



Deliverable Task 6

ELRC Workshop Report for Estonia



Author(s):	Martin Luts (Tilde) Margit Kurm (Tilde)
Dissemination Level:	Public
Version No.:	V1.0
Date:	07-03-2016



Contents

<u>1</u>	<u>ELRC Workshop in Estonia</u>	<u>3</u>
<u>2</u>	<u>Workshop Agenda</u>	<u>4</u>
<u>3</u>	<u>Workshop Participants</u>	<u>6</u>
<u>4</u>	<u>Summary of the Content of Sessions</u>	<u>7</u>
4.1	Session 1: Opening addresses	7
4.2	Session 2: “Multilingualism in Europe”	7
4.3	Session 3: “The Estonian Language in the Digital Age”	7
4.4	Discussion 1: “Multilingual Public e-Services in Estonia”	8
4.5	Session 4. “Future of Translation”	8
4.6	Session 5: “Automated Translation: How does it work and what is it used for? Use cases”	8
4.7	Discussion 2: “Bilingual Data in the Information Systems of Institutions”	9
4.8	Session 6: “Legal Framework”	9
4.9	Session 7: “How Can Public Institutions Benefit from Automated Translation Infrastructure?”	9
4.10	Session 8: “How to Prepare Data for Automated Translation”	10
4.11	Session 9: “Closing remarks”	10
<u>5</u>	<u>Workshop Presentation Materials</u>	<u>11</u>
<u>6</u>	<u>Appendix</u>	<u>12</u>
6.1	Invitation Letter (in Estonian)	12
6.2	Agenda (in Estonian)	14
6.3	ELRC Workshop Brochure (in Estonian)	16

ELRC Workshop Report for Estonia

1 ELRC Workshop in Estonia

The Estonia ELRC Workshop took place in Tallinn, on the 11th of February 2016, at the office of the European Commission Representation in Estonia. The workshop was co-organized by the ELRC Representative in Estonia – Tilde, Estonian Language Resource Centre and Office of the European Commission Representation in Estonia.

An event-dedicated web part in Estonian was set up at the ELRC website before the event, at <http://lr-coordination.eu/et/avaleht>. The page was populated with the agenda (http://lr-coordination.eu/estonia_agenda) and the online registration form (http://lr-coordination.eu/et/estonia_registration).

Invitations to the participants were sent out a couple of weeks before the event. The invitation text was a localized version of the respective template, signed by the EC (Saila Rinne, signed by hand) and countersigned by all co-organizers (Hannes Rumm, Kadri Vider, Margit Kurm; all digitally signed), cf. Appendix 6.1. The personalized invitations were sent through the account tallinn@lr-coordination.eu which was created for this event communication purposes. The list of persons invited to the event is provided in Appendix **Error! Reference source not found.** The invitation was followed by a personal e-mail and calls to the persons to whom invitations were sent. A reminder and updated information was sent out to all registered participants a couple of days before event.

The event was voice recorded and videos (slides synchronized with the voice recordings, both in original, Estonian and English) are available on Workshop dedicated Web.

2 Workshop Agenda

The workshop agenda was based on a “master” agenda agreed with the EC, the agenda was adopted for the local needs.

**Agenda for CEF.AT - ELRC Training Workshop
(European Language Resource Coordination)
Office of the European Commission Representation in Estonia
Rävala pst 4, Tallinn**

9:00 – 9:30	Registration and coffee
	Opening and welcome Saila Rinne , EC DG CONNECT
9:30 – 9:55	Jana Laurend , European Commission Representation in Estonia Kadri Vider , Center of Estonian Language Resources Andrejs Vasiljevs , ELRC/Tilde
9:55 – 10:30	Multilingualism in Europe Saila Rinne , EC DG CONNECT
10:30 – 11:00	The Estonian Language in the Digital Age Kadri Vider , Center of Estonian Language Resources
	Panel: Multilingual Public Services in Estonia
11:00 – 11:35	Katrin Ringo , State Information System Authority, state portal eesti.ee Remo Tiigirand , Enterprise Estonia, e-Residency project Aleksandr Belousov , Centre of Registers and Information Systems
11:35 – 12:05	Coffee break
12:05 – 12:30	Future of Translation Rihards Kalniņš , ELRC/Tilde
	Automated translation: How does it work? Use cases
12:30 – 12:50	Mark Fishel , Tartu University Martin Luts , ELRC/Tilde
	Panel: Bilingual Data in the Information Systems of Institutions
12:50 – 13:25	Kaja Kell , Bank of Estonia Helve Trumann , Chancellery of the Riigikogu Kadri Vider , Center of Estonian Language Resources Margit Kurm , ELRC/Tilde

ELRC Workshop Report for Estonia

13:25 – 13:45	Legal Framework Viljar Peep , Estonian Data Protection Inspectorate
13:45 – 14:45	Lunch
14:45 – 15:15	How Can Public Institutions Benefit from CEF .AT? Saila Rinne , EC DG CONNECT
15:15 – 15:30	How to Prepare the Data of Institutions for Automated Translation? Andrejs Vasiljevs , ELRC/Tilde
15:30 – 16:15	Teamwork and coffee
16:15 – 16:30	Wrap-up

ELRC Workshop Report for Estonia

3 Workshop Participants

The Estonian ELRC workshop received 74 registrations spanning a wide range of ministries and public organizations. The workshop was well-attended – there were 66 participants. Before the workshop we organized cold calling activities, as a result we got more contacts and suggestions who would be better candidates considering the theme of the workshop. Furthermore, in this way we were able to make sure that important public administration institutions were represented.

