The CEF Automated Translation Platform

Vilmantas LIUBINAS, computational linguist (DG Translation, European Commission)

24 June 2021



Connecting Europe Facility platform language tools

Translation (to 30 languages)



Speech-to-Text (EN, FR, DE)



Linguistic analysis of documents

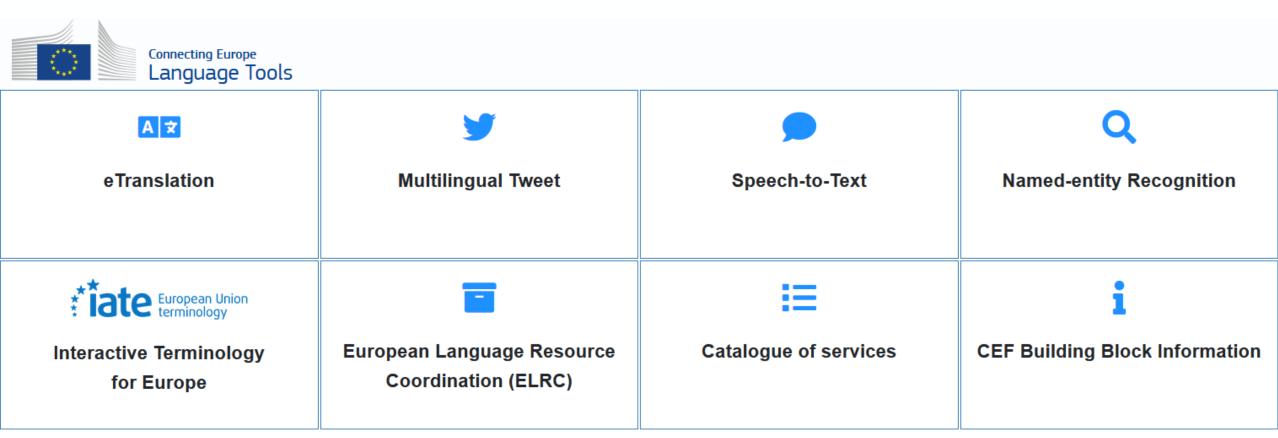


-> Recognize specific categories of words (EN, DE) -> Determine Eurovoc category





HTTPS://LANGUAGE-TOOLS.EC.EUROPA.EU







Who's IT FOR?





ONE SYSTEM — TWO AUDIENCES

- Public administrations in Member States, Iceland and Norway
- Universities
- SMEs
- CEF digital service infrastructures

CEF eTranslation

- Translators and staff of the EU Institutions
- Digital services of the EU Institutions

DGT eTranslation





A BIT OF HISTORY

MT@EC

- Legalese
- Statistical MT

eTranslation

- Neural MT
- Domain adaptation
- Integration with online public services

CEF.AT

- More LT tools, not just translation
- Supporting actions for data collection – ELRC
- Supporting funded projects





ETRANSLATION







eTranslation now available for English - Irish!







