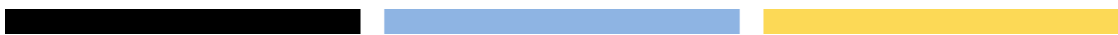


**European Language
Resource Coordination**
Connecting Europe Facility

ELRC Workshop Report for Finland



Author(s):	Imre Bartis (FIN-CLARIN)
Dissemination Level:	Public
Version No.:	V1.1
Date:	2016-03-21



Contents

<u>1</u>	<u>Executive Summary</u>	<u>3</u>
1.1	Workshop Presentations	3
1.2	Dissemination activities	4
<u>2</u>	<u>Workshop Agenda</u>	<u>5</u>
<u>3</u>	<u>Summary of Content of Sessions</u>	<u>7</u>
3.1	Morning session	7
3.2	Afternoon session	7
<u>4</u>	<u>Synthesis of Workshop Discussions</u>	<u>9</u>
4.1	Panel Multilingual Public Services in Finland	9
4.2	Panel Data and Language Resources in Finland	9
4.3	Panel How can Public Institutions benefit from the CEF.AT Platform	10
4.4	Localized agenda	11

1 Executive Summary

The Finnish ELRC Workshop took place in Helsinki, on the 19th of February 2016 at the University of Helsinki.

The event was attended by more than 30 participants from public administration, the Finnish language technology industry as well as from different universities.

The opening addresses were delivered by Kimmo Koskeniemi (FIN-CLARIN), Kimmo Rossi (European Commission), and Andrejs Vasiljevs (ELRC/Tilde).

The program featured a number of panel discussions with local experts 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 Finnish public administration can provide language resources to the European Commission in order to develop CEF.AT.



Picture 1 Andrejs Vasiljevs presenting objectives of ELRC



Picture 2 Krister Linden presents LT industry in Finland

1.1 Workshop Presentations

All presentations are provided on ELRC Finland Workshop website (<http://lr-coordination.eu/fi/finland>). This website contains all presentations in .pdf format and are supplemented with video recordings of presentations.

ELRC Workshop Report for Finland

1.2 Dissemination activities

The persons listed in the following subchapter have been sent an invitation to the workshop and, in case they didn't register, a reminder as well. After the workshop those who haven't attended the event were sent an email with links to the [press release](#) of the workshop and to the agenda containing the videos and presentations of the workshop. The participants received a shorter email with a link to the agenda.

FIN-CLARIN also tweeted during the workshop about the different ideas expressed there. After the workshop FIN-CLARIN tweeted about the event's press release.

In order to reach that segment of the Finnish media that might be interested in the topic of the workshop FIN-CLARIN contacted the Press Officers of the Faculty of Arts of the University of Helsinki and of the Department of Modern Languages of the university and asked them to send the email containing the link to the press release to their contacts. The Press Officers however suggested to have a meeting with FIN-CLARIN on March 16 about a communication strategy regarding possible dissemination activities. FIN-CLARIN accepted the suggestion, thus after March 16 we will be able to give more information regarding this issue.



KIELIPANKKI
The Language Bank of Finland

KIELIPANKKI KÄYTTÄJÄKSI AINEISTOT TYÖKALUT FOORUMI ORGANISAATIO TUKI IN ENGLISH PÅ SVENSKA

Suomen ELRC-työpaja 19.2.2016

EU:ssa on 24 virallista kieltä, joilla kansalaisia on voitava tasavertaisesti palvelua. Käännettävien tekstien määrä kasvaa jatkuvasti niin julkishallinnossa, liike-elämässä kuin muallakin yhteiskunnassa. Tämä lisää paineita monikielisen käännesteknologian kehittämiseen, etenkin kun Euroopan komission vuoden 2016 tavoitteena on rakentaa Euroopan yhtenäinen digitaalinen markkina-alue (YDMA). Monien kielten kohdalla riittämätön kieliteknologinen tuki haittaa julkisten palveluiden käyttöä ja liiketoimintaa.

Helsingissä järjestettiin 19.2.2016 Eurooppalaisten kielivarojen koordinaatio (The European Language Resource Coordination, ELRC) työpaja. Tilaisuuden tavoitteena oli saada suomalaiset julkishallintoimet huomaamaan, että niiden hallussa on runsaasti kiellaineistoja, joiden avulla monikielisen konekäännösjärjestelmän toimivuutta voitaisiin merkittävästi parantaa. Samalla käännösalue voitaisiin räätälöidä hallinnon omiin käännöstarpeisiin sopivaksi.