The full participants list can be viewed in the Appendix **Error! Reference source not found.**



4 Summary of the Content of Sessions

4.1 Session 1: Opening addresses

Jana Laurend, the language specialist of the EC Representation in Estonia, opened the event by welcoming the audience and introducing the key persons in conceiving and organizing the event, namely the ELRC consortium and the EC/DGT representatives. She also remembered from her work experience as an EU translator how machine translation helped translators of other languages, but unfortunately not yet for the Estonian language.

Saila Rinne, programme officer for EU Policies, DG Connect, introduced to the audience the European Commission's project for building CEF.AT and the role of language resources. She expressed her thanks to the local organizers for helping to organize the workshop.

Andrejs Vasiljevs, the local ELRC representative, welcomed the audience on behalf of the consortium. He described the crucial role that language technology development has for smaller languages in order for them to survive in the digital era and the importance of cooperation in collecting language resources for building CEF.AT. He also gave an overview of the ELRC consortium's activities and expressed his thanks for seeing so many participants.

Kadri Vider, the head of the Center of Estonian Language Resources, also welcomed the audience on behalf of her organization, being one of the co-organizers of the event.

4.2 Session 2: “Multilingualism in Europe”

Saila Rinne, programme officer for EU Policies, DG Connect, began the first session following the opening addresses: “Multilingualism in Europe.” Ms. Rinne described the importance of linguistic diversity at the EU, home to 24 official EU languages and a total of about 60 major “regional/minority” languages. She stressed that the equality of the EU official languages is enshrined in the European legal basis, therefore the EU is committed to supporting multilingualism.

This was also the basis for the EU's strong initiative to support language technologies, such as machine translation. This support is exemplified in the creation of the MT@EC system for public administrations, as well as in the deployment of mature language technologies through the Connecting Europe Facility (CEF) programme.

Rinne went on to describe the EU's Digital Single Market strategy and the ways in which language barriers affect public and private services. According to Rinne, pan-European public services currently face a multilingual challenge. There is no lingua franca among public administrations in Europe and 90% of EU web users prefer to use their own language for online services. The traditional method used to counter this – human translation – is too expensive and too slow with the intended text volumes. Therefore, human translation is not a solution in all use scenarios.

The EU's solution to the challenge is to build a pan-European automated translation platform, CEF.AT, which will make European public online services multilingual by deploying mature language technologies (such as MT@EC) in a secure platform. The goals are to make public digital services equally usable by all EU users, irrespective of their working language and language skills, as well as to facilitate cross-border information exchange in public administration.

Rinne concluded by detailing the role of member states in the creation of CEF.AT, namely, the sharing of language resources as part of the ELRC project. Rinne said that by contributing resources, member states could help improve the quality of the CEF.AT systems and then enjoy the benefits of automated translation: the reduction of language barriers in Europe.

4.3 Session 3: “The Estonian Language in the Digital Age”

This session featured one speaker: Kadri Vider, head of the Center of Estonian Language Resources.

ELRC Workshop Report for Estonia

This session discussion focused on the Estonian language in the digital age – examining the status of “digital extinction” for Estonian and the measures that are available to support the language today.

In her presentation, Kadri Vider gave an overview about language technology and what has been done for the Estonian language so far. She mentioned that Estonia is very fortunate as we have the Estonian law and language development program (which also contains a language technology support development measure). And in order to fulfill the objectives, we have the National Language Program and the results of its projects are free for everybody. She described different ongoing projects. Also, she introduced the work of the Center of Estonian Language Resources – CLARIN and METANORD networks partner. She described how the Center is collecting language resources and what resources they already have and called for cooperation in sharing and donating language resources.

4.4 Discussion 1: “Multilingual Public e-Services in Estonia”

The panelists were: Katrin Ringo, State Information System Authority, state portal eesti.ee; Remo Tiigirand, Enterprise Estonia, e-Residency project; Aleksander Belousov, Centre of Registers and Information Systems.

Mr. Belousov kicked off the session by introducing the wide range of registers managed by the Centre of Registers and Information Systems. These registers include: e-Business Register, e-Notary, e-Land Register, Court Information System, Criminal Records Database, e-File, electronic State Gazette. He pointed out which e-services are available for the end-users not only in Estonian but also in English. He outlined the obstacles for (automated) multilingual services.