eTranslation language coverage

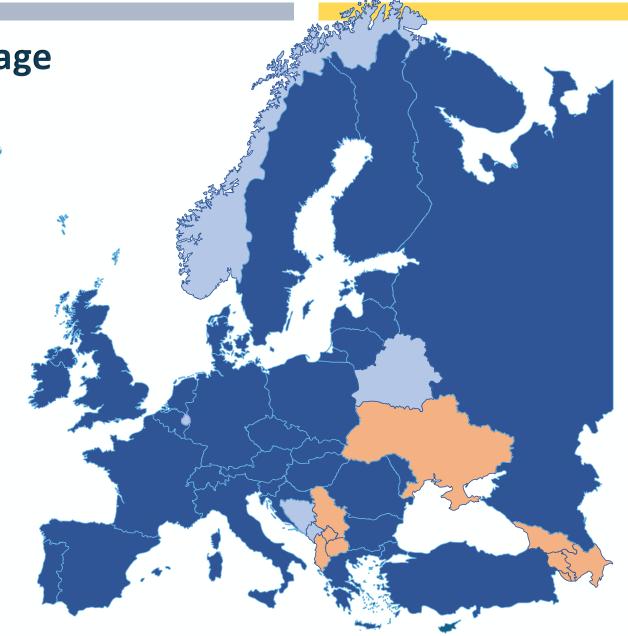




Partially covered

Not covered

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





eTranslation translation volume

2.2 miopages of
DGT human
translation

4.8 mio

additional machine translation pages for the Commission

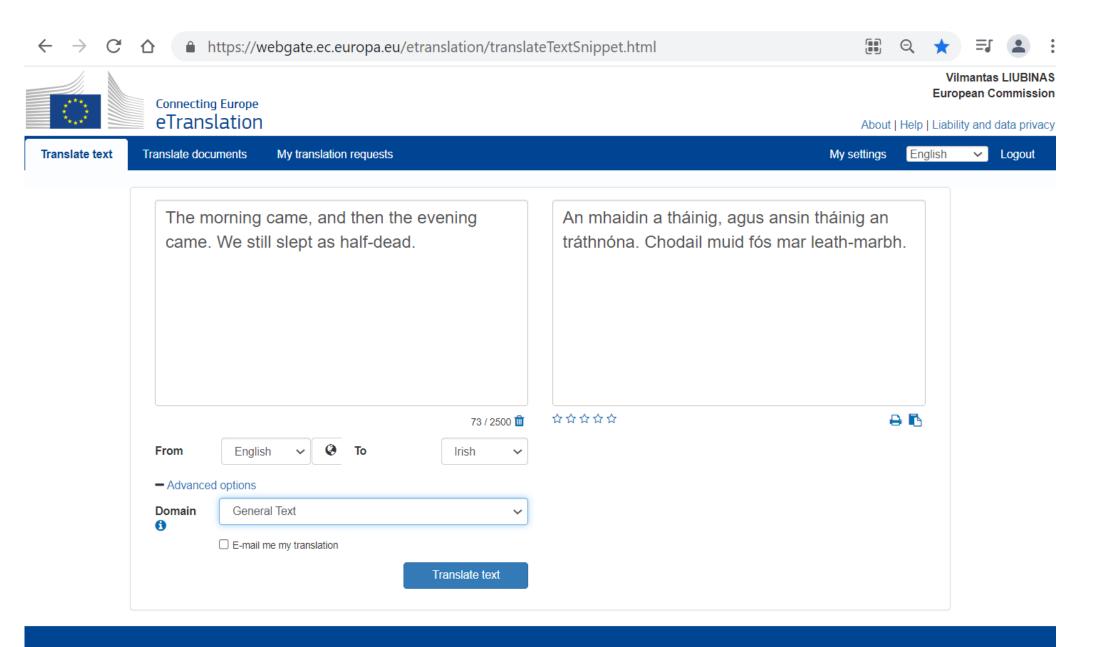
70 mio

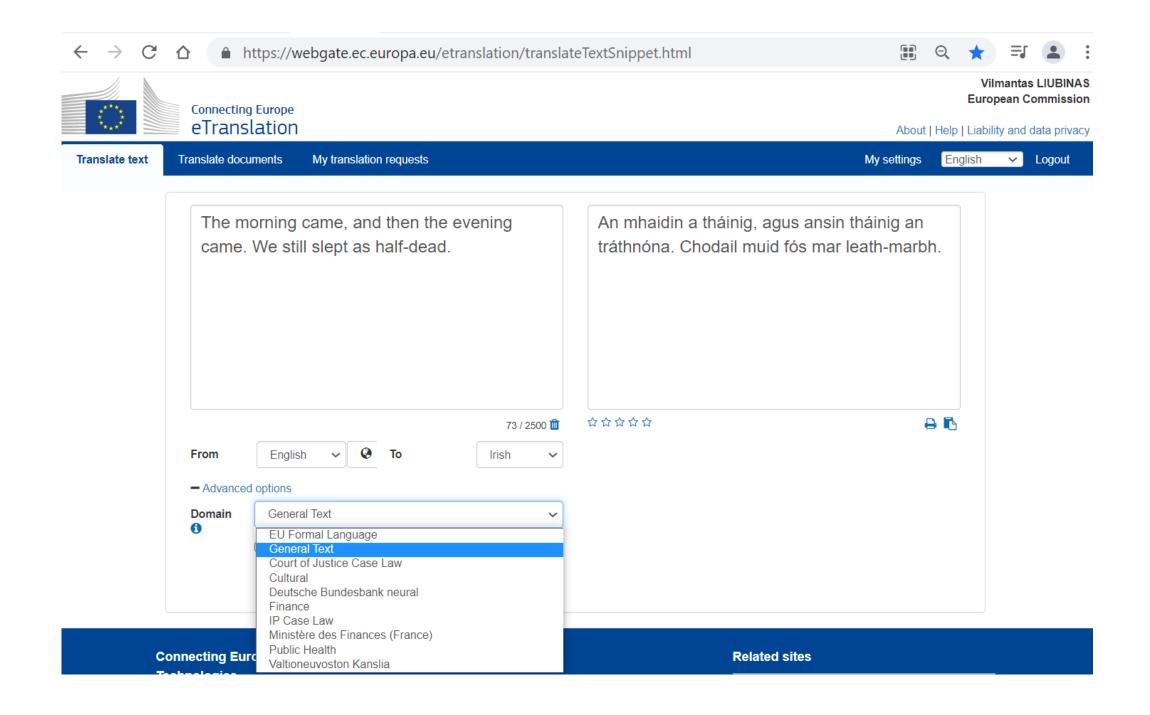
total eTranslation output in 2020



100 mio

total eTranslation output in 1st half 2021









eTranslation: two main translation domains:

★ 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



About | Help | Liability and data privacy

My translation requests English Logout Translate text **Translate documents** My settings Bulgarian act Bulgarian act Bulgarian 1.docx 2.docx Annex.xlsx **12.5** KB 12.5 KB m **12.5** KB Supported formats: 🔎 🕮 🔑 🤝 🗎 From * Bulgarian ~ To * Irish 🗙 Advanced options Domain 🕕 EU Formal Language XLIFF QE (Beta) Same as source TMX Output format E-mail me my translation Delete after download. Translate document

Scrios ón bhfreastalaí é

An tSeicis

☐ An tSlóivéinis

Eolas | Cabhair | Dliteanas agus príobháideachas sonraí

righ an doiciméad

Doiciméid a aistriú Mo chuid aistriúchán Mo chuid socruithe Gaeilge Logáil amach Téacs a aistriú Bulgarian act Bulgarian act Bulgarian 1.docx 2.docx Annex.xlsx **12.5** KB **12.5** KB 12.5 KB Formáidí a nglactar leo: 💹 💷 🔑 🤝 🗎 🔳 Ó* An Bhulgáiris \checkmark Go* Tá 23 dteanga as 30 roghnaithe X An Araibis An Béarla ☐ An Chróitis ☐ An Danmhairgis - Ardroghanna An Eastóinis An Fhionlainnis An Fhraincis An Ghaeilge Réimse 🕕 An Ghearmáinis An Ghréigis ✓ An Iodáilis An Iorua ✓ An Íoslainn ☐ An Laitvis ☐ An Liotuáinis An Mháltais Formáid an aschuir An Ollainnis An Pholainnis An Phortaingéilis An Rómáinis Seol an t-aistriúchán ch An Rúisis An Seapáinis An Sínis (simpli... An Spáinnis

An tSlóvaicis

An tSualainnis



Scrios ón bhfreastalaí é tar éis a íoslódála.

Eolas | Cabhair | Dliteanas agus príobháideachas sonraí

Aistrigh an doiciméad

Doiciméid a aistriú Mo chuid aistriúchán Mo chuid socruithe Gaeilge Logáil amach Téacs a aistriú Bulgarian act Bulgarian act Bulgarian 1.docx 2.docx Annex.xlsx **12.5** KB **12.5** KB 12.5 KB Formáidí a nglactar leo: 💹 🕮 🔑 🤝 🗎 📗 Ó* An Bhulgáiris ~ Go* Tá 23 dteanga as 30 roghnaithe 🗙 - Ardroghanna Réimse 🕕 EU Formal Language Mar an gcéanna leis an bhfoinse TMX XLIFF QE (Beta) Formáid an aschuir Seol an t-aistriúchán chugam i ríomhphost.

