

Identifying and managing your data and how ELRC can assist and help



European Language Resource Coordination

ELRC Finland Workshop | 24.10.2018

Aivars Bērziņš | ELRC/Tilde



European
Commission



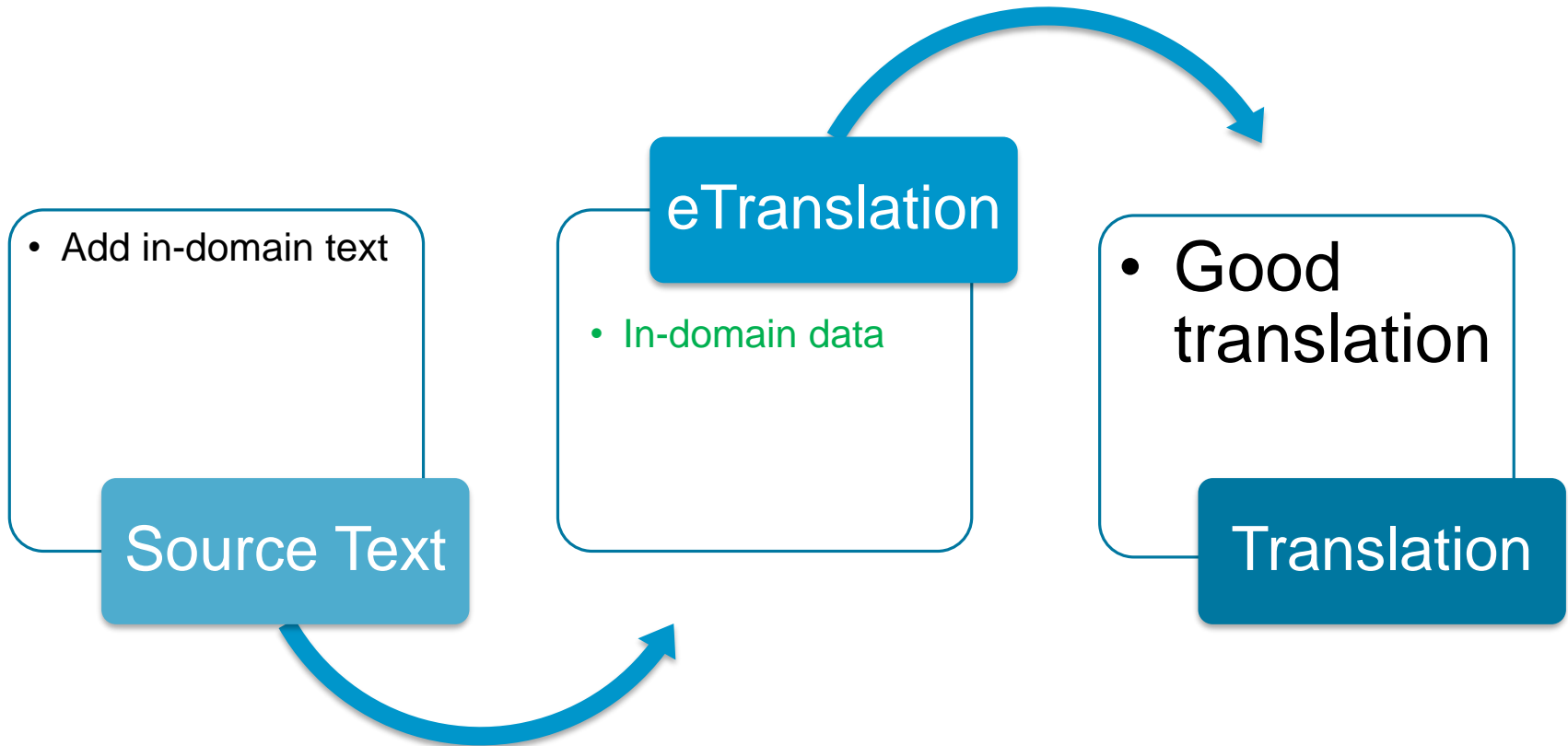
European Language
Resource Coordination
Connecting Europe Facility