Mr. Tiigirand continued the session by speaking about his organization’s effort regarding the e-Residency program: project timeline, number of applicants etc. He elaborated on the variety of motivators for applying for an e-Residency in Estonia, the two major ones being setting up a business in Estonia and investing here. He discussed the language barriers for these activities encountered by the e-Residents.

Finally, Ms. Ringo presented her work with the state portal eesti.ee, focusing on multilingual public e-services: a mixture of information sharing, requests, e-forms, e-signature etc. She described shortly the end-users of the portal: citizens, businesses, officials. The short descriptions of the services are human-translated into English and Russian, only one service is fully multilingual (e-prescriptions, trilingual). E-forms are only available in Estonian. Translation practice and the future scenarios were discussed shortly.

4.5 Session 4. “Future of Translation”

This session featured one speaker: Rihards Kalniņš, ELRC/Tilde.

Mr Kalniņš started by talking about the role of machine translation in the European and world context, giving examples about e-services and e-commerce from all over the world and also about translating user feedback. He then moved on from the business sphere to the sphere of governance. He described the approach of automated translation that is used in Latvia and was used during Latvia’s presidency of the Council of the EU, giving examples of use situations: information to journalists, guests and tourists. Mr Kalniņš then delved into depths of the corresponding machine translation system, describing the data needed for creating such a system. He outlined the channels through which the service can be used. He raised the following question: how prepared is Estonia for fulfilling its role of presidency and offering bilingual information and services?

4.6 Session 5: “Automated Translation: How does it work and what is it used for? Use cases”

This session featured two speakers: Martin Luts, ELRC/Tilde and Mark Fishel, Tartu University.

Mr Luts introduced a study conducted among Estonian public officials about their translation practices.

ELRC Workshop Report for Estonia

Dr. Fishel gave a brief overview of the technical basics of automated translation. He introduced modern statistical machine translation, giving examples of statistical systems and how they are used in the decoding process. He showed some examples and counter-examples of good machine translation outputs, discussing the root causes. He outlined the use cases where using MT is beneficial (postediting among others). He talked about how sentence and word alignment functions, giving some interesting examples to help explain the complex points. Mr Fishel also introduced the MOSES core.

In conclusion, Dr. Fishel stressed the importance of language resources – monolingual data, bilingual data, glossaries, etc. – in building high-quality MT engines. He explained how much data was necessary for building systems, as well as the clear benefits of training domain-based MT systems. Also, segregation of duties (MT trainer and client) were proposed.

4.7 Discussion 2: “Bilingual Data in the Information Systems of Institutions”

This session featured four speakers: Kaja Kell, Bank of Estonia; Helve Trumann, Chancellery of the Riigikogu (Parliament); Kadri Vider, Center of Estonian Language Resources; Margit Kurm ELRC/Tilde.

Margit Kurm from ELRC/Tilde reminded that a good source of bilingual data is translation procurements, each year there are 15-23 different tenders from different state institutions and it is good to ask back bilingual files. She also talked about the usage of translation memories.

Helve Truman from the Chancellery of the Parliament talked how translation is organized in their office, what tools they use and what they are translating.

Kaja Kell from Bank of Estonia shared their experience and mentioned how important it is to review old translations and terminology among other state classifiers as it affects terminology and understanding on such a large scale.

Kadri Vider called the participants to think what resources they may have to improve MT and donate them to the Language Resources Center. All material is good for teaching a MT system. She also gave several samples.

4.8 Session 6: “Legal Framework”

In his presentation at the session on “Legal Framework,” Viljar Peep (the director of the Estonian Data Protection Inspectorate) concentrated on the legal aspects of reusing data in automated translation. First he gave an overview about the legal framework, e.g. Personal Data Protection Act, Public Information Act. He didn’t concentrate on copyright law or license law. He demonstrated how personal data can be made impersonal and showed other legal means for promoting the reuse of data.

4.9 Session 7: “How Can Public Institutions Benefit from Automated Translation Infrastructure?”

In this session, a representative of the European Commission, Saila Rinne, of DG Connect detailed how public institutions can benefit from automated translation infrastructure.

Ms. Rinne discussed the importance of enabling multilingualism; the role of machine translation in doing so; the European Commission’s current MT service MT@EC; and future plans to build the CEF.AT platform. ELRC is working toward building CEF.AT. Rinne talked about the current interaction between actors in the member states and the EU.

Rinne then discussed MT, which as she explained, is the only viable solution for the following: quick and cheap access to information in foreign languages; understanding information received in a foreign language that otherwise could not be used or would require substantial time and costs to translate; making multilingual use of websites possible; and facilitating cross-lingual information search and analytics.

ELRC Workshop Report for Estonia

According to Rinne, these benefits are apparent in the EC's current MT service, MT@EC, launched in 2013. MT@EC is a SMT system based on Moses. The system can be used to translate multiple documents in multiple languages and is currently used by EU institutions and bodies.

