European Language Resource Coordination (ELRC) is a service contract operating under the EU’s Connecting Europe Facility SMART 2015/1091 programme.

Deliverable D3.2.24
Task 8

ELRC Workshop Report for Norway

Dissemination Level: Public
Version No.: V1.0
Date: 2019-04-25
European Language Resource Coordination
ELRC Workshop Report for Norway

Contents

1 Executive Summary 3
2 Workshop Agenda 4
3 Summary of Content of Sessions 7
  3.1 Welcome and introduction 7
  3.2 Norway’s participation in CEF Digital 7
  3.3 How does eTranslation work? 7
  3.4 The European Language Resource Coordination – actions and services 8
  3.5 TermBank and other actions for eTranslation in Sweden 8
  3.6 Usage of eTranslation in the Online Resolution Portal 8
  3.7 eTranslation and Norwegian language policy 8
  3.8 Panel discussion 1: How to ensure representation of both forms of Norwegian in eTranslation and other language technology services? 9
  3.9 Collecting language resources in Norway for ELRC – challenges and possibilities 9
  3.10 Translating for public sector clients 10
  3.11 Coordinating public procurement of translation services in Sweden 10
  3.12 Translation memories, privacy, and confidentiality 10
4 Synthesis of Workshop Discussions 12
  4.1 Success stories and lessons learnt 12
5 Workshop Participants 13
6 Appendix 14
  6.1 Localised invitation letter 14
  6.2 Dissemination of the workshop 15
  6.3 Localized agenda 16
1 Executive Summary

The second ELRC workshop in Norway took place on 20 November 2018 in the National Library’s premises in Oslo.

Information about the workshop was disseminated through the National Library’s usual communication channels (website, Facebook and printed brochure). Personal invitations were sent to key people in the areas of public information, translation and language technology. The Norwegian Language Council and the Norwegian Agency for Public Management and eGovernment assisted in spreading invitations to their networks.

Forty-five participants from a wide and relevant range of organizations, both public and private, attended the event. The workshop programme followed the ELRC template. The morning session was devoted to the topic of "Connecting a multilingual Europe: European context and local needs", while the afternoon session was devoted to "Engage: hands-on data". Both sessions ended with panel discussions, which included questions from the audience. During lunch and coffee breaks, there were lively conversations and networking among the participants.

The workshop was streamed online for virtual participation, and a complete video recording is available at https://bit.ly/2uKq2gw and on ELRC Webs site: http://www.lr-coordination.eu/l2norway_agenda.
# Workshop Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Session Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30 – 10:00</td>
<td><strong>Registration and coffee</strong></td>
</tr>
<tr>
<td>10:00 – 10:10</td>
<td><strong>Welcome</strong>&lt;br&gt;Presenter: Hege Høsøien (National Library of Norway)</td>
</tr>
<tr>
<td>10:10 – 10:30</td>
<td><strong>Session 1. Connecting a multilingual Europe: European context and local needs</strong></td>
</tr>
<tr>
<td>10:10 – 10:30</td>
<td><strong>Norway’s participation in CEF Digital</strong>&lt;br&gt;Presenter: Sarah Hails Gjelsten&lt;br&gt;(Norwegian Agency for Public Management and eGovernment)</td>
</tr>
<tr>
<td>10:30 – 10:50</td>
<td><strong>How does eTranslation work?</strong>&lt;br&gt;Presenter: Koenraad De Smedt (University of Bergen)</td>
</tr>
<tr>
<td>10:50 – 11:10</td>
<td><strong>The European Language Resource Coordination – actions and services</strong>&lt;br&gt;Presenter: Andrejs Vasiljevs (Tilde, ELRC representative)</td>
</tr>
<tr>
<td>11:10 – 11:30</td>
<td><strong>Coffee Break</strong></td>
</tr>
<tr>
<td>11:30 – 11:50</td>
<td><strong>TermBank and other actions for eTranslation in Sweden</strong>&lt;br&gt;Presenter: Simon Dahlberg&lt;br&gt;(Swedish Language Council)</td>
</tr>
<tr>
<td>11:50 – 12:10</td>
<td><strong>Usage of eTranslation in the Online Resolution Portal</strong>&lt;br&gt;Presenter: Ragnar Wiik&lt;br&gt;(European Consumer Centre Norway)</td>
</tr>
</tbody>
</table>
12:10 – 12:30  eTranslation and Norwegian language policy  
Presenter: Åse Wetås  
(Language Council of Norway)