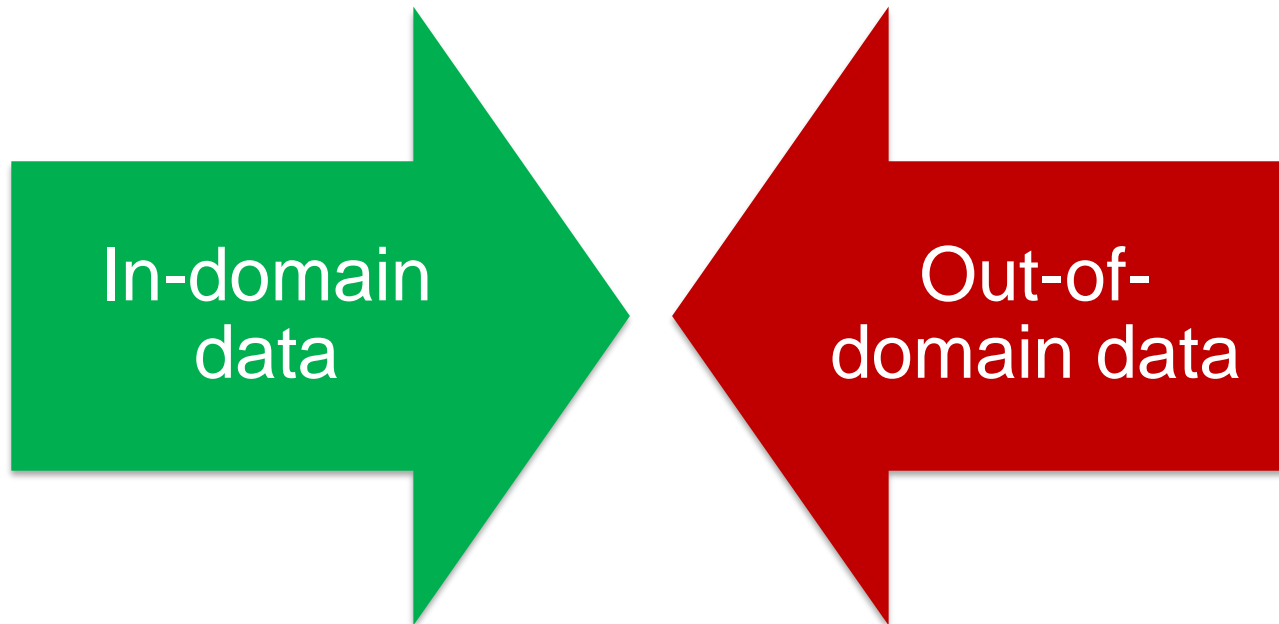


How to provide smaller language communities with the same opportunities that the largest language communities enjoy?

Why ELRC...?







→ Your language resources make a difference!

What has been achieved in the first phase?



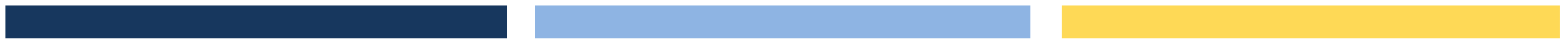
**Collection of 225
language resources
overall**