About | Help | Privacy policy

Translate text

Translate documents

My translation requests

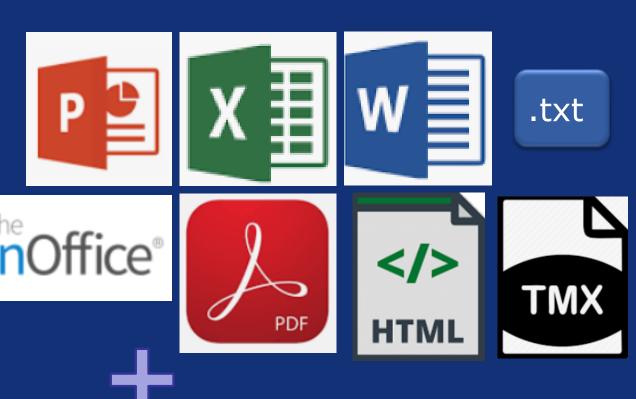
My settings

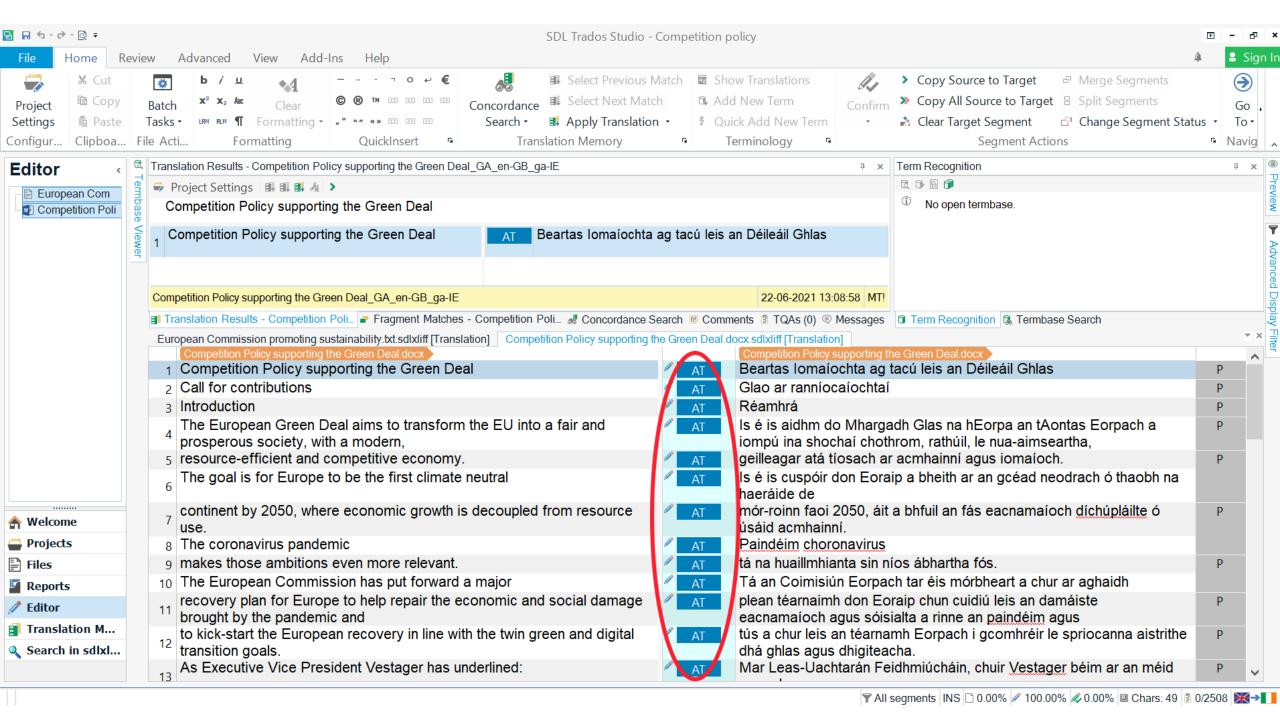
English



Logout

Supported formats:





About | Help | Privacy policy

Translate text Translate documents My translation requests My settings English V Logout

0

Documents are available for download for 24 hours

Translated file	Status	Message	Document format	Target format	Source language	Target language
Rapport de mission_FR.docx	Ongoing		docx	docx	English	French
Another document in Bulgarian_FR.d	Ongoing		docx	docx	English	French
Document in Bulgarian_FR.docx	Ongoing		docx	docx	English	French
PMO-2020-00018-00-00-PT-ORI-00_FR	Translated		pdf	docx	Portuguese	French
PMO-2020-00018-00-00-PT-PRT-00_FR	Translated		rtf	rtf	Portuguese	French
	Rapport de mission_FR.docx Another document in Bulgarian_FR.d Document in Bulgarian_FR.docx PMO-2020-00018-00-00-PT-ORI-00_FR	Rapport de mission_FR.docx Ongoing Another document in Bulgarian_FR.d Ongoing Document in Bulgarian_FR.docx Ongoing PMO-2020-00018-00-00-PT-ORI-00_FR Translated	Rapport de mission_FR.docx Ongoing Another document in Bulgarian_FR.d Ongoing Document in Bulgarian_FR.docx Ongoing PMO-2020-00018-00-00-PT-ORI-00_FR Translated	Rapport de mission_FR.docx Ongoing docx Another document in Bulgarian_FR.d Ongoing docx Document in Bulgarian_FR.docx Ongoing docx PMO-2020-00018-00-00-PT-ORI-00_FR Translated pdf	Rapport de mission_FR.docx Ongoing docx docx Another document in Bulgarian_FR.d Ongoing docx docx Document in Bulgarian_FR.docx Ongoing docx docx PMO-2020-00018-00-00-PT-ORI-00_FR Translated pdf docx	Rapport de mission_FR.docx Ongoing docx docx English Another document in Bulgarian_FR.d Ongoing docx docx English Document in Bulgarian_FR.docx Ongoing docx docx English PMO-2020-00018-00-00-PT-ORI-00_FR Translated pdf docx Portuguese

(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>GA</target-language>
- <domain>GEN</domain>
- <email-destination>firstname.lastname@ucd.ie/email-destination>



Client application in Dublin



<translated-text>Dia dhuit domhan!</translated-text>







WHAT ABOUT THE TRANSLATION QUALITY?





Most challenging areas:

- Creative texts
- Rare words
- Long sentences (100+ words)
- Individual words or phrases that depend on the context

What does translation engine take into account while translating?

Previous knowledge:	V	EN: The Winter war was harsher than the Cold War. DE: Der Winterkrieg war härter als der Kalte Krieg. (not "Winter Krieg") EN: Sun is a lovely dog. FR: Le soleil est un beau chien. (Sun est un beau chien)
Document context:	X	EN:a careful approach towards the control of <u>ISA</u> is needed. FR:une approche prudente du contrôle de <u>I'ISA</u> est nécessaire. (l'AIS - l'anémie infectieuse du saumon)
Paragraph context:	X	 EN:proof that a person has either: been vaccinated; recovered from COVID-19 ES:proporciona pruebas de que una persona tiene: vacunados contra COVID-19; (ha sido vacunada) recuperado de COVID-19. (se ha recuperado)
Sentence itself:		EN: Recovery did not take long for Thomas. FR: <u>La reprise</u> n'a pas pris de temps pour Thomas. EN: Recovery did not take long for creditors. FR: <u>Le recouvrement</u> n'a pas pris de temps pour les créanciers.