In the future, Rinne said, MT@EC will evolve into CEF.AT. CEF.AT will build on the existing MT@EC service but not be limited to it; put emphasis on secure, high-quality, customizable MT for pan-European online services but not be limited to them; and be a multilingualism enabler – not only MT.

4.10 Session 8: “How to Prepare Data for Automated Translation”

In this final session, Andrejs Vasiljevs of ELRC/Tilde began by illustrating how the volume of data impacts machine translation quality. Then, he listed various data sources in organizations – from dictionaries to bulky publications, from archives and intranet to translation companies and open data. He went on to describe the process of managing the life cycle of data, enriching it with metadata, aligning data and also talked about the terminology service TaaS and machine translation platforms. Finally, he encouraged users to share data and recommended the first steps on that road, including the ELRC helpdesk service.

4.11 Session 9: “Closing remarks”

The closing session consisted of brief closing remarks from the main organizers of the event: Andrejs Vasiljevs of ELRC/Tilde and Saila Rinne of DG Connect.

The speakers once again stressed the need for collaboration not only between EU member states but also by individual institutions within each country. Only by collaborating and sharing resources can the vision of CEF.AT finally be realized, reducing language barriers in Europe.

ELRC Workshop Report for Estonia

5 Workshop Presentation Materials

Presentations of the workshop can be divided into two categories – based on master slides (adopted for local auditory) and presentations prepared by panel discussion participants. All presentations are provided on the ELRC Tallinn Workshop website (http://lr-coordination.eu/estonia_agenda). This website contains all presentations in .pdf format that are supplemented with video recordings of presentations. Two types of video are provided – in original language and voice over translations into English for presentations originally in Estonian and voice over translations into Estonian for presentations in English. In addition .pdf format presentations are stored in the respective Dropbox folder: Dropbox\ELRC European Commission\T6 Workshops\Trial Workshop Estonia.

To provide more information about the workshop's topic, a dedicated ELRC Tallinn Workshop brochure was prepared. The brochure was adjusted to the style of the ELRC web page. The brochure provides information about multilingualism, CEF.AT and CEF activities, ELRC (activities and consortium), important resources. In addition, there is information about activities in Estonia regarding the automated translation solution for public administration. The brochure is provided in Appendix 6.3.

	Avamine ja tervitussõnad (Video - ORIGINAL, ENG, EE)
9:30 – 9:55	Saila Rinne, EC DG CONNECT Jana Laurend, Euroopa Komisjoni esindus Eestis Kadri Vider, Eesti Keeleressursside Keskus Andrejs Vasiljevs, ELRC/Tilde
9:55 – 10:30	Mitmekeelsus Euroopas Saila Rinne, EC DG CONNECT (Video - ORIGINAL, ENG, EE)
10:30 – 11:00	Eesti keel digiajal Kadri Vider, Eesti Keeleressursside Keskus (Video - ORIGINAL, ENG, EE)
11:00 – 11:35	Arutelu: mitmekeelsed avalikud teenused Eestis (Video - ORIGINAL, ENG, EE) Katri Ringu, RIA riigiportaal.eesti.ee Remo Tiigirand, EAS, e-residentsuse projekt Aleksandr Beloussov, RIK
11:35 – 12:05	Kohvipaus
12:05 – 12:30	Tölkimise tulevik Rihards Kalniņš, ELRC/Tilde (Video - ORIGINAL, ENG, EE)
12:30 – 12:50	Automaattõlge: kuidas see toimub? Kasutusstsenariumid (Video - ORIGINAL, ENG, EE) Mark Fishel, Tartu Ülikool Martin Luts, ELRC/Tilde
	Arutelu: kakskeelsed andmed seotud infrastruktuurides (Video - ORIGINAL, ENG, EE)

Picture 1 Snapshot of Workshop Web page

6 Appendix

6.1 Invitation Letter (in Estonian)

Following invitation letter was sent out to list of 112 invites. The invitations were sent through the account Tallinn@lr-coordination.eu which was created for this event communication purposes.



Hea kolleeg

Kutsume teid Eesti avaliku sektori automaattõlkeseminarile, mis toimub 11. veebruaril 2016 Euroopa Komisjoni Eesti esinduses aadressil Rävala 4, Tallinn.

Keeltele ja keelelisel mitmekesisusel on Euroopa kultuuris, ettevõtluses ja koostöös keskne roll. Selleks et hõlbustada nii Euroopa Komisjoni ja Euroopa riigiasutuste vahelist kui ka Euroopa riigiasutuste omavahelist mitmekeelset suhtlust ning dokumendivahetust, on Euroopa Komisjon loonud programmi „Connecting Europe Facility“ (CEF, Euroopa ühendamise rahastu) raames automaattõlkeplatvormi CEF AT. Avaliku sektori asutused saavad seda kasutada tasuta. Platvorm põhineb MT@EC masintõlkesüsteemil, mille on välja töötanud Euroopa Komisjoni kirjaliku tõlke peadirektoraat ja mis on nüüd antud liikmesriikide riigiasutuste käsutusse. CEF-i programmi raames rakendamine avardab tõlkesüsteemi kasutusala, parendab selle kvaliteeti ning valmistab süsteemi ette selleks, et seda saaks edaspidi kasutada üleeuroopalistes avaliku sektori veebiteenustes.