**More than 2 billion
words in all EU official
languages, Norwegian
and Icelandic**

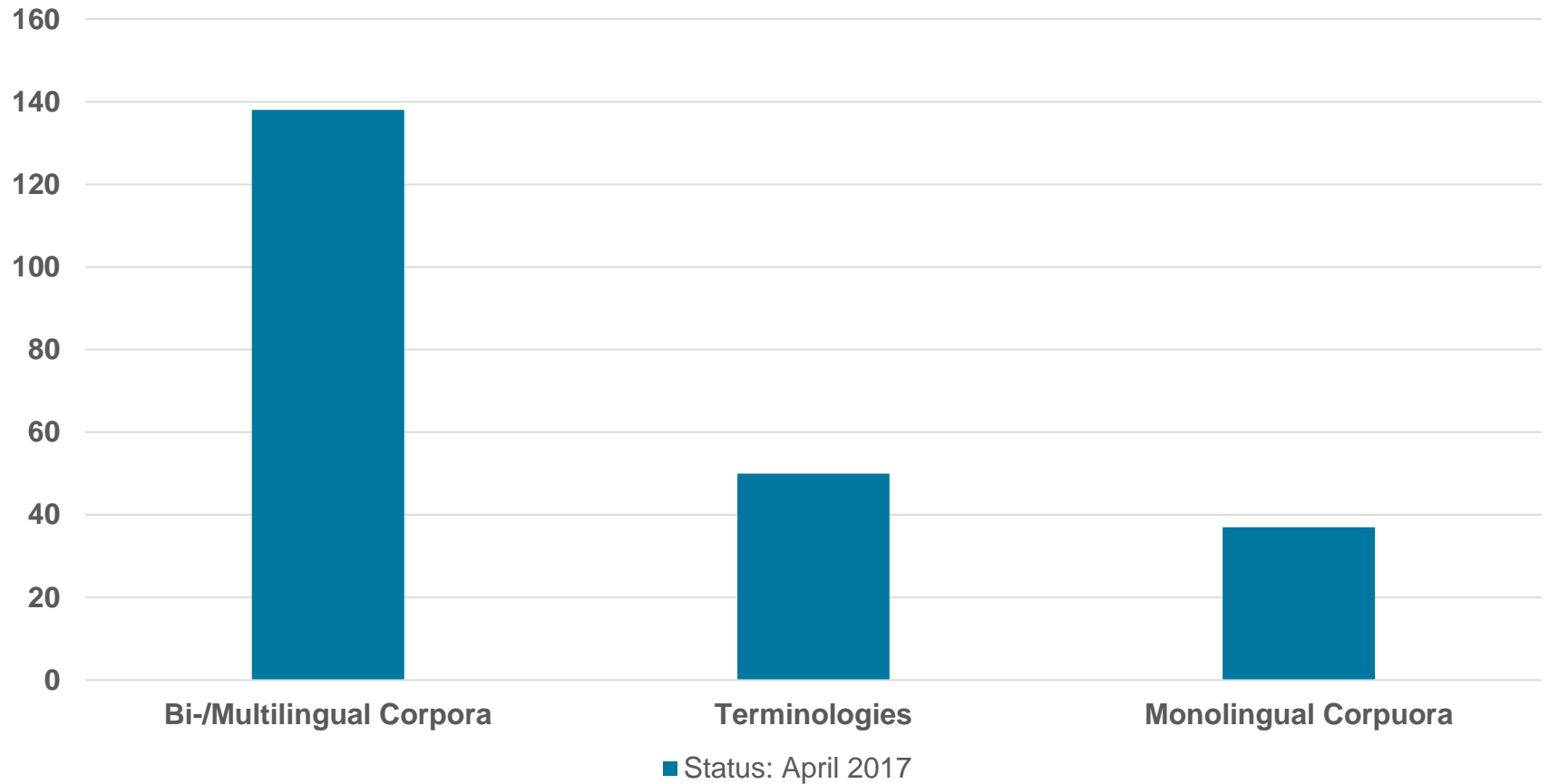
**More than 91 language
resources to be used by
you!**