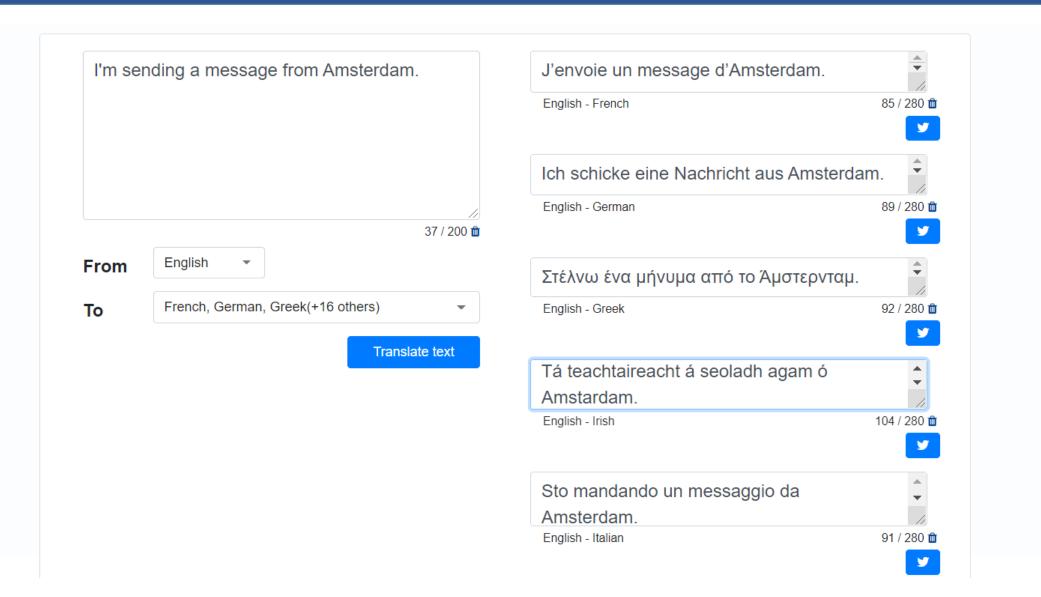




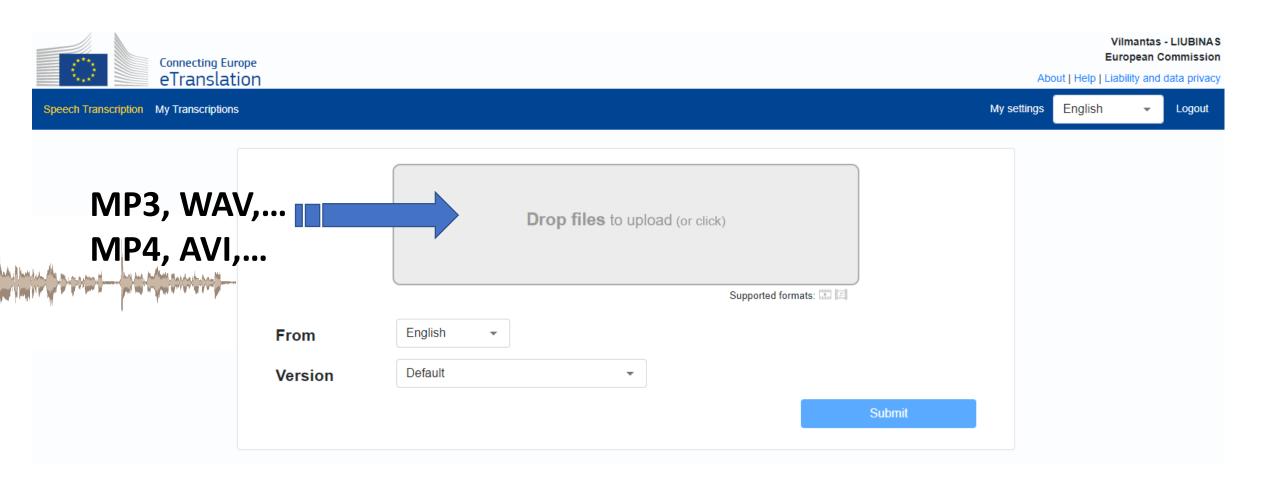
OTHER LANGUAGE TECHNOLOGY SERVICES

eTranslation multilingual tweet

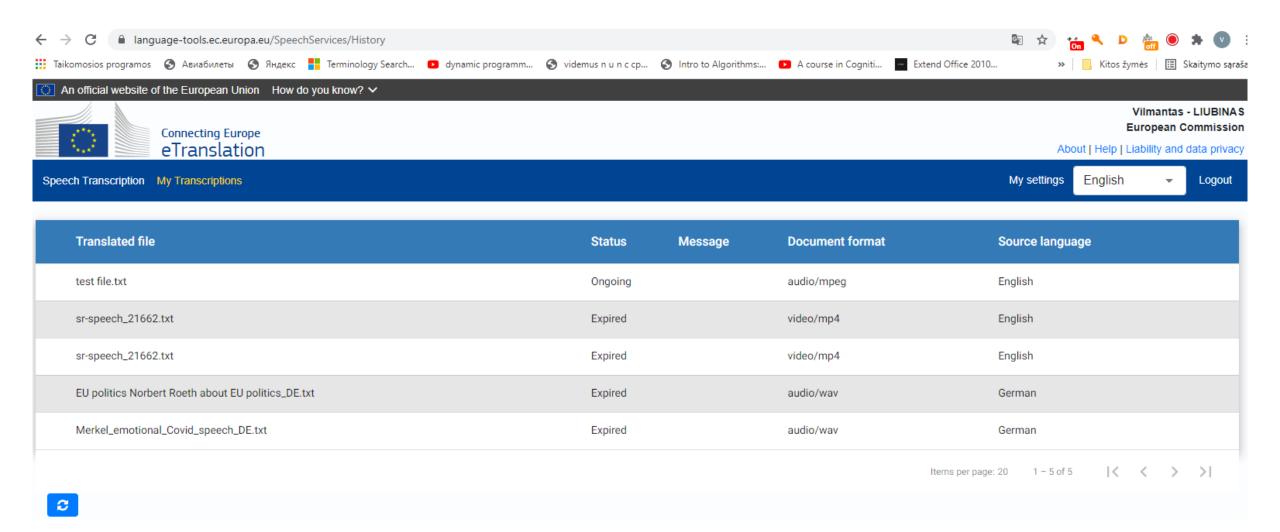




Speech-to-Text service



Speech-to-Text service



Speech-to-Text service

BBC ORIGINAL ARTICLE:

A teenage girl who donated her hair to a cancer charity found out a year later she has cancer herself. Kiya, 15, sat a GCSE exam on the same day she was diagnosed with stage four Hodgkin's lymphoma after discovering a lump in her neck. Twelve months after donating her hair to the Little Princess Trust, a charity that provides wigs to children with cancer, Kiya, from New Lubbesthorpe in **Leicestershire**, received her own wig from the charity following chemotherapy treatment. She's now raising money for them and other cancer charities and has had some famous supporters. Scotland captain and Liverpool left-back Andy Robertson told Kiya in a message: "It's so important for you to keep strong and keep fighting. You'll never walk alone.