Soovime seminaril osalejad kaasata aktiivselt aruteludesse ja loodame, et nad aitavad meil paremini mõista mitmekeelses töökkeskkonnas töötavate inimeste vajadusi. Meie eesmärk on viia eestikeelne automaattõlge nende vajadustega vastavusse. Täpse ajakava leiab kirja manusest. Seminari viivad läbi Euroopa Komisjoni selle ala eksperdid, keeletehnoloogid ja keeleteenuste pakkujad. Soovime saada selgust selles, mida ootate automaattõlkel, ning ühtlasi välja selgitada, millised on teie asutuse mitmekeelsed keeleressursid. Koos Andmekaitse Inspektsiooni esindajaga arutame ka automaattõlgetega seotud tehnilisi ja õiguslikke küsimusi. Seetõttu on teie osalemine väga oluline – tänu teie tagasisidele saame automaattõlke kujundada just teie vajadustele vastavaks.

Kui olete huvitatud automaattõlke kasutamisest oma asutuses või soovite olla kursis selle praeguste võimalustega ja viimase aja edusammudega, olete oodatud osalema. Registreeruge aadressil http://www.lr-coordination.eu/events/eesti_registration/. Küsimuste korral võtke palun ühendust seminari kohaliku koordinaatoriga: tallinn@lr-coordination.eu. Täpsemat infot seminaride sarja ja antud seminari kohta vaadake veebilehelt <http://www.lr-coordination.eu/events/>. Jääme ootama teie vastust.

Lugupidamisega

Saila Rinne, European Commission, DG CONNECT

Hannes Rumm, Euroopa Komisjoni esindus Eestis

Kadri Vider, Eesti Keeleressursside Keskus

Margit Kurm, Tilde

6.2 Agenda (in Estonian)

Workshop agenda was based on “master” agenda agreed with EC, agenda was adopted for the local needs.

**Agenda for CEF.AT - ELRC Training Workshop
(European Language Resource Coordination)
Office of the European Commission Representation in Estonia
Rävala pst 4, Tallinn**

9:00 – 9:30	Registreerimine ja kohv Avamine ja tervitussõnad Saila Rinne, EC DG CONNECT
9:30 – 9:55	Jana Laurend, Euroopa Komisjoni esindus Eestis Kadri Vider, Eesti Keeleressursside Keskus Andrejs Vasiljevs, ELRC/Tilde
9:55 – 10:30	Mitmekeelsus Euroopas Saila Rinne, EC DG CONNECT
10:30 – 11:00	Eesti keel digiajastul Kadri Vider, Eesti Keeleressursside Keskus
11:00 – 11:35	Arutelu: mitmekeelsed avalikud teenused Eestis Katrin Ringo, RIA riigiportaal eesti.ee Remo Tiigirand, EAS, e-residentsuse projekt Aleksandr Belousov, RIK
11:35 – 12:05	Kohvipaus
12:05 – 12:30	Tõlkimise tulevik Rihards Kalniņš, ELRC/Tilde
12:30 – 12:50	Automaattõlge: kuidas see toimub? Kasutusstenaariumid Mark Fishel, Tartu Ülikool Martin Luts, ELRC/Tilde
12:50 – 13:25	Arutelu: kakskeelsed andmed asutuste infosüsteemides Kaja Kell, Eesti Pank Helve Trumann, Riigikogu Kantselei Kadri Vider, Eesti Keeleressursside Keskus Margit Kurm, ELRC/Tilde

ELRC Workshop Report for Estonia

- 13:25 – 13:45 **Õigusraamistik**
Viljar Peep, Andmekaitse Inspektsioon
- 13:45 – 14:45 **Lõuna**
- 14:45 – 15:15 **Kuidas saab automaattõlkeplatvorm CEF AT riigiasutustele kasulik olla?**
Saila Rinne, EC DG CONNECT
- 15:15 – 15:30 **Kuidas asutuse andmeid automaattõlkeks ette valmistada?**
Andrejs Vasiljevs, ELRC/Tilde
- 15:30 – 16:15 **Rühmatöö ja kohv**
- 16:15 – 16:30 **Kokkuvõte**

6.3 ELRC Workshop Brochure (in Estonian)



Keelte mitmekesisus – Euroopa nurgakivi

Euroopa keeleressurside koordineerimise seminar

Tallinn, Eesti
11. veebruar 2016



Figure 1 Front cover of Brochure



Pilt Euroopa Komisjon



Teel ühtse mitmekeelse Euroopa poole

Viimastel aastatel on Euroopa kõrvaldanud palju piiranguid, mis takistavad digitaalse ühtse turgu kujunemist. See on Euroopa üks olulisemaid prioriteete. Raha, kaubad ja inimesed saavad nüüd vabalt liikuda üle riigipiiride, rikastades kogu Euroopa Liitu.

