce mai diverse.**”

# eTranslation: NMT fluență și reordonare sintactică naturală

## Source: EN

Rules to prevent the misuse of hosting services for the dissemination of terrorist content online in order to guarantee the smooth functioning of the internal market **are set out in this Regulation** in full respect of the fundamental rights ...

In nominal terms, the VAT Gap decreased by €10.5 billion to €147.1 billion in 2016, **a drop to 12.3% of total VAT revenues** compared to 13.2% the year before.

## Target: RO

**Prezentul regulament stabilește** normele menite să prevină utilizarea abuzivă a serviciilor de găzduire pentru difuzarea de conținut cu caracter terorist în mediul online, pentru a garanta buna funcționare a pieței interne, cu respectarea deplină a drepturilor fundamentale ...

În termeni nominali, deficitul de încasare a TVA a scăzut cu 10.5 miliarde EUR, ajungând la 147.1 miliarde EUR în 2016, **ceea ce reprezintă o scădere a veniturilor totale din TVA cu 12,3 %** față de 13,2 % în anul precedent.



# eTranslation: Probleme posibile

## Source: EN

In addition, the definition of "energy savings" in Article 4(16) of Romanian Law No 121/2014 **deviates** from the definition of the same term in the EED ...

It is this compulsive need for **unanimity** that is keeping us from being able to act credibly on the global stage.

As an attractive tourist destination in mainland Croatia thanks to its history and culture, Hrvatsko Zagorje **has also been drawing in scores of visitors** with its gastronomy based on traditional local products.

## Target: RO

În plus, definiția „economiiilor de energie” de la articolul 4 alineatul (16) din Legea nr. 121/2014 **nu se abate** de la definiția termenului din DEE ...

În **unanimitate**, este obligatoriu să se mențină posibilitatea de a acționa în mod credibil pe scena mondială.

Ca destinație turistică atractivă în Croația continentală datorită istoriei și culturii sale, Hrvatsko Zagorje **a obținut, de asemenea, scoruri ale vizitatorilor** cu gastronomia proprie pe baza produselor locale tradiționale.

# Cheia succesului: resursele lingvistice

- Mai multe date pentru toate limbile
- O acoperire lexicală mai bună pentru toate limbile
- Traducere automată adaptată domeniului / subiectului dumneavoastră de interes

# Următorul obiectiv: CEF.AT

- Va fi clădit pe serviciile existente în platforma eTranslation – dar cu facilități mult extinse și cu viteză sporită
- Va pune un accent deosebit pe traducerea automată **securizată, de calitate înaltă și adaptabilă** pentru serviciile online pan-europene – dar nu numai
- Asigurarea unei traduceri de calitate superioară în domenii specifice (începând cu domeniile relevante pentru DSI)
- Va deschide oportunități pentru sectorul privat, prin granturi și servicii generice, prin creșterea cererii pentru tehnologii lingvistice avansate

# Următorul obiectiv: CEF.AT

- Noi limbi (AR, RU, TK, ZH)
- Formate de ieșire suplimentare (pdf, json vs. tmx și xliif)
- Recunoașterea automată a entităților denumite
- Transliterare
- Datele colectate de ELRC vor fi puse la dispoziția celor interesați

→ Pentru atingerea acestor obiective, sunt necesare mai multe date de calitate pentru toate limbile în cât mai multe domenii.



# Cum vă puteți conecta la eTranslation

- Accesul deschis doar pentru infrastructuri de servicii digitale și administrații publice
  - Contactați [CEF-AT@ec.europa.eu](mailto:CEF-AT@ec.europa.eu) cu solicitarea de acces și credențialele de acces
  - Personalul CEF-AT vă va furniza documentația tehnică a procedurii de conectare, atât programatică (cereri SOAP) cât și prin interfața RESTful
  - Adaptați serviciul pentru utilizare multilinguală!

# CONCLUZII

- Dacă doriți ca eTranslation să lucreze optimal în limba română pentru domeniul dumneavoastră, acest lucru se poate face doar cu ajutorul resurselor lingvistice specifice!
- Refolosirea datelor existente (traduceri umane, memorii de traducere, dicționare și terminologii multilingve, etc.) este cel mai bun mijloc de a îmbunătăți calitatea traducerii automate
- ELRC are nevoie de implicarea noastră pentru a face, și pentru **limba română**, din CEF.AT un success
- Împreună vom reuși să obținem datele potrivite