Over 450.000 terms

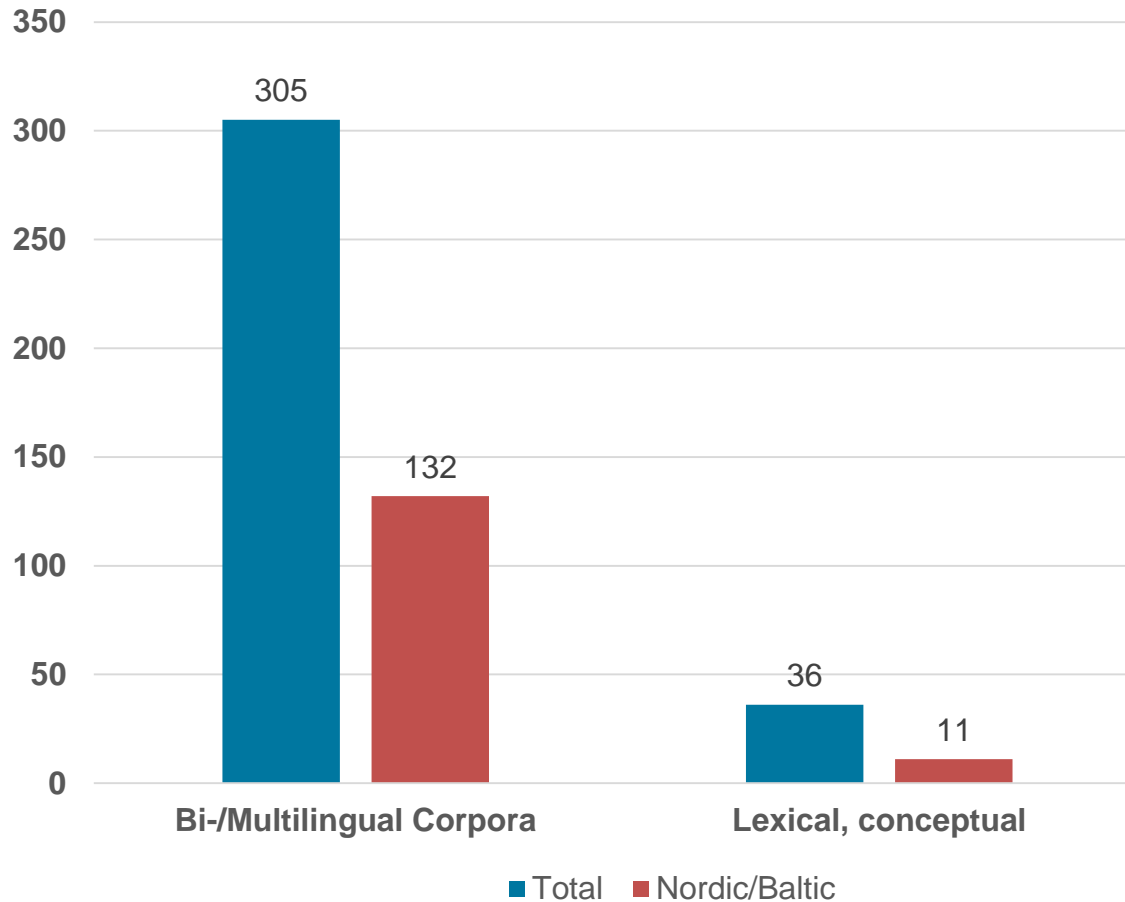
**More than 2 million
translation units**



LR contributions by type



What has been achieved so far...?




Status: October 2018


Count	Corpus	Lexical conceptual resource	Translation units
Danish	12	1	91 335
Estonian	6	1	128 401
Finnish	8		200 307
Icelandic	19	1	490 958
Latvian	5	1	66 550
Lithuanian	8	1	204 176
Norwegian	59	2	1 850 253
Swedish	15	4	103 057

- **The ELRC-SHARE Repository**

- Access to, sharing and contribution of language resources
- Access to tools and services catalogue (upcoming)
- Visit <http://www.lr-coordination.eu/resources>



The screenshot shows the homepage of the ELRC-SHARE Repository. At the top, there is a navigation menu with links for Home, Browse Resources, Help, About, Register, and Login. The main header features the ELRC-SHARE logo and a large image of various national flags. Below the image, the text "ELRC-SHARE Repository" is displayed. A search bar is present with the placeholder text "Type in your keywords, please..." and a "Search" button. Below the search bar, the text "52 language resources at your disposal (and counting)" is shown. The main content area includes a welcome message: "Welcome to the ELRC-SHARE repository!" followed by a brief description of the repository's purpose and a link to the Language Resources section. At the bottom, there is a note about contributing resources, mentioning the "register" and "login" options.



Danish

Search

Filter by:

▼ Language

- ◆ Danish (11)
- ◆ English (11)
- ◆ Finnish (1)
- ◆ French (1)
- ◆ German (1)

more

▸ Resource Type

▸ Media Type

▸ Licence

▸ Conditions of Use

▸ Linguality Type

▸ Multilinguality Type

▸ Data Format

▸ Domain

▸ Appropriateness For DSI

11 Language Resources

Order by: Resource Name A-Z ▼

📁

DA-EN Danish Health Authority

Danish | English

↓ 0 👁 32

Non-standard/ Other Licence/ Terms

📁

DA-EN Danish Ministry of Economic Affairs and the Interior

Danish | English

↓ 0 👁 27

Non-standard/ Other Licence/ Terms

📁

DA-EN Danish Ministry of Finance

Danish | English

↓ 0 👁 27

Non-standard/ Other Licence/ Terms

📁

DA-EN Danish Ministry of Finance 2

Danish | English

↓ 0 👁 18

Helpdesk for Language Resources

Helpdesk for Language Resources

We are happy to answer any questions on the technical or legal aspects related to the use, production, collection, processing, and sharing of language resources.