Endiselt on aga veel palju piiranguid, mis takistavad digitaalse ühtse turu loomist ega võimalda kõigil inimestel kasu saada paremini lõimitud Euroopast. Üks viimaseid takistusi on keelebarjäär.

Kuigi keelte mitmekesisus on EL-i nurgakivi, tekitab mitmekeelsus rahvaste vahel ka barjääre. Keelebarjäärid ei võimalda Euroopa tarbijatel kasutada täisväärtuslikult ära interneti pakutavaid võimalusi ja piiravad juurdepääsu paljudele digiteenustele.

Nende barjääride tõttu on EL-i internetiturg lükkunud. 43% eurooplastest ei osta kunagi internetist kaupu ega teenuseid, kui neid pakutakse võõrkeeles. Keeleoskus mõjutab juurdepääsu riigiasutuste e-teenustele ning keelekogukondade piirid takistavad EL-i rikkalikku kultuurivara levikut.

Digitaalses keskkonnas aitab keelebarjääre ületada keeletehnoloogia.

Tehnoloogilised saavutused (näiteks automaattõlge) loovad ulatuslikud võimalused mitmekeelseks suhtlemiseks digikeskkonnas. Automaattõlge tugevdab EL-i elanike, riigiasutuste ja vabariikide vahelist sidet, elavdades suhtlust ning soodustades üksteisemõistmist.

Nii tekib mitte ainult ühtne mitmekeelne digitaalne turg, vaid ka omavahel tihedamalt seotud ja paremini lõimitud mitmekeelne Euroopa.



Picture 2 Page 1 and page about multilingualism in Europe



Euroopa automaattõlkeplatvorm CEF.AT

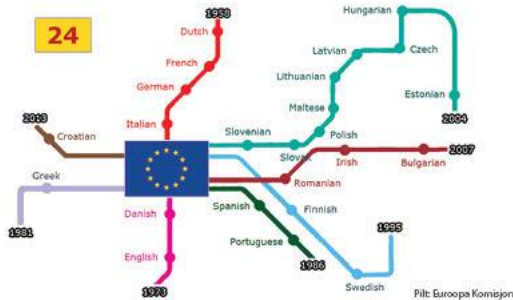
Euroopa ühendamise rahastust (Connecting Europe Facilities – CEF) finantseeritakse projekte, mille abil luuakse (seni puudunud) side Euroopa transpordi-, energia- ja digitaalvõrkude põhistruktuuride vahel. Samuti muudab see Euroopa majanduse keskkonnasõbralikumaks, soodustades säästlike transpordiliikide, liirete lairühenduste ja taastuvate energiaressursside kasutamist vastavalt strateegiale „Euroopa 2020“. Keskkonduses omavahel seotud nutikatele ja kestlikele transpordi-, energia- ja digitaalvõrkudele, soodustab Euroopa ühendamise rahastu jätkusuutliku Euroopa digitaalse ühtse turu loomist.

Euroopa ühendamise rahastu pöörab suurt tähelepanu IKT-sektorile. IKT-sektoris plaanitakse luua lairühenduste ja digiteenuste taristu (Digital Service Infrastructure – DSI). See tagab ulatusliku piiritlete teenuste valiku nii elanikele, ettevõtjatele kui ka riigiasutustele. Üks olulisemaid valdkondi selles töös on kõiki EL-i liikmesriike hõlmav automaattõlkesüsteemide virtuaaltaristu (CEF.AT).

CEF.AT eesmärk on pakkuda ühtsel platvormil kvaliteetseid valdkonnaomaseid automaattõlkesüsteeme kõigi EL-i ametlike keelte jaoks. CEF.AT platvormi saab kasutada nii tarbijate õiguste, tervisekaitse, avalike hangete, sotsiaalhoolekande, kultuuri- kui ka muudes valdkondades.

CEF.AT püüab rakendada parimaid EL-is ja selle liikmesriikides välja töötatud lahendusi ning tehnoloogiaid. Eesmärk on luua kordumatu ja kõikehõlmav lahendus, mis võimaldaks ületada keelebarjääri ja kujundada ühtsema Euroopa.

Mitmekeelne keskkond soodustab Euroopa riikide internetiteenuste kasutamist. Automaattõlkeplatvorm võimaldab kõigil Euroopa riigiasutustel astuda pika sammu keelebarjääri töökeskkonna loomise teel.