Panel discussion 1
How to ensure representation of both official forms of Norwegian in eTranslation and other language technology services?  
Moderator: Jon Arild Olsen  
(National Library of Norway)

12:30 – 13:00  Panellists: Hege Høsøien (National Library of Norway), Sarah Hails Gjelsten (Norwegian Agency for Public Management and eGovernment), Koenraad De Smedt (University of Bergen), Ragnar Wiik (European Consumer Centre Norway), Åse Wetås (Language Council of Norway), Jan Olav Fretland (National Association of Nynorsk Municipalities)

13:00 – 13:40  Lunch

Session 2. Engage: hands-on data

13:40 – 14:00  Collecting language resources in Norway for ELRC – challenges and possibilities  
Presenter: Jon Arild Olsen  
(National Library of Norway)  
(Presentation)

14:00 – 14:20  Translating for public sector clients  
Presenter: Bjørn Ove Asprem  
(Semantix)  
(Presentation)

14:20 – 14:40  Coordinating public procurement of translation services in Sweden  
Presenter: Klas Ericsson (Kammarkollegiet)
14:40 – 15:00  
*Coffee Break*

**15:00 – 15:20**

**Translation memories, privacy, and confidentiality**

Presenter: **Øyind Grinde**  
(Norwegian Agency for Public Management and eGovernment) ([Presentation](#))

**15:20 – 15:50**

*Panel discussion 2*

**Outsourcing of public translations: security and public access**

Moderator: **Hege Høsøien**  
(National Library of Norway)

Participants: **Jon Arild Olsen** (National Library of Norway), **Bjørn Ove Asprem** (Semantix), **Klas Ericsson** (Kammarkollegiet), **Øyind Grinde** (Norwegian Agency for Public Management and eGovernment)

**15:50 – 16:00**

**Conclusions**

Presenters: **Hege Høsøien** and **Jon Arild Olsen** (National Library of Norway)
3 Summary of Content of Sessions

3.1 Welcome and introduction

Hege Høsøien, director of the department for research and dissemination at the National Library of Norway, welcomed the participants and presented the background for the National Library’s involvement in eTranslation and ELRC. In addition to having Norway’s largest digital media collection, the National Library hosts the Norwegian Language Bank. The Language Bank is a political initiative designed to further the development of technology solutions based on Norwegian. The goal is to prevent domain loss of Norwegian language in technology dependent areas. The Language Bank offers large collections of text and speech in machine-readable format offered to commercial developers, academic researchers, and public organisations.

The National Library of Norway also participates in the research infrastructure project CLARINO with funding from the Research Council of Norway. CLARINO constitutes the Norwegian part of the European CLARIN project. The ultimate goal is to make existing and future language resources readily accessible to researchers and to bring e-science to the humanities.

3.2 Norway’s participation in CEF Digital

Sarah Hails Gjelsten from the Norwegian Agency for Public Management and eGovernment (Difi) presented Norway’s participation in the digital part of the Connecting Europe Facility program. In order to support the development of a digital single market in Europe, CEF Digital promotes two types of components:

- Technological building blocks, which can be integrated into digital public services, either alone or in combination: eDelivery, eInvoicing, eID, eSignature and eTranslation, with eArchiving in development.
- Digital public services based on the technological building blocks: eProcurement, Cybersecurity, eHealth, eJustice, Online Dispute Resolution, Europeana, Safer Internet and Open Data.

Difi coordinates the Norwegian participation in CEF Digital. The Agency has set up a network of resource people who will help public agencies and private businesses with adopting the various CEF components.