CEF eTranslation platform TRANSCRIPTION:

A teenage girl who donated her hair to a cancer charity found out. A year later she has cancer herself. Kia, 15 years old girl, said at GCSE exam on the same day she was diagnosed with stage four Hodgkin's lymphoma after discovering a lump in her neck. 12 months after donating her hair to the Little Princess Trust, a charity that provides wigs to children with cancer. Kia from new Lubbesthorpe in life. Chester Shire received her own wig from the charity following chemotherapy treatment. She's now raising money for them and other cancer charities and has had some famous supporters, Scotland captain and Liverpool left back. Andy Robertson told Kia in a message. It's so important for you to keep strong and keep fighting. You'll never walk alone.

Document Categorization Tool

OUTPUT:

EuroVoc categories



INPUT:
Source document
(22 EU languages)

Zoo (0.12998)

Cooperation policy (0.11468)

Directory (0.11312)

Environmental protection (0.11241)

Community programme (0.11026)

Protection of animal life (0.09938)



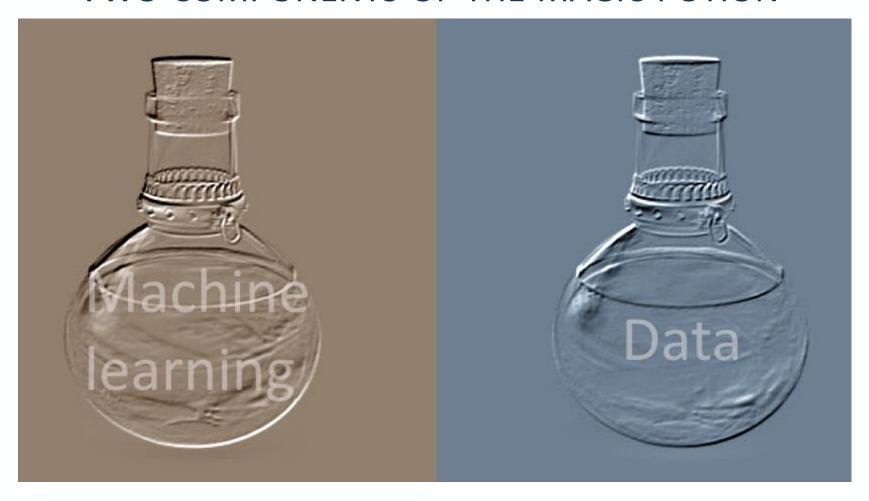


How is that possible?