Please feel free to contact us through one of the following channels:

Telephone*	+33 970 440 522
Secretariat Support	+49 681 857 7552 85
Skype	ELRC Helpdesk
E-mail	help@lr-coordination.eu

lr-coordination.eu/helpdesk

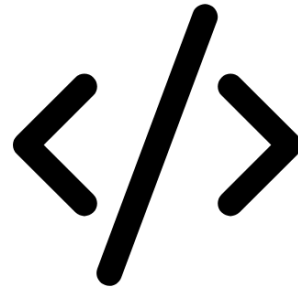


If you encounter any legal or technical issues with your language resources, ELRC and our broad network of experts can help!



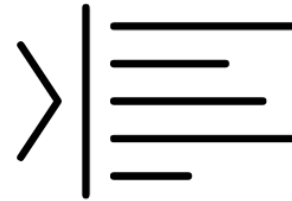
Data conversion

If your data isn't converted to the proper formats, we can help convert it



Tag removal

Does your data contain unneeded tags? We can assist in removing them!



Re-formatting

Need to re-format DOCX to XML, or PDF to WORD? Let us do it for you!



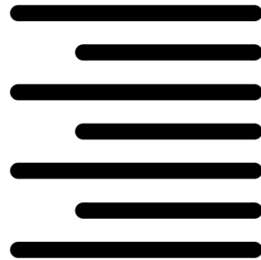
Data extraction

If your data is trapped in archives and databases, we can help extract it



Cleaning

If your data is messy (i.e., lots of noise), we will clean it up



Alignment

Translations aren't aligned? We'll do it for you with our tools!



Metadata validation

Metadata is crucial!
We can validate and organize metadata for your team



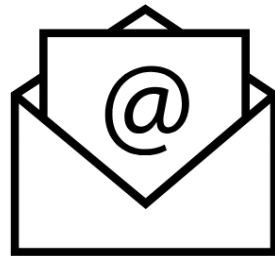
Anonymisation

Does your data contain private info?
We can help to fully anonymise



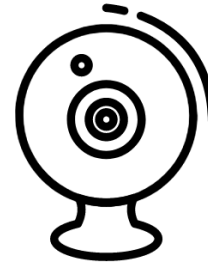
Phone

Pick up the phone and call the ELRC helpdesk



Email

Send us off an email briefly explaining your technical/legal issue



Skype

Skype us! We are online and ready to help with any questions




On-site

Our team of experts will travel directly to assist you at your own offices

ELRC consortium provides language processing services and assistance directly on-site to data contributors





We will fix your data issues and return the processed data directly to you. We can also help to improve your data management processes. Just ask!

How to request services



Submit a request for on-site assistance by filling out the form below. See a list of services [here](#).

First name *

Last name *

Institution *

Country *

Email *

Types of assistance required *

- Legal assistance
- Data processing
- Anonymisation
- Other

Description of assistance required

Submit

[Ir-coordination.eu/request-onsite-assistance](https://ir-coordination.eu/request-onsite-assistance)



Problem: Need to learn how to handle PDFs in order to produce parallel corpora from multilingual PDF files.

Solution: One day focused meeting with data owner team to share best practices of better data handling and processing practices.





Problem: In some Spanish governmental departments, archives were only available in PDF

Solution: ELRC helped provide good converters to get usable documents



Problem: Data owner needed help with anonymization, as databases contained personal info. Another need: cleaning up 'junk' data (URLs, numbers, fragments)

Solution: ELRC helped provide anonymization services and data cleaning



Case studies

1. Amesto Translations Ltd.

- Norway's largest commercial translation company with many public sector clients.
- Normally agreements with public agencies do not say anything about ownership or access to TM.
- The National Library has received TM from all public sector assignments since 2011, ca. 1,5 million units.
- The National Library takes responsibility for all further use.