Työpajassa esiteltiin Euroopan komission tulevaa CEF-AT-konekäännös-alustaa, joka helpottaa mm. asiakirjojen ja muun kielellisen sisällön vaihtoa Euroopan kansallisten julkishallintojen sekä EU:n hallintoelinten, kansalaisten ja yritysten välillä. Työpajaan osallistui kutsuttuja puhujia Euroopan komissiosta, kansallisesta julkishallinnosta ja hallintoelimestä, kielipalveluja tarjoavista yrityksistä sekä kieliteknologian alalta.

CEF-AT-käännösalue perustuu ns. tilastolliseen konekääntämiseen, jossa järjestelmä oppii kääntämistä esimerkiksi perusteella. Käännösalueen kehittämisen kannalta hyödyllinen aineisto voi sisältää esimerkiksi yksikielisiä tekstejä, tekstien käännöksiä eri kielille, käännösmuisteja, sanastoja tai terminologioita, kunhan nämä ovat digitaalisessa muodossa. Aineistojen käytöstä pitää vain sopia.

Työpajassa sanottua

Työpajan yhteenveto-osuudessa puhujilta kysyttiin, millaisia tuloksia he työpajalta odottavat. He vastasivat näin:

Toivon, että avoin data lisääntyä myös julkishallinnon puolella. Tästä seuraisi mm. konekäännösalueen parantaminen ja siitä olisi hyötyä kansalaisille ja yrityksille, sillä avoimen datan avulla on mahdollista rakentaa monikielisiä liiketoimintaa. – Kimmo Rossi, Euroopan komissio

Toivon, että suomalaisten julkishallinnon organisaatioiden verkkosivujen alanurkkaan ilmestyisi CC BY -merkki. Vain tällä ehdolla verkkosivujen sisältöä saa käsitellä avoimena datana. – Kristofer Lindén, FIN-CLARIN

Toivon, että Suomessa olisi ryhmä julkishallinnon ihmisiä, joille on tärkeää, että suomen kieli saa riittävästi kieliteknologista tukea ja voi näin suoritus digitaalisen ojan haasteista. – Andreja Vaasilevs, ELRC

Katso työpajan videot ja esitykset osoitteessa http://lr-coordination.eu/finland_agenda

[Tutustu ELRC:hen!](#)

© 2015–2016 Kielipankki, FIN-CLARIN ja CSC – Tieteen tietekniikan keskus

Uutisia

- Maaliskuun Kielipankki-tutkija 10.3.2016
- ELRC-työpaja 19.2.2016
- Kipoc-korpus LÄT:ssä 27.11.2015

Lisää uutisia

Yhteystiedot

Kielipankin tekninen ylläpito:
kielipankki@iki.csc.fi

Aineistoihin ja palveluihin liittyvät asiat:
fin-clarin@iki.helsinki.fi
p. 029 4140599 / 029 4129317

2014
DSA
2015

Picture 3 Snap-shot of press release

2 Workshop Agenda

Agenda of the workshop was based on Master Agenda for the Workshop series and was adopted to suite best for the ELRC purpose in Finland.

08:15 – 09:00	Registration
09:00 – 09:30	Opening, Welcome, Aims and Objectives Kimmo Koskenniemi , FIN-CLARIN Kimmo Rossi , European Commission Andrejs Vasiljevs , ELRC
09:30 – 09:45	The EU and Multilingualism Kimmo Rossi , European Commission
09:45 – 10:10	Languages and Language Technologies in Finland Krister Lindén , FIN-CLARIN
10:10 – 10:50	Panel: Multilingual Public Services in Finland Moderator: Niko Papula , Kites Language Cluster Finland Speakers: Marjaana Kulovesi , Yle, the Finnish Broadcasting Company Aki Hietanen , Ministry of Justice Juhani Reiman , Lingsoft ltd Taru Virtanen , Prime Minister’s Office
<i>10:50 – 11:10</i>	<i>Coffee Break and Networking</i>
11:10 – 11:40	Automated Translation: How does it Work? Jörg Tiedemann , University of Helsinki
11:40 – 12:05	What data is needed? Mietta Lennes , FIN-CLARIN
<i>12:05 – 13:05</i>	<i>Lunch break and networking</i>
13:05 – 13:15	The Multilingual City of Helsinki Eila Ratasvuori , The City of Helsinki
13:15 – 14:00	Panel: Data and Language Resources in Finland Moderator: Lauri Carlson , University of Helsinki Speakers:

