



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable Task 6

ELRC workshop report for Sweden



Author(s):	Lars Borin (Språkbanken) Rickard Domeij (Language Council of Sweden)
Dissemination Level:	Public
Version No.:	<V1.1>
Date:	2016/04/27

Contents

Contents.....	2
1 Executive summary.....	3
2 Dissemination of Workshop.....	4
3 Workshop agenda.....	5
4 Summary of content of sessions	7
4.1 Morning session.....	7
4.2 Afternoon session	8
5 Synthesis of workshop discussions	11
5.1 Workshop presentations.....	12
5.2 Invitation email (in Swedish).....	13
5.3 Press release for the event (In Swedish).....	15

European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.

1 Executive summary

The Swedish ELRC Workshop took place in Stockholm, on the 10th of March 2016 at Europahuset ("The House of Europe"), where the local EC representation has its office in Sweden.

The event was attended by nearly 50 participants mostly from public agencies.

The program featured presentations from the governmental, translation and language technology sectors. Both the audience and the speakers had a positive attitude towards the CEF platform and showed interest in how Swedish public administration can provide language resources to the European Commission in order to develop CEF.AT.

All presentations are provided on ELRC Sweden Workshop [website](#). This website contains all presentations in pdf format and are supplemented with video recordings of presentations.

Information about the workshop was disseminated through the information channels of the Language Council. Invitations to the workshop were sent out to key persons in the area of public information, communication and translation at public agencies and other organizations. A press release was sent out by the EC representation's office, which also video broadcast the event.



European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.

2 Dissemination of Workshop

Localized invitations were sent out to potential workshop participants, where objectives of the event were explained and importance of ELRC activities highlighted.

Invitation email (in Swedish) is attached in the Appendix.

Information about the workshop was published on the EC representation office website and respective press release was distributed (see in Appendix).

3 Workshop agenda

9:30–10:00	Registration and coffee
10:00–10:30	Welcome and introduction Katarina Areskoug Mascharenhas, Head of the European Commission in Sweden Rickard Domeij, The Language Council of Sweden Andrejs Vasiljevs, ELRC
10:30–11:00	EU and multilingualism Maria Sjöström-Gisslén, EC in Sweden
11:00–11:30	Languages and language technologies in Sweden Lars Borin, Språkbanken (The Swedish Language Bank)
11:30–12:00	Translation services at the Swedish MFA Jeffrey Ganellen, Ministry for Foreign Affairs
12:00–13:00	Lunch
13:00–13:30	How can public institutions benefit from CEF.AT? Spyridon Pilos, DGT, European Commission (on Skype)
13:30–14:00	Machine translation—how does it work? Anna Sågvall-Hein, Uppsala University and Convertus
14:00–14:30	Machine translation: enabling multilingual communication for e-government Rihards Kalniņš, ELRC
14:30–15:00	Coffee break

European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.

15:00–15:30	Legal framework for contributing data Mikael Vall, The Ministry of Finance
15:30–16:00	Need of data and language resources Andrejs Vasiljevs, ELRC
16:00–16.30	Discussion and wrap up Lars Borin, Rickard Domeij och Maria Sjöström-Gisslén

4 Summary of content of sessions

4.1 Morning session

The morning session contained the welcoming on behalf of the local EC representation, The Language Council and The Swedish Language Bank, as well as a brief presentation of the aims of the workshop by Andrejs Vasiljevs, the ELRC representative.

The welcome speech was held by Katarina Areskoug Mascharenhas, Head of the European Commission's representation in Sweden.

Rickard Domeij (The Language Council of Sweden) presented information from the local organizers together with Maria Sjöström-Gisslén (DGT representative in Sweden), Lars Borin (The Swedish Language Bank in Göteborg).

Andrejs Vasiljevs from the ELRC presented the aim of the workshop. 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.

Maria Sjöström Gisslén from the local EC office held a presentation on the EU and multilingualism. The European union now has 24 official languages. According to the Treaty, regulations and other documents of general application shall be drafted in the official languages. According to Council Regulation 1/1958, citizens have a right to address the official EU bodies in any of the EU's official languages and to receive a reply in that language. There are approximately 4300 translators and 1000 interpreters working for the Institutions.