LATVIA
TRANSLATES
WITH

HUGO.LV

- **MT service** for the Latvian public sector
- Enables **multilingual communication** in the public sector and **empowers access** to information and e-services
- **Securely** translates texts, documents, websites
- **Adapted** for the Latvian language and public sector texts
- **Integrated** into e-services and government websites
- Languages
 - Latvian-English
 - English-Latvian
 - Latvian-Russian

Developed
by

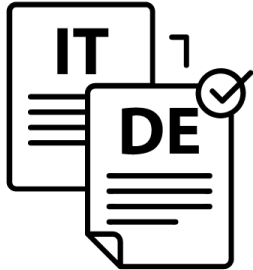


Culture Information
Systems Centre
Republic of Latvia

Integrating eTranslation for the 2017-2018 EU Council Presidency

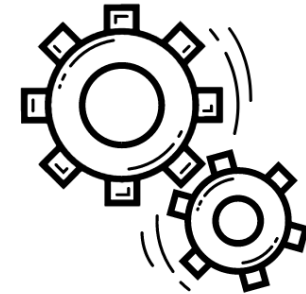


CEF eTranslation

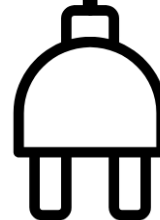


MT systems for the **24 official EU languages** enabling translation of full documents, preserving text formatting

AI-powered Neural MT



AI-powered **custom Neural MT** providing superior-quality translation adapted for the Presidency requirements





Translate 2017

in cooperation with the Estonian Presidency of the Council of the EU

Menu EN EE

Translate text

Translate document

Translate website

From: English German **Estonian** ▼



To: **English** ▼

Translate

Eesti Euroopa Liidu Nõukogu eesistumise moto on „Ühtsus tasakaalu kaudu”. Usume, et Euroopal on piisavalt palju ühist, leidmaks edukalt lahendused kõigile väljakutsetele ning kasutamaks ära kõiki meie ees avanevaid võimalusi. Peame hõlbustama avatust nii meie majanduses kui ka ühiskonnas, kindlustades samas turvalisuse ja julgeoleku. Eesti kui eesistujariigi ülesanne on leida tasakaal liikmesriikide erinevate seisukohtade, traditsioonide ja huvide vahel, et saavutada Euroopa kodanike jaoks parim võimalik tulemus.

The motto of the Estonian presidency of the Council of the European Union is “Unity through balance”. We believe that Europe has enough in common to successfully find solutions to all challenges and to exploit all the capabilities that we face. We need to facilitate openness in both our economies and in society, while securing security and security. The task of Estonia as the presidency is to find a balance between the various views, traditions and interests of member states in order to achieve the best possible outcome for European citizens.

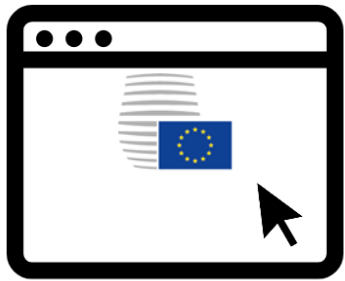
Powered by **eTranslation**, the EU's automated translation service

Developed by **Tilde**

translate2017.eu

EU2017.EE

EU PRESIDENCY TRANSLATOR



Website

For staff, widely enabling automated translation of texts, full documents, websites



Desktop

For journalists, enabling easy translation of documents, press releases, local news sites