ELRC – Euroopa keeleressursside koordineerimise programm

Euroopa keeleressursside koordineerimise programmi (ELRC) raames alustab Euroopa Komisjon enneolematult ulatuslikku keeleandmete kogumise tööd. See võimaldab kõigil EL-i liikmesriikides, Islandil ja Norras kohandada CEF.AT platvormi sobivaks riiklike teenuste osutamise jaoks, soodustades sel kombel mitmekeelsete teenuste pakkumist Euroopa elanikele, riigiasutustele ja ettevõtetele ning põgenikele. Nii kõrvaldab ELRC lõhe, mis vaitseb praegu Euroopa Komisjoni poolt riigiasutuste käsutusse antud masintõlkesüsteemide võimaluste ja kõigis Euroopa riikides pakutavate teenustega seotud tegelike igapäevaste vajaduste vahel. Lisaks annab see võimaluse Euroopa rahvaste keeli laiemalt kasutada.

ELRC eesmärk on luua pidev keeleressursside koordineerimise mehhanism, mis tagaks Euroopa automaattõlkeplatvormi jaoks vajalike keeleressursside olemasolu. Ainult sel juhul, kui kõik EL-i liikmesriigid kaasatakse keeleressursside kogumisse ja masintõlketehnoloogia arendamisse, on võimalik luua võrdväärsed tõlkesüsteemid kõigi EL-i keelte jaoks. ELRC õhutab riike selles valdkonnas aktiivselt osalema ja panustama ühtse Euroopa taristu loomisse.

ELRC tegeleb EL-i kõigi 24 ametliku keele ning norma ja islandi keelega. Muutes mahukad ning mitmekeelsed keeleressurssid kättesaadavaks, aitab ELRC parandada automaattõlke kvaliteeti ja eri valdkondades kasutatavust. Eriti peab CEF silmas digiteenuste mitmekeelseks muutmise vajadust.

Selle eesmärgi saavutamiseks pöörab ELRC tähelepanu riigiasutustes loodavatele ja hoitavatele keeleressurssidele. Neis asutustes luuakse väga mitmesuguseid dokumente, millest paljud tõlgitakse teistesse keeltesse. Need dokumendid on väärtuslik materjal automaattõlkesüsteemide täiustamiseks. Ka akadeemilistes asutustes luuakse uurimis- ja õppetöö käigus mitmekeelseid keelematerjale.

ELRC hoolitseb selle eest, et neid keeleandmeid saaks avatud andmete põhimõtet järgides kasutada nii automaattõlke parandamiseks kui ka võimaluse korral muuks otstarbeks.

ELRC aitab avaliku sektori esindajatel ja teistel keeleressursside omanikel neid andmeid paremini hallata ning lähendada nende kättesaadavuse ja kasutamiseiga seotud tehnilisi ja juriidilisi küsimusi. Kasulikud nõuan ded ja tugikeskuse kontaktandmed on saadaval ELRC veebisaidil www.lr-coordination.eu.

Kord aastas korraldab ELRC üle-euroopalisi keelekonverentse, kus osalevad Euroopa Komisjoni esindajad ning EL-i riikide avaliku sektori, uurimisasutuste ja tehnoloogiaettevõtete töötajad. Esimene selline konverents toimus Riias 2015. aasta aprillis mitmekeelse digitaalse ühtse turu tippkohtumise raames. Konverentsi materjalid on saadaval Riia tippkohtumise (www.rigasummit2015.eu) ja ELRC veebisaidil.

Picture 3 CEF.AT and ELRC activities



Automaattõlkesüsteemi keeleressursid

CEF-i automaattõlkeplatvormi teenuste kvaliteedi parandamiseks vajavad tõlkesüsteemid kõigi CEF-i programmis osalevate riikide ametlike keelte jaoks mahukaid keeleressursid: tekste, sõnastikke (sh terminoloogiasõnastikke), geograafiliste jt nimetuste, märgendite, lühendite jms andmebaase ja muud taolist.

Liikmesriikide ja CEF-iga seotud riigiasutused, vabaühendused ning eraorganisatsioonid loovad iga päev suures mahus väärtuslikke keeleandmeid. Enamikku neist andmetest saaks kasutada CEF-AT platvormi täistamiseks.

Need keeleressursid aitavad parandada nii masintõlkesüsteemide üldist kvaliteeti kui ka valdkonnaomaste masintõlgete kvaliteeti. Vajalike ressursside hulka kuuluvad nii suured rööpkorpused (näiteks õigusaktid ja muud dokumendid ning nende tõlked) kui ka ühekeelsed korpused (näiteks eestikeelsed tekstikogumid), mida paljud EL-i liikmesriigid loovad oma keelte jaoks. Spetsialiseeritud masintõlkesüsteemide väljatöötamiseks kuluvad ära mitte ainult üldised ressurssid, vaid ka spetsiifilised konkreetse valdkonna keeleandmed – kindlale valdkonnale iseloomulikud tekstid, nende tõlked, terminid, klassifikaatorid jms. Eriti olulised on avalike internetiteenustega seotud keeleandmed.

Tekstikorpused – see on suuremahuline mitmesuguste tekstide kogum, mis on tavaliselt kokku kogutud elektrooniliselt ning seotud tarkvaraga, mis hõlbustab selle lingvistilist analüüsi. Tekstikorpuse loomiseks kogutakse kokku suure hulga eri dokumentide, tekstiühikute, veebilehtede jms sisu.