Lars Borin from the the Swedish Language Bank talked about the languages and language technologies in Sweden. He presented an overview of the state of language technology support for the Swedish language with reference to the META-NET report *The Swedish Language in the Digital Age*. He emphasized the importance of sharing language resources and developing language technology in Sweden to secure the future of the Swedish language and other languages in Sweden.

European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.



Jeffrey Ganellen from the Ministry for Foreign Affairs talked about Translation services at the Swedish Ministry of Foreign Affairs.

4.2 Afternoon session

Spyridon Pilos, DGT, European Commission, talked (via Skype) about how CEF.AT will build on the existing MT@EC service, putting emphasis on secure, quality, customizable MT. It will provide public institutions with faster and more secure translation, better quality translation in general and include translation for specific domains. However, for this to become a reality, a great deal of linguistic data is necessary.



Anna Sågvall Hein from Uppsala University and company Convertus delivered a presentation titled *Automated Translation: How does it work?* Anna Sågvall Hein explained how different types of MT-systems are constructed and used in a historical context: rule-

European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.

based, statistical and interlingua systems. She talked about pros and cons of the different types, with a focus on the statistical systems that are most often used today. She gave concrete examples of what can and cannot be done by machine translation.



Rihards Kalniņš, ELRC, had a presentation *Machine translation: enabling multilingual communication for e-government*. As a concrete example of how MT can be used within public agencies, Rihards Kalniņš described the use of automated translation by the government in Latvia, giving examples of use situations: web information to citizens, information to journalists, guests and tourists. He also talked about the data needed for creating such a system.

Mikael Vall, Ministry of Finance, presented an overview of the legal framework relevant for contributing data, focusing on the PSI-directive and the obligation of public authorities to give access to information.

In the last presentation Andrejs Vasiljevs talked more about the need of data and language resources for improving MT@EC and realizing CEF.AT, focusing on the practicalities of contributing data to ELRC.



European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 programme.

The last presentation was followed by a discussion where the audience had the chance to ask questions to Andrejs Vasiļjevs and Lars Borin as representatives for the ELRC. The discussion is summarized in the next section.

Rickard Domeij ended the workshop by thanking everybody for his or her contributions.

5 Synthesis of workshop discussions

The general discussion period at the end of the day was dominated by issues of language data and how these could be contributed. The following questions were raised.

Q: *You have discussed mainly MT between English and the other EU languages, but translations are also made between other EU language pairs. Which kind of translation data do you need?*

A: For practical reasons, the focus will be on translation to and from English, and a few other language pairs where the demand is high. Note that there is also a need for domain-specific translation data.

Q: *I have translation memories which may contain confidential material on the one hand, and which on the other hand reflect various stages of the translation process, but normally not the end result, meaning that they are full of typos and question-marked terms. How can I contribute?*

A: As for confidential data, it will be good to know that the original data will never be made available, only translation models built on it, but in cases of confidentiality it is probably better not to submit the data if in any kind of doubt. When it comes to the issue of partly erroneous data, even this kind of data can be extremely valuable, especially for domain adaptation purposes, and errors will probably be cancelled out by the large amounts of correct data coming from many sources.

Q: *What about copyrighted text?*

A: The EU does not have a fair-use clause (like the US), so there are obstacles, but there is also ongoing work aiming at changing European legislation in a more favorable direction for use of data in research and development. Current ELRC activities are focused on collecting public data but we are also interested in identifying copyrighted resources that may be of potential use for CEF.AT.

Q: *How much work will it entail on our part to contribute data?*

A: With small amounts, not much at all. A few minutes for uploading individual documents to the ELRC website is all that it takes, and even small amounts of data are extremely valuable, if they come from many sources. With larger document collections, there are various ways of streamlining the process, but these will generally have to be dealt with on a case-by-case basis. The first step should generally be to contact the ELRC helpdesk, which is manned by legal and technical experts to support data collection activities.

Q: *Why will the availability of CEF.AT be restricted to public administrations? Why will it not be available to everybody?*

A: The EC does not want to compete with the private sector. The existing MT@EC service is currently offered only to public services as part of the cooperation between the EU and its Member States of the EU, as well as Iceland and Norway to support e-Government and interoperability between public administrations. The possibility to offer CEF.AT to the private sector will be examined as part of a financial sustainability analysis which is ongoing.

5.1 Workshop presentations

Click on the links below to see the presentations.