Mobile

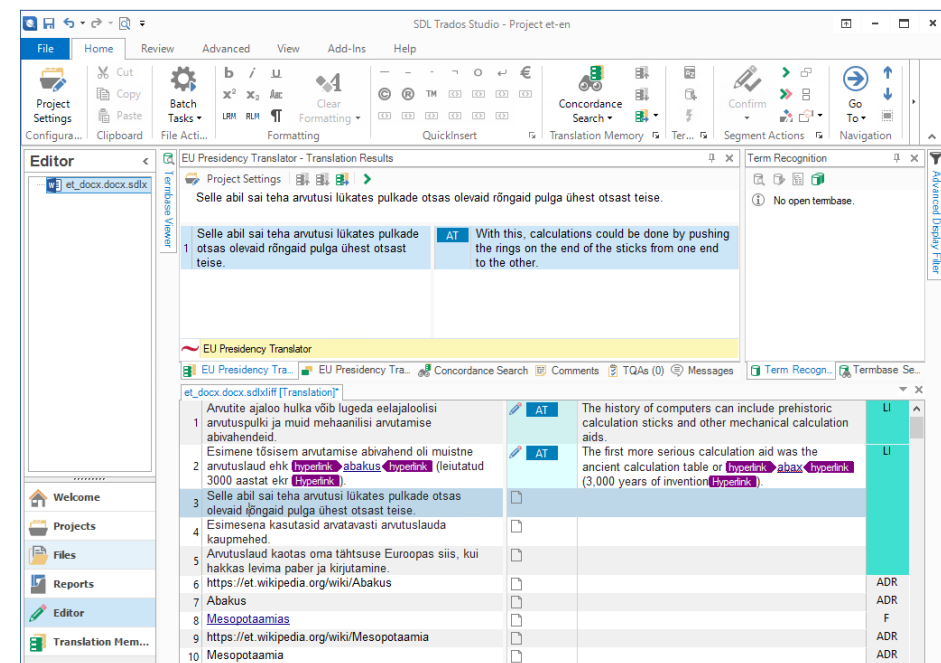
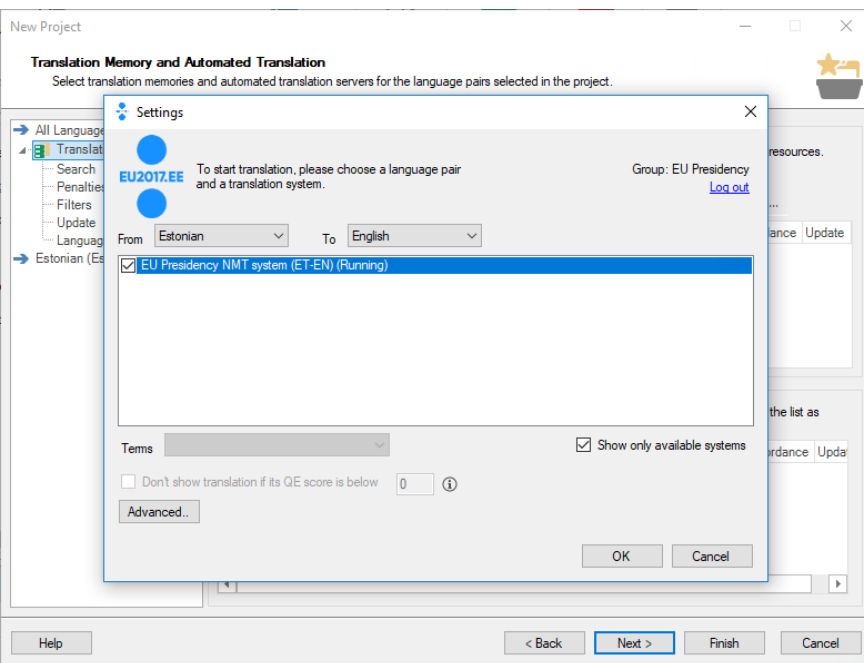
For delegates, enabling quick translation of words and phrases on the fly during meetings



Plugin

For translators, helping boost translation productivity by with MT and post-editing

PLUG-IN FOR PROFESSIONAL TRANSLATORS



translate2017.eu

BENEFITS FOR ESTONIA, BULGARIA, AUSTRIA



- Enables Presidency staff to quickly translate documents
- Empowers visiting journalists and delegates to access info in the local language, e.g., press releases, local news sites
- Supports staff translators in their work by boosting translation productivity up to 35%
- Lowers costs of translation for documents by utilizing post-edited machine translation
- Allows public sector organizations to translate content and websites into multiple languages





Estonia

July-December 2017



Bulgaria

January-June 2018



Austria

July-December 2018





Bulgarian Presidency of the Council of the European Union



- Translate text
- Translate document
- Translate website

From: English German French



To: German French Bulgarian

Translate

Enter the text you want to translate.

Machine translation results help to understand the meaning of a source text, but do not equal translation by a human.

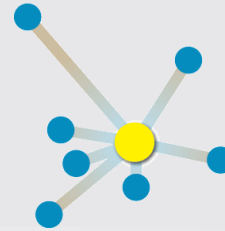
European Parliament Resolution
Language equality in the digital age
Sept 11, 2018



Kiitos!

www.lr-coordination.eu

info@lr-coordination.eu



**European Language
Resource Coordination**

Connecting Europe Facility