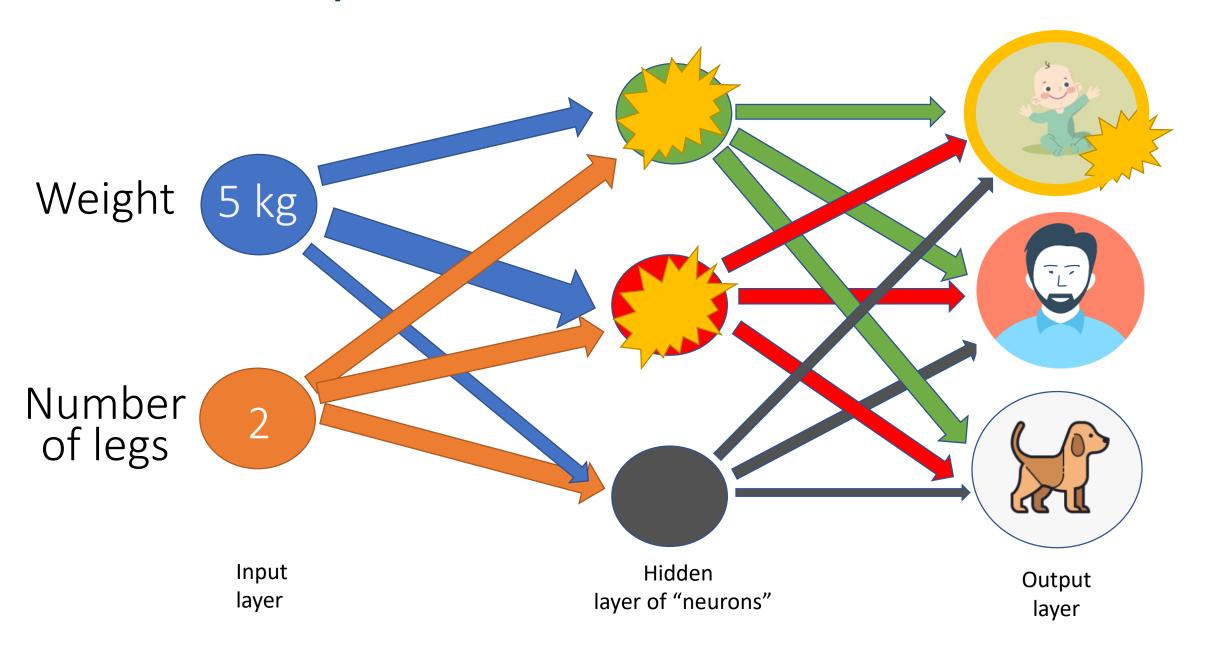




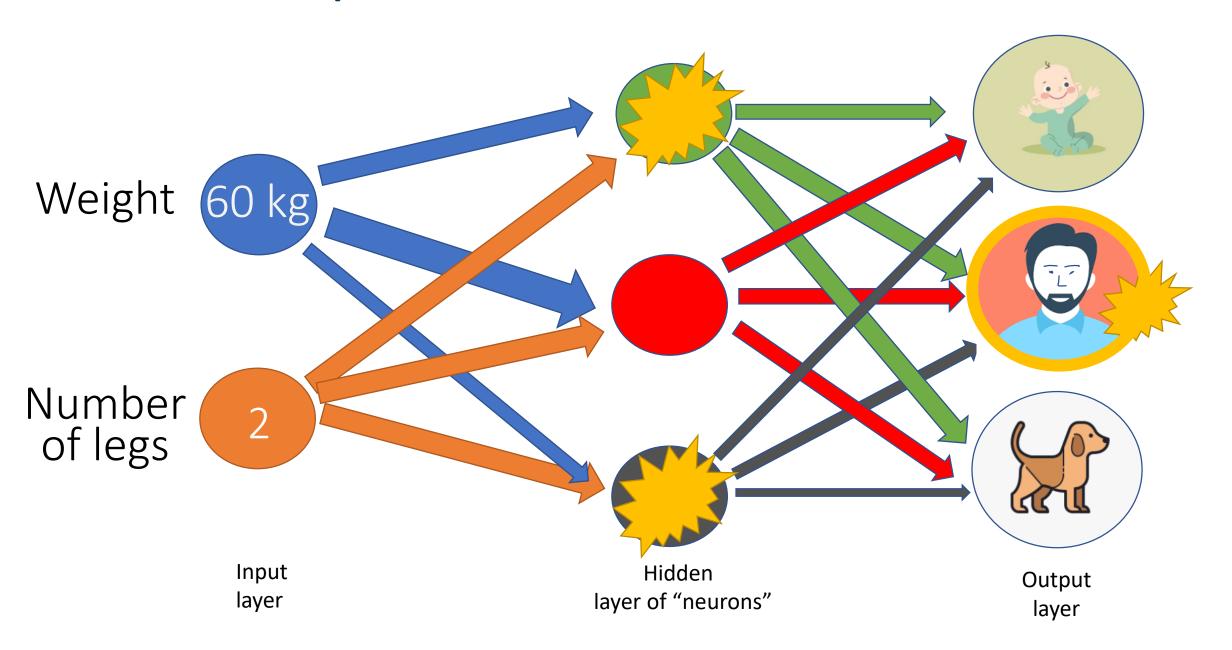
TWO COMPONENTS OF THE MAGIC POTION

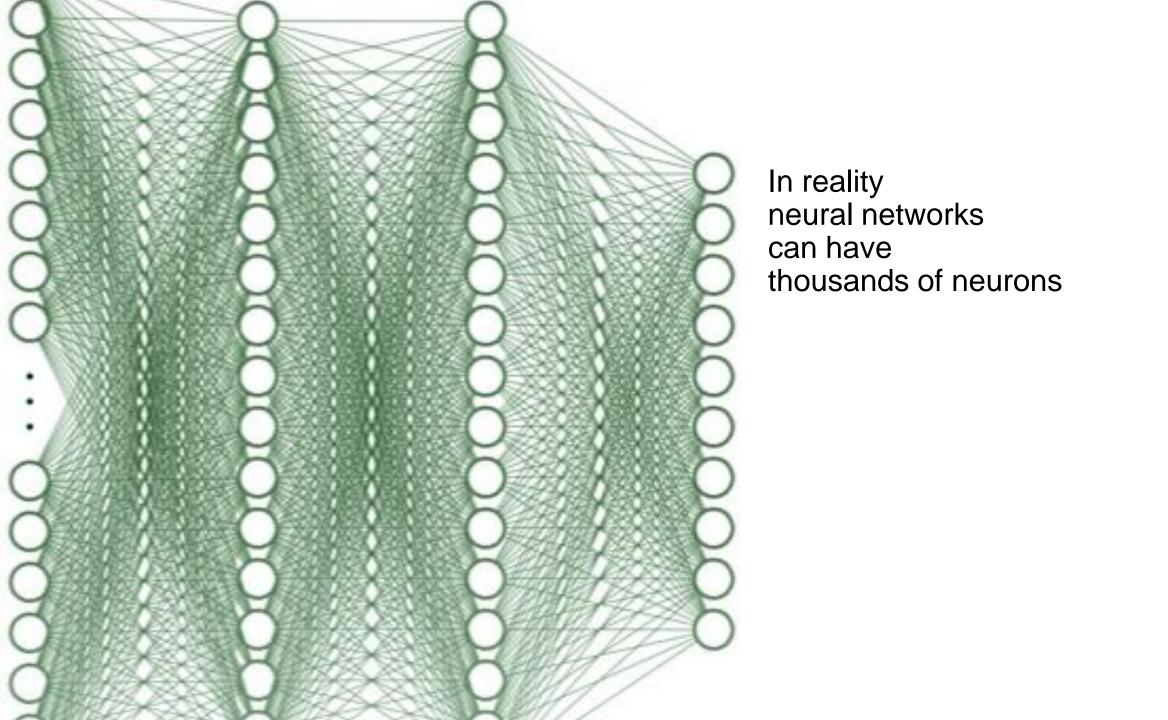


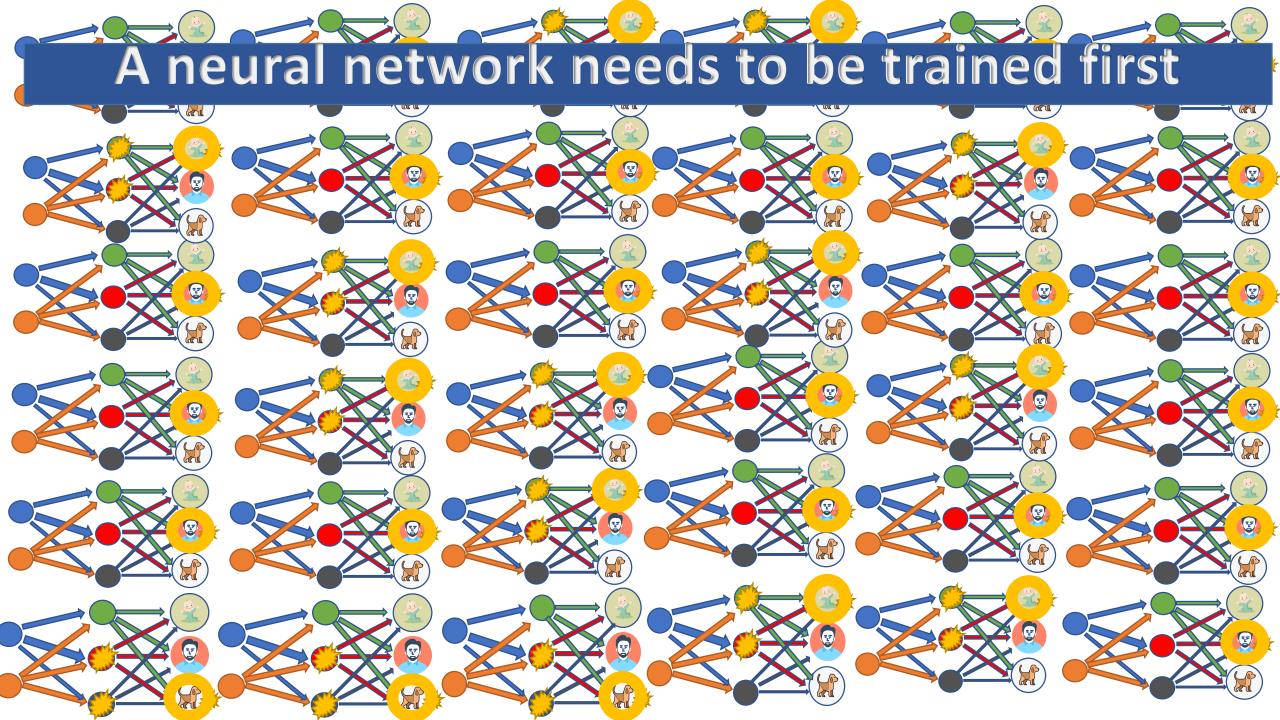
The simplest "trained" neural network in action



The simplest "trained" neural network in action











THE NEURAL TRANSLATION ENGINE

- Its training lasts several days
- Two separate translation engines for each language pair are trained:

English -> Irish

Irish -> English

- 254 new or upgraded engines in 2020
- Needs massive amounts of DATA!!!