ELRC Workshop Report for Finland

Kaisamari Kuhmonen, Prime Minister's Office
Leena Peltomaa, AAC Global Ltd
Eila Ratasvuori, The City of Helsinki
Timo Simell, The Finnish Information Society
Development Centre

14:00 – 14:20	Data and Language Resources: Technical Issues Andrejs Vasiljevs , ELRC
14:20 – 14:50	Legal framework for Contributing Data Maria Rehbinder , Aalto University
<i>14:50 – 15:10</i>	<i>Coffee Break and Networking</i>
15:10 – 15:25	Eurooppaa yhdistävä automaattinen käännösalausta ja Avoimen datan portaali Kimmo Rossi , European Commission
15:25 – 15:50	The Future of Translation Rihards Kalniņš , ELRC
15:50 – 16:30	Panel: How can Public Institutions benefit from the CEF.AT Platform Moderator: Rihards Kalniņš , ELRC Speakers: Marja Hämäläinen-Herjo , Finnish Tax Administration Tanja Niemi , Ministry of Justice Tarja Varho , European Parliament
16:30 – 16:45	Interactive Session: Best Practice for the Future: Capitalize on your Valuable Data Krister Lindén , FIN-CLARIN Kimmo Rossi , European Commission Andrejs Vasiljevs , ELRC
16:45 – 17:00	Wrap-up and on site Conclusions Moderator: Tarja Salonen , Kites Language Cluster Finland

3 Summary of Content of Sessions

3.1 Morning session

The morning session contained the welcoming on behalf of FIN-CLARIN, ELRC and EC as well as a brief presentation of the aims of the workshop, EC's presentation on the EU and multilingualism, FIN-CLARIN's presentation on language technology in Finland, the panel discussion on multilingual public services in Finland, **Jörg Tiedemann**'s (University of Helsinki) presentation on how automated translation works, as well as **Mietta Lennes**' (FIN-CLARIN) presentation on what data is needed.

The presentation on language technology in Finland delivered by **Krister Lindén**, the Research Director of FIN-CLARIN, was centered around the various reasons why language technology is more and more vital for the Finnish society. Lindén pointed out that even though Finland has only two official languages (Finnish and Swedish), there are many other languages used on everyday basis in Finland as well: Sami languages, Romany, sign languages, Karelian, Russian, Estonian, Somali, English, Chinese, etc. Communication in various foreign languages is important e.g. in regional cooperation with Finland's neighboring countries. The need for language technology services might increase also because of the growing number of immigrants in Finland. Lindén talked also about FIN-CLARIN's contribution to the improvement of Finnish language technology by gathering language resources and making them available for public or for research use. The results of research, e.g. language technology tools, might be of benefit also for the language technology industry in Finland.

In his presentation on automated translation, **Jörg Tiedemann** (University of Helsinki) stressed that there has to be a serious increase in the amount of open Finnish language resources if we want to develop machine translation for Finnish.

3.2 Afternoon session

The afternoon session contained **Eila Ratasvuori**'s (City of Helsinki) presentation on the multilingual services of the City of Helsinki, the panel discussion Data and Language Resources in Finland, **Andrejs Vasiljevs**' (ELRC) presentation Data and Language Resources: Technical Issues, **Maria Rehbinder**'s (Aalto University) presentation on the legal framework for contributing data, **Kimmo Rossi**'s (European Commission) presentation on CEF.AT and the Open Data Portal, the panel discussion How can Public Institutions benefit from the CEF.AT Platform, as well as the Interactive session Best Practice for the Future: Capitalize on your Valuable Data, which was merged with the Wrap-up slot of the workshop.

Maria Rehbinder (Aalto University) pointed out in her presentation about the legal framework for contributing data that you can treat web content as open data only when the webpages have a Creative Commons (e.g. a CC-BY) logo.

The four speakers of the **Interactive Session: Best Practice for the Future: Capitalize on your Valuable Data** tackled from various angles the issue of engaging to ELRC's effort to gather language resources that could be used for improving CEF.AT for Finnish.

Krister Lindén (FIN-CLARIN) expressed his wish that Finnish public administration organizations will start using the Creative Commons tags on the websites that are meant to contain open data. **Kimmo Rossi** (European Commission) stated that hopefully this workshop would help increase the amount of open data in Finland. More open data for Finnish, he said,

ELRC Workshop Report for Finland

means: 1) a machine translation that better suits the translation needs of our language and 2) improved business opportunities for Finnish citizens and enterprises. **Tarja Salonen** (Kites Language Cluster Finland) suggested that many of those dealing with language resources in public administration might not be aware of the fact that language resources are valuable and that they can be used e.g. for developing a machine translation that works well for Finnish, too. Because of this it's important to get the workshop's message through to these persons. Finally **Andrejs Vasiljevs** (ELRC) envisioned that as a result of this workshop there would hopefully appear a group of people from the public sector who are passionate about their language and are willing to help their country and Europe to tackle the huge challenge that the digital age poses to us.