Rööpkorpus – see on tekstikorpused kahes keeles (lähte- ja sihtkeeles), kus igal lähtekeelsel tekstil on sihtkeelne tõlge. Näiteks moodustavad rööpkorpuse Eesti Vabariigi eestikeelsed õigusaktid ja nende ingliskeelsed tõlked. Tänapäevased automaattõlkesüsteemid suudavad rööpkorpustest statistiliste meetodite abil „õppida“ tõlkimise seoseid.

Terminoloogiasõnastikud – kakskeelsed valdkonnaspetsiifilised sõnastikud, mis sisaldavad termineid ja nende tõlkeid (st lähte- ja sihtkeele termineid). Terminoloogiasõnastikud võimaldavad automaattõlkesüsteeme kohandada konkreetsete valdkondade jaoks, nii et vastava valdkonna tekstide tõlgetes kasutataks termineid järjekindlalt ja õigesti.



ELRC konsortium

Euroopa Komisjon



Saksamaa tehisintellekti uurimise keskus



DFKI-see kuuluvad Kaiserslauternis, Saarbrückenis ja Bremenis paiknevad uurimis- asutused, samuti Berliinis asuv projektbüroo. DFKI on Saksamaa juhtiv uurimiskeskus uuendusliku äritarkvara tehnoloogia valdkonnas, kus kasutatakse tehisintellekti.

Tilde



Tilde on keeletehnoloogia ettevõtte, mis on spetsialiseerunud kohandatud masintõlkesüsteemide ja pilvepõhiste terminologiateenuste väljatöötamisele ning tõlke- ja lokaliseerimisteenuste pakkumisele. Ettevõtte on asutatud 1991. aastal ning sellel on filiaalid Riias, Tallinnas ja Viinuses.

Euroopa keeleressursside hindamise ja levitamise amet



ELDA tegutseb Prantsusmaal Pariisis ning selle eesmärk on soodustada keeleressursside kasutamist inimkeeletehnoloogia (Human Language Technology – HLT) sektoris ja hinnata keeletehnoloogiat.

Keele- ja kõnetöötuse instituut



Keele- ja kõnetöötuse instituut on Kreeka haridus- ja usuaasjade ministereeriumi uurimistöö ja tehnoloogia peasekretariaadi egiidi all tegutsev uurimisasutus.

Automaattõlke kasutajate ühendus



TAUS on üleilmsete keele- ja tõlketööstuse ressursside keskus, mille peakorter asub Amsterdams. Selle eesmärk on saavutada parem tõlke kvaliteet uuendusliku lähenemise ja automatiseerimisega.



Picture 4 Language resources and ELRC consortium



ELRC Eesti seminari kava

Aeg: 11. veebruar 2016
Kohd: Tallinn, Euroopa Liidu Maja, Rävälä pst 4

9.00–9.30	Registreerimine ja kohv
9.30–9.55	Avamine ja tervitussõnad Saila Rinne, EC DG CONNECT Jana Laurend, Euroopa Komisjoni esindus Eestis Kadri Vidar, Eesti Keeleressurside Keskus Andrejs Vasiljevs, ELRC/Tilde
9.55–10.30	Mitmekeelsus Euroopas Saila Rinne, EC DG CONNECT
10.30–11.00	Eesti keel digiajastul Kadri Vidar, Eesti Keeleressurside Keskus
11.00–11.35	Arutelu: mitmekeelsed avalikud teenused Eestis Katrin Ringo, RIA riigiportaal.eesti.ee Remo Tiigirand, EAS, e-residentsuse projekt Aleksandr Belousov, RIK
11.35–12.05	Kohvipaus
12.05–12.30	Tõlkimise tulevik Rihards Kalniņš, ELRC/Tilde
12.30–12.50	Automaattõlge: kuidas see toimub? Kasutusstsenariumid Mark Fishel, Tartu Ülikool Martin Luts, ELRC/Tilde



ELRC Eesti seminari kava

12.50–13.25	Arutelu: kakskeelsed andmed asutuste infosüsteemides Kaja Kell, Eesti Pank Helve Trumann, Riigikogu Kantsel Kadri Vidar, Eesti Keeleressurside Keskus Margit Kurm, ELRC/Tilde
13.25–13.45	Õigusraamistik Viljar Peep, Andmekaitse Inspektsioon
13.45–14.45	Lõuna
14.45–15.15	Kuidas saab automaattõlkeplatvorm CEF AT riigiasutustele kasulik olla? Saila Rinne, EC DG CONNECT
15.15–15.30	Kuidas asutuse andmeid automaattõlkeks ette valmistada? Andrejs Vasiljevs, ELRC/Tilde
15.30–16.15	Rühmatöö ja kohv
16.15–16.30	Kokkuvõte



Pilt: Euroopa Komisjon

Picture 5 Agenda



Picture 6 Back cover