Connecting Europe Facility platform language tools

Translation (to 30 languages)



Speech-to-Text (EN, FR, DE)



Linguistic analysis of documents



-> Recognize specific categories of words (EN, DE) -> Determine Eurovoc category

Connecting Europe Facility platform language tools

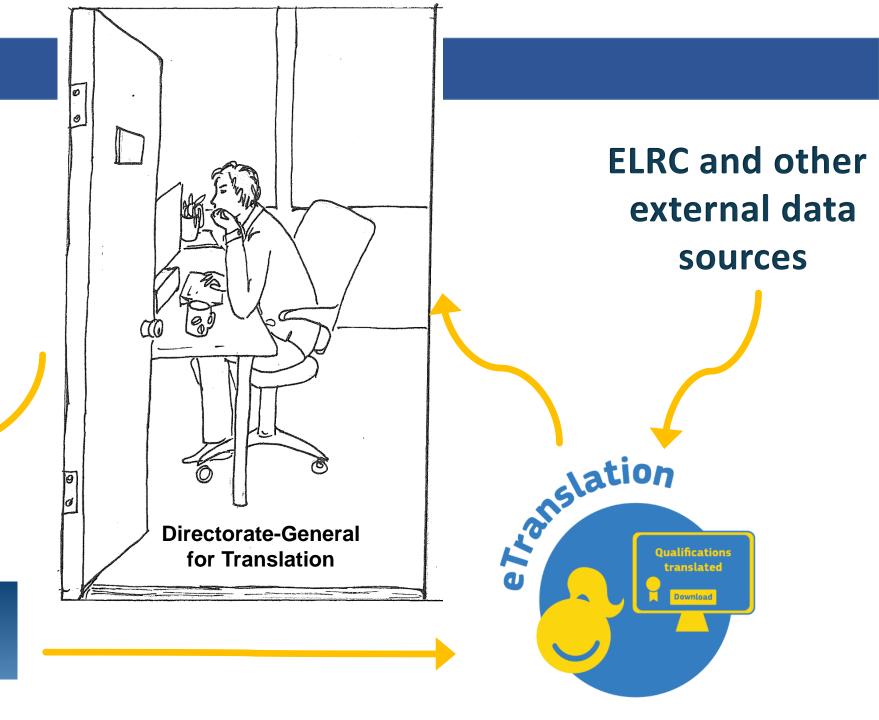


It all starts here:

Tosaíonn sé ar fad anseo:

2 200 000 pages in 2020

EURAMIS







Bilingual data: 10 000 000 Irish-English sentences from Euramis

Showing 30 hits / 30 ((more results available)				
☐ TM: Policies Doc. No.: 800000101 Req. Serv.: ECFIN Year: 2021 SL: EN (Direct)	Trans: foxshan				
EN: Directive 2008/56/EC establishing a framework for community action in the field of marine environmental policy ('Marine Strategy Framework' Directive) (OJ L 164, 25.6.2008, p. 19).	GA: Treoir 2008/56/CE lena mbunaítear creat do ghníomhaíocht chomhphobail i réimse an bheartais comhshaoil mhuirí ('an Treoir Réime um Straitéis Mhuirí') (IO L 164, 25.6.2008, lch. 19).				
☐ TM: Policies Doc. No.: 800000001 Req. Serv.: ECFIN Year: 2021 SL: EN (Direct)	Trans: hiarlao				
EN: Respectively Directive 2008/56/EC establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) (OJ L 164, 25.6.2008, p. 19);	GA: Na Treoracha seo faoi seach: Treoir 2008/56/CE lena mbunaítear creat do ghníomhaíocht chomhphobail i réimse an bheartais comhshaoil mhuirí ('an Treoir Réime um Straitéis Mhuirí') (IO L 164, 25.6.2008, lch. 19);				
TM: Policies Doc. No.: 400140000 Req. Serv.: JUST Year: 2021 SL: EN (Direct) Trans: omuirpo					
EN: Academic/research institution • Government authority in charge of environmental policy • National prosecutors and professional networks for prosecutors • National judges and professional network for judges • National law enforcement professionals or professional networks of law enforcement • Local/regional authorities • European or international bodies dealing with environmental crime • Non-governmental organisations (NGO) • Business/industry or business/industry associations • Consumer organisations and citizens	GA: Institiúid acadúil/taighde • Údarás an rialtais atá i gceannas ar bheartas an chomhshaoil • Ionchúisitheoirí náisiúnta agus líonraí gairmiúla le haghaidh ionchúisitheoirí • Breithiúna náisiúnta agus líonra gairmiúil le haghaidh breithiúna • Gairmithe náisiúnta um fhorfheidhmiú an dlí nó líonraí gairmiúla um fhorfheidhmiú an dlí • Údaráis áitiúla/réigiúnacha • Comhlachtaí Eorpacha nó idirnáisiúnta a bhíonn ag déileáil leis an gcoireacht chomhshaoil • Eagraíochtaí neamhrialtasacha (NGO) • Lucht gnó/tionscail nó comhlachais ghnó/tionscail • Eagraíochtaí tomhaltóirí agus saoránaigh				
☐ TM: Policies Doc. No.: 800280101 Req. Serv.: MARE Year: 2021 SL: EN (Direct)	Trans: haogaea				
EN: 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)	GA: Treoir 2008/56/CE ó Pharlaimint na hEorpa agus ón gComhairle an 17 Meitheamh 2008 lena mbunaítear creat do ghníomhaíocht Chomhphobail i réimse an bheartais comhshaoil mhuirí (An Treoir Réime um Straitéis Mhuirí)				
☐ TM: Policies Doc. No.: 800280000 Req. Serv.: MARE Year: 2021 SL: EN (Direct)	Trans: haogaea				
EN: 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)	GA: Treoir 2008/56/CE ó Pharlaimint na hEorpa agus ón gComhairle an 17 Meitheamh 2008 lena mbunaítear creat do ghníomhaíocht Chomhphobail i réimse an bheartais comhshaoil mhuirí (An Treoir Réime um Straitéis Mhuirí)				
☐ TM: Policies Doc. No.: 801120301 Req. Serv.: MARE Year: 2021 SL: EN (Direct)	Trans: niciogi				
EN: 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) (OJ L 164, 25.6.2008, p. 19).	GA: Treoir 2008/56/CE ó Pharlaimint na hEorpa agus ón gComhairle an 17 Meitheamh 2008 lena mbunaítear creat do ghníomhaíocht chomhphobail i réimse an bheartais muirthimpeallachta (an Treoir Réime um Straitéis Mhuirí) (IO L 164, 25.6.2008, lch. 19).				
☐ TM: Policies Doc. No.: 800000301 Req. Serv.: ECFIN Year: 2021 SL: EN (Direct)	Trans: dgtmany				
EN: Directive 2008/56/EC establishing a framework for community action in the field of marine environmental policy ('Marine Strategy Framework' Directive) (OJ L 164,	GA: Treoir 2008/56/CE lena mbunaítear creat do ghníomhaíocht chomhphobail i réimse an bheartais comhshaoil mhuirí ('an Treoir Réime um Straitéis Mhuirí') (IO L 164, 25.6.2008,				