4 Synthesis of Workshop Discussions

4.1 Panel Multilingual Public Services in Finland

The participants in the panel presented the various multilingual services their organizations offer and talked about the machine translation related experience of their organizations.

Aki Hietaniemi (Ministry of Justice) talked about the <http://www.finlex.fi> site, containing at the moment legislation in Finnish, Swedish and English, and about the plans to make the website available as open data starting with spring 2016. The Finnish Ministry of Justice has tested the machine translation of the European Commission (MT@EC) that has been in use before CEF.AT, but the results hasn't been satisfactory.

Taru Virtanen (Prime Minister's Office) informed the audience that the Prime Minister's Office PMO provides all in all translations from and to 24 languages. The organization has tested MT@EC, but hasn't used it in production yet. According to her machine translation could help however in the near future to tackle the problem of the increasing amount of documents to be translated and of the tight translation deadlines.

Marjaana Kulovesi (Yle, the Finnish Broadcasting Company) told that besides the AV-translations the company does to and from Finnish and Swedish, the subtitling for hearing disabled people and TV-news in such languages as Sami languages, sign languages, English and Russian, Yle has also web pages in different languages, and that is where machine translation could be perhaps of help for the company. Machine translation hasn't been however in use at Yle.

Niko Papula (Kites Language Cluster Finland) pointed out that there seems to be a chicken or the egg problem with the Finnish machine translation. Right now machine translation for Finnish doesn't work that well, and because of that, people are not really interested in investing into it. Due to this lack of interest Finnish machine technology cannot really develop. When thinking about what kind of texts machine translation for Finnish is useful at the moment (templates, simplified language) the question of the source language of machine translation has a crucial role, that is why public administration organizations should consider what are the texts in which the use of templates and simple language could be of use both for the citizens and for the organization itself. Machine translation would probably result in quite reliable translations in the case of such texts.

According to **Juhani Reiman** (Lingsoft ltd) in the field of language services people in Finland are not yet aware of the changes that the increasing amount of open data on the internet will bring. However there will come a time when certain types of texts will be translated only with machine translation (just as there will be a time when driving a car won't need a human driver).

4.2 Panel Data and Language Resources in Finland

The participants in the panel presented the language resources their organizations have and expressed their thoughts about the need for more open data, the usefulness of machine translation and about how machine translation fitting the needs of Finnish public administration could be developed.

ELRC Workshop Report for Finland

Lauri Carlson (University of Finland) stated that texts to be used for the development of machine translation should be of a good quality, so that we would not end up with “garbage in, garbage out”. He also deemed that it might be that the major beneficiaries of a possible increase in the number of Finnish open data would be large multinational enterprises.

Leena Peltomaa (AAC Global Ltd) suggested that Finnish public administration should gather the translations its various organizations have, as well as their source texts (given they are allowed to be used publicly) into one place, in order to solve the problems fragmentation causes in the process of developing a machine translation to be used by all these organizations. She also pointed out that machine translation is just a tool to be used by translator, just like e.g. translation memories. According to her the improvement of machine translation will lead to an increased need for translators.

Kaisamari Kuhmonen (Prime Minister’s Office) affirmed that the Prime Minister’s Office is considering to open this year for public use some of its translation memories and termbases. She also stated that the Prime Minister’s Office will test MT@EC to see if it could be used in translating webpages of the organization.

Eila Ratasvuori (City of Helsinki) stated that machine translation could be useful for the City of Helsinki when communicating about small but urgent things (e.g. letting parents know what their children should bring to school next day) especially in the case of rare languages.

Timo Simell (TIEKE The Finnish Information Society Development Centre) drew attention to the large context of the issue of developing a machine translation for the needs of the Finnish public administration, saying that immigration, multiculturalism and the growing number of texts to be translated increase the pressure on translation services.

4.3 Panel How can Public Institutions benefit from the CEF.AT Platform

The participants in the panel focused on the question of the possible use of machine translation in general and of CEF.AT in particular in their organizations.

Marja Hämäläinen-Herjo (Tax Administration) considered that machine translation might be used in the near future by the Finnish Tax Administration, given that the number of translators will probably drop while the amount of documents to be translated will most likely increase.