EU and multilingualism [\(presentation\)](#)

Maria Sjöström-Gisslén, EC in Sweden

Languages and language technologies in Sweden [\(presentation\)](#)

Lars Borin, Språkbanken (The Swedish Language Bank)

Translation services at the Swedish MFA [\(presentation\)](#)

Jeffrey Ganellen, Ministry for Foreign Affairs

How can public institutions benefit from CEF.AT? [\(presentation\)](#)

Spyridon Pilos, DGT, European Commission (on Skype)

Machine translation—how does it work? [\(presentation\)](#)

Anna Sågvall-Hein, Uppsala University and Convertus

Machine translation: enabling multilingual communication for e-government [\(presentation\)](#)

Rihards Kalniņš, ELRC

Legal framework for contributing data [\(presentation\)](#)

Mikael Vall, The Ministry of Finance

Need of data and language resources [\(presentation\)](#)

Andrejs Vasiljevs ELRC

5.2 Invitation email (in Swedish)

Välkommen till seminarium om maskinöversättning inom EU!

Plats: Europahuset i Stockholm, Regeringsgatan 65, plan 2.

Datum: torsdag 10 mars 2016

Tid: kl. 9–c:a 16.30

Sista dag för anmälan: fredag 26 februari

Inom programmet Connecting Europe Facility utvecklar EU-kommissionen en digital plattform för offentliga e-tjänster. Plattformen erbjuder maskinöversättning för att göra det möjligt att kommunicera över språkgränserna inom EU. Myndigheter i medlemsländerna uppmanas att bidra till insamlingen av språkligt material för att träna maskinöversättningssystemet. Datainsamlingen samordnas av nätverket European Language Resource Coordination (ELRC). Syftet med seminariet är att presentera och diskutera EU-kommissionens planer och betydelsen av att myndigheterna delar med sig av data till projektet och hur det i praktiken ska göras.

Ur programmet:

- Seminariet öppnas av Katarina Areskoug Mascharenhas, chef för EU-kommissionen i Sverige.
- Maria Gisslén från EU-kommissionen i Sverige berättar om flerspråkigheten inom EU
- Lars Borin, chef för Språkbanken i Göteborg, talar om språk och språkteknologi i Sverige.
- Jeffrey Ganellen från UD berättar om översättningsarbetet på UD och på svenska myndigheter.
- Mikael Vall från Finansdepartementet går igenom juridiken för hur myndigheter kan dela med sig av data för träning av maskinöversättning och annat vidareutnyttjande.
- Representanter från ELRC berättar om initiativet: Varför? Vilka data behövs? Hur ska insamlingen gå till? Hur kan du bidra? Vad får du ut av det?

Seminariet är gratis och arrangeras av EU-kommissionen i samverkan med Språkrådet, Språkbanken i Göteborg och Europahuset i Stockholm. Skicka gärna vidare till intresserade och berörda personer!

Registrera dig på ELRC:s webbplats: <http://lr-coordination.eu/sv/sweden>. Fullständigt program kommer att publiceras inom kort. Obs! Ange om du vill ha specialkost i fältet Annan information.

Vänliga hälsningar

Rickard Domeij

Språkrådet

Institutet för språk och folkminnen

www.sprakochfolkminnen.se



Är EU:s språkliga mångfald ett hinder för kommunikation mellan myndigheter, näringsliv och medborgare?

Språken och den språkliga mångfalden är basen för europeisk kultur, handelsförbindelser och samarbete. Vi arbetar, studerar, reser, lever och driver företag över landsgränser och geografiska barriärer. Kan den digitala inre marknaden bli flerspråkig?

Hur kan vi samla in språkdata för att bygga maskinöversättningssystem och skapa bättre förutsättningar för flerspråkig offentlig service inom EU?

Bakgrund:

Den digitala inre marknaden (DIM), ett prioriterat mål för EU-kommissionen till 2020, är i grunden flerspråkig. Hittills har dock de flesta av Europas språk inte haft tillräckligt språkteknologistöd i DIM. I själva verket kan den så omhuldade europeiska flerspråkigheten emellanåt framstå som ett hinder när det handlar om fri kommunikation eller tillgång till offentliga tjänster och service, samt om möjligheter att driva företag och att få jobb. EU-kommissionen har nu tagit ett viktigt steg i riktning mot en genuint flerspråkig digital inre marknad.