Data matters and cannot be left to chance



Is cineál faoi leith iad, iad níos cothaithí agus níos saibhre i Vitimín A Bíonn siad ar oscailt idir 10.00 a.m. agus 4.30 p.m. ar ghnáthlaethanta oil Déantar iniúchadh agus grinnscrúdú orthu trí iniúchadh inmheánach (link to Tá siad eagraithe de réir bailte fearainn. Deartháireacha agus deirfiúracha linn iad. Ina theannta sin, íocann Éire €75 milliún in aghaidh na bliana le buiséad o Déantar iad a choigeartú tar éis an íocaíocht dheireanach a fháil, nuair a Déantar iad a earcú ó chomórtas oscailte, nach féidir le státseirbhísigh pá Tá siad ag fáil réidh leis na seanmhódhanna cócaireachta, lenar bhain pota Tá an fostaí sin ina gcónaí i dtír a bhfuil comhaontú um Chánachas Dúbailte Tá siad freagrach as áiteanna poiblí atá faoina rialú, lena n-áirítear bóid Tá freagrach orthu maidir le riarachán laethiúil na Comhairle. Bíonn siad ag imeacht ina naonar nó i mbeirteanna. Tá siad cosúil leis na horduithe ar féidir iarratas a dhéanamh orthu i bpró Tá siad cosúil leis na horduithe ar féidir iarratas a dhéanamh orthu i bpró Tá siad cosúil leis na horduithe ar féidir iarratas a dhéanamh orthu i bpro Úsáidtear iad ar gach suíomh gréasáin beagnach, agus ní chuireann siad iste Foirmeacha a éilítear le clárú Tá siad faoi réir athchóirithe nuair a bheidh figiúirí dheireadh na bliana Tá siad tacaithe agus maoinithe ag an mBord um Fhaisnéis do Shaoranaigh. That is one of the advantages of having a body like this to discuss such ma Is é atá iontu ná taifid charntha de chuid duine aonair nó institiúide, ar Is iad na bonnchlocha iad, agus na dearbhchlocha chomh maith, ar a mbunaíte Is de shliocht na bhfianna buí iad, a tugadh isteach sa pháirc sa seachtú l Scaoiltear amach anisin ar bannaí iad áit ar dócha go nglacfaidh siad páir Ansin measctar le chéile go maith iad agus sórtáiltear iad de réir na vóta: Ansin measctar le chéile go maith iad agus sórtáiltear iad de réir na vóta: Ullmhófar iad de réir fhorálacha an Achta Toghcháin, 1997 (arna leasú), de Beidh siad mar uain íobartacha a thabharfar don mhac tíre i mBaile Átha Cl Tá siad oilte chun:

Bíonn siad ina n-aonar agus ní bhíonn siad faiteach.

Tá diansaothar ag baint leo agus tógann siad am.

D'oirfeadh sé dóibh iad a chur i ndiaidh na praisí (brocailí nó cabáis Baintear úsáid astu go coitianta chun suíomhanna gréasáin a chur ag obair,

Tá siad scríofa i bhformáid ceist agus freagra agus freagraíonn siad na cei
Is iad sin:
Is iadsan:

Déanann siad an argóint nach féidir le hAirteagal 8 na dualgais a éiríonn f
Rinne siad araon cur síos ar na dúshláin a bhaineann le reachtáil Chumann G
Cheannaigh siad teach i Londainagus bhí seisear páistí acu.

Tugann siad breithiúnas neamhspleách maidir le ceisteanna um straitéis, fei
Mionnaím duit nach ionann teaghlach nach bhfuil slándáil bia acu agus teagh
Sháraigh siad fórsaí an Rí agus bhog siad ó Chnoc Staball síos Sráid Thomái

They are of a special variety, more nutritious and rich in Vitamin A, which has h They are open between 10.00 a.m. and 4.30 p.m. on normal working days. They are open to audit and scrutiny by both internal audit (link to Accountabili They are organized by townland. They are our brothers our sisters. They are our way of making sure that Irish companies are supported in every way. They are readjusted after the last payment is received, when the Maternity Benefit They are recruited from open competition, which excludes civil servants, and are They are replacing their old cooking methods, which involve sitting a pot on the They are resident in a country with which Ireland has a Double Taxation agreement They are responsible for keeping public places under their control, including public They are responsible for the day-to-day administration of the Council. They are seen alone or sometimes in pairs. They are similar to the orders that can be applied for in the process of divorce They are similar to the orders that can be applied for in the process of divorce They are similar to the orders that can be applied for in the process of judicial They are standard across almost all websites and they do not affect your device They are subject to change but notice will be given. They are subject to revision when the end-year figures become available. They are supported and funded by the Citizens Information Board. They are supported by a section of Irish-America whose Stone Age politics have no They are the accumulated records of an individual or institution which merit pres They are the foundation stones, but also the touchstones, upon which we build our They are the great-great-great (and so on) grandchildren of the Fallow deep They are then released on bail where they are likely to engage in further crimins They are then thoroughly mixed and sorted according to the first preferences reco They are then thoroughly mixed and sorted according to the first preferences reco They are to be prepared in accordance with the provisions of the Electoral Act 19 They are to be the sacrificial lambs to appeare the Dublin wolves". They are trained to: They are usually alone and quite cheeky. They are very labour intensive and take time. They are well suited to follow the brassicas (broccoli or cabbage) that have used They are widely used in order to make websites work, or work more efficiently, as They are written in a question and answer format and answer the questions most of They are: They are: They argue that Article 8 cannot avail the respondent in his argument on the oblic

They are:
They argue that Article 8 cannot avail the respondent in his argument on the oblight They both described the challenges of running multi-activity Active Retirement Associated They bought a house in London and had six children.
They bring independent judgement to bear on issues of strategy, performance, resource.

They bring peace and harmony in their families and a nation at large.

They broke through the Crown forces and moved from Staball Hill down Thomas Street





How to register?







- Eligible individual users
 - Self-registration via <u>https://webgate.ec.europa.eu/etranslation/public/welcome.html</u>

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







- Eligible individual users
 - Public Administrations
 - Universities
 - CEF-funded projects
 - SMEs





READY? LET'S GIVE IT A TRY!

HTTPS://LANGUAGE-TOOLS.EC.EUROPA.EU/





THANK YOU FOR YOUR ATTENTION!

Vilmantas LIUBINAS

(DG Translation, European Commission) vilmantas.liubinas@ec.europa.eu