According to **Tanja Niemi** (Ministry of Justice) machine translation might be useful in the communication between courts or justice ministries from different EU countries in cases when the legal issues concerned cannot be sufficiently explained by using the forms that have been specifically designed for tackling problems raised by multilingualism in cross-border scenarios.

Tarja Varho (European Parliament) talked about her experience with MT@EC as a Finnish translator of the European Parliament and stated that the output quality (which still varies though according to the type of document to be translated) has improved considerably since her first test in 2013. At the moment the Finnish translators of the European Parliament can choose whether they want to do their translations with the help of machine translation or not.

4.4 Localized agenda

Aika: 19.2.2016

Paikka: Metsätalon sali 1, Unioninkatu 40, Helsinki

08:15 – 09:00 **Ilmoittautuminen**

Työpajan tavoitteet ([video](#))

09:00 – 09:30 **Kimmo Koskenniemi**, FIN-CLARIN
Kimmo Rossi, Euroopan komissio
Andrejs Vasiljevs, Eurooppalaisten kielivarojen
koordinaatio ([esitys](#))

09:30 – 09:45 **Euroopan Unioni ja monikielisyys ([video](#))**
Kimmo Rossi, Euroopan komissio ([esitys](#))

09:45 – 10:10 **Kielet ja niiden kieliteknologinen tuki Suomessa ([video](#))**
Krister Lindén, FIN-CLARIN ([esitys](#))

**Paneelikeskustelu: Monikieliset julkiset palvelut
Suomessa ([video](#))**

10:10 – 10:50 **Moderaattori:**
Niko Papula, Kites ry
Puhujat:
Marjaana Kulovesi, Yle
Aki Hietanen, Oikeusministeriö
Juhani Reiman, Lingsoft Oy
Taru Virtanen, Valtioneuvoston kanslia

10:50 – 11:10 Kahvi- ja verkottumistauko

11:10 – 11:40 **Konekääntäminen: miten se toimii? ([video](#))**
Jörg Tiedemann, Helsingin yliopisto ([esitys](#))

ELRC Workshop Report for Finland

11:40 – 12:05	Millaista dataa tarvitaan? (video) Mietta Lennes , FIN-CLARIN (esitys)
12:05 – 13:05	<i>Lounas- ja verkottumistauko</i>
13:05 – 13:15	Monikielinen Helsingin kaupunki (video) Eila Ratasvuori , Helsingin kaupungin tietokeskus (esitys)
	Paneelikeskustelu: Kieliaineistoja Suomessa (video) Moderaattori: Lauri Carlson , Helsingin yliopisto Puhujat: 13:15 – 14:00 Kaisamari Kuhmonen , Valtioneuvoston kanslia Leena Peltomaa , AAC Global Oy Eila Ratasvuori , Helsingin kaupungin tietokeskus Timo Simell , TIEKE Tietoyhteiskunnan kehittämiskeskus Ry
14:00 – 14:20	Kieliaineistoihin liittyviä teknisiä ja käytännön kysymyksiä (video) Andrejs Vasiljevs , Eurooppalaisten kielivarojen koordinaatio (esitys)
14:20 – 14:50	Kieliaineistojen jakamisen oikeudellinen kehys (video) Maria Rehbinder , Aalto-yliopisto (esitys1 , esitys2)
14:50 – 15:10	<i>Kahvi- ja verkottumistauko</i>
15:10 – 15:25	Eurooppaa yhdistävä automaattinen käännoalusta ja Avoimen datan portaali (video) Kimmo Rossi , Euroopan komissio (esitys)
15:25 – 15:50	Kääntämisen tulevaisuus (video) Rihards Kalniņš , Eurooppalaisten kielivarojen koordinaatio (esitys)

Paneelikeskustelu: Konekäännösinfrastruktuurin

hyödyt ([video](#))

Moderaattori:

Rihards Kalniņš, Eurooppalaisten kielivarojen
koordinaatio

15:50 – 16:30

Puhujat:

Marja Hämäläinen-Herjo, Verohallinto

Tarja Varho, Euroopan parlamentti

Tanja Niemi, Oikeusministeriö

Interaktiivinen osuus: Hyödynnä omaa arvokasta dataasi

(tulevaisuuteen suuntautuneita parhaita käytäntöjä) ([video](#))

Krister Lindén, FIN-CLARIN

Kimmo Rossi, Euroopan komissio

Andrejs Vasiļjevs, Eurooppalaisten kielivarojen
koordinaatio

16:30 – 16:45

Avoin keskustelu, johtopäätökset, yhteenveto

Moderaattori: Tarja Salonen, Kites Ry

16:45 – 17:00