Inom Fonden för ett sammanlänkat Europa finns maskinöversättningsplattformen FSE.AÖ (*Connecting Europe Facility Automated Translation platform (CET.AT)*) som ska möjliggöra flerspråkig och tvärspråklig kommunikation och dokumenthantering mellan nationella myndigheter och organisationer, företag och medborgare i EU. EU-kommissionen har beslutat att göra FSE.AÖ tillgängligt gratis för alla offentliga organ i EU. De nationella offentliga förvaltningarna måste dock aktivt bidra i arbetet för att göra FSE.AÖ till ett verktyg som ska kunna hantera det löpande behovet av flerspråkig offentlig service i Europa.

Projektet *European Language Resource Coordination* (ELRC) har startats i syfte att samla in språkliga data för att bygga maskinöversättningssystem åt offentliga institutioner över hela EU samt Island och Norge, med målet att bättre möta de dagliga behoven av flerspråkig offentlig service i hela Europa. ELRC är en offentlig datainsamlingsinsats av hittills oanade proportioner som syftar till att förse FSE.AÖ med språk- och översättningsdata (en- och flerspråkiga texter) med relevans för den dagliga verksamheten i de offentliga förvaltningarna i Europas länder.

Länk till maskinöversättningsplattformen FSE.AÖ:

https://joinup.ec.europa.eu/community/cef/og_page/catalogue-building-blocks#AT

5.3 Press release for the event (In Swedish)



Stockholm den 2 mars 2016

Är EU:s språkliga mångfald ett hinder för kommunikation mellan myndigheter, näringsliv och medborgare?

Språken och den språkliga mångfalden är basen för europeisk kultur, handelsförbindelser och samarbete. Vi arbetar, studerar, reser, lever och driver företag över landsgränser och geografiska barriärer. Kan den digitala inre marknaden bli flerspråkig?

Hur kan vi samla in språkdata för att bygga maskinöversättningssystem och skapa bättre förutsättningar för flerspråkig offentlig service inom EU? Detta och mycket annat kommer vi att diskutera på ett seminarium den 10 mars. Här kan du se hela [programmet](#)

Tid: Torsdag den 10 mars kl. 09.30-16.00

Plats: Europahuset, Regeringsgatan 65, plan 2, Stockholm

Anmäl dig [här](#)

Seminarier kommer även att [webbsändas](#).

Bakgrund:

Den digitala inre marknaden (DIM), ett prioriterat mål för EU-kommissionen till 2020, är i grunden flerspråkig. Hittills har dock de flesta av Europas språk inte haft tillräckligt språkteknologistöd i DIM. I själva verket kan den så omhuldade europeiska flerspråkigheten emellanåt framstå som ett hinder när det handlar om fri kommunikation eller tillgång till offentliga tjänster och service, samt om möjligheter att driva företag och att få jobb.

EU-kommissionen har nu tagit ett viktigt steg i riktning mot en genuint flerspråkig digital inre marknad.

Inom Fonden för ett sammanlänkat Europa finns maskinöversättningsplattformen [FSE.AÖ](#) (*Connecting Europe Facility Automated Translation platform (CET.AT)*) som ska möjliggöra flerspråkig och tvärspråklig kommunikation och dokumenthantering mellan nationella myndigheter och organisationer, företag och medborgare i EU. EU-kommissionen har beslutat att göra FSE.AÖ tillgängligt gratis för alla offentliga organ i EU. De nationella offentliga förvaltningarna måste dock aktivt bidra i arbetet för att göra FSE.AÖ till ett verktyg som ska kunna hantera det löpande behovet av flerspråkig offentlig service i Europa.

Projektet *European Language Resource Coordination* (ELRC) har startats i syfte att samla in språkliga data för att bygga maskinöversättningssystem åt offentliga institutioner över hela EU samt Island och Norge, med målet att bättre möta de dagliga behoven av flerspråkig offentlig service i hela Europa. ELRC är en offentlig datainsamlingsinsats av hittills oanade proportioner som syftar till att förse FSE.AÖ med språk- och översättningsdata (en- och flerspråkiga texter) med relevans för den dagliga verksamheten i de offentliga förvaltningarna i Europas länder.

Mer information:

Maria Sjöström Gisslén

E-post: maria.sjostrom-gisslen@ec.europa.eu

Tel: 08 563 444 42, 070 669 99 95

Therese Domisch

E-post: therese.domisch@ec.europa.eu

Tel: 08 562 444 29