3.3 How does eTranslation work?

Koenraad De Smedt, professor in computational linguistics at the University of Bergen and head of CLARINO, the Norwegian part of the CLARIN project, gave an in-depth presentation of the development of machine translation, from the first rule-based programs in the 60’s and 70’s to the statistically based approaches beginning in the late 80’s until the ongoing introduction of machine learning and neural networks. Koenraad emphasized the importance of domain adaptation, showing that machine translation based on an EU legal corpus gives very good results with legalese, but fails partially in other areas. Giving many examples, he demonstrated the improvements obtained with neural machine translation compared to purely statistical approaches.
3.4 The European Language Resource Coordination – actions and services

Andrejs Vasiljevs from Tilde and the ELRC consortium presented ELRC: how it is organised, the goals it aims to achieve, the actions undertaken and, finally, the services provided for data holders: extraction, cleaning, reformatting, alignment, anonymization and more. Andrejs stressed that ELRC can also give advice on legal questions and assist in improving data management processes. Assistance is provided by phone and email or directly by on-site visits from one of ELRC’s many experts, all of this free of charge for the data holders.

3.5 TermBank and other actions for eTranslation in Sweden

Simon Dahlberg from the Swedish Language Council presented the TermBank project funded by CEF Digital. Simon stressed the importance of domain-specific data and demonstrated the role of terminological data in adapting machine translation to specific domains. The TermBank project focuses on three domains related to different digital services infrastructures provided by CEF Digital: health (eHealth), business law (eJustice) and consumer rights (Online Dispute Resolution Portal). Simon also presented other measures taken by Swedish authorities in order to strengthen Swedish language in eTranslation, for instance collecting parallel texts from the web sites of public organisations.

3.6 Usage of eTranslation in the Online Resolution Portal

In Norway, eTranslation is so far only used in the Online Dispute Resolution Portal (ODR). This service permits consumers to send complaints to an approved dispute resolution body. The consumers can write their complaint in Norwegian, and eTranslation will provide an automatic translation into the language of the country where the disputed goods or services have been bought. Ragnar Wiik, director of the European Consumer Center Norway, presented experiences so far with machine translation in ODR. It quickly became clear that eTranslation do not translate well complaints written in a relatively informal language. The reason is that all Norwegian language data in eTranslation has been provided by the law department in the Norwegian Ministry of Foreign Affairs. To help ODR make better translations of Norwegian informal language, the European Consumer Center Norway translated a selection of authentic, but anonymized complaints originally sent by email. The resulting translation memories have been sent to Tilde, and the European Consumer Center Norway now hopes to see significant improvements in the results from eTranslation.

3.7 eTranslation and Norwegian language policy

Åse Wetås, director of the Norwegian Language Council, discussed some of the challenges that language technology in general and eTranslation in particular pose with regard to Norwegian law. The Norwegian Language Act requires that Norwegian citizens should be free to address national public bodies in any of the two official written languages, Bokmål and Nynorsk, and receive answers in their chosen written form. Today, eTranslation only translates from or into Bokmål. It is crucial that eTranslation and other language technologies used by public services follow the requirements of the Norwegian Language Act. The current expansion of language technologies from written to spoken language further increases these challenges.
3.8 Panel discussion 1: How to ensure representation of both forms of Norwegian in eTranslation and other language technology services?

Compared to major European languages, the general state of Norwegian language technology is not brilliant, but compared to languages of similar importance, e.g. the other Nordic languages, Norway has a reasonably well-developed infrastructure for both research and development, cf. The META-NET report from 2012. However, Norway has particular challenges due to the existence of two official forms of written Norwegian (Bokmål and Nynorsk) that are sufficiently different to be considered as different languages in most technology dependent contexts.

Why isn’t Nynorsk better represented in eTranslation and other language technology tools, and what can we do to change this? Nynorsk is a minority language used by approximately 12% of the population. The legal requirements for the use of Nynorsk in public administrations only apply to published text. Consequently, the amount of Nynorsk in internal documents (e.g. translations) is significantly lower.

In order to improve the functionality of Norwegian language, Bokmål as well as Nynorsk, in eTranslation, public administrations need to be made aware of the value of the textual data they produce. Today, this awareness is too often limited to statistical data related to administrative core tasks. It is also necessary to raise the awareness of the value of licensing web sites and other forms of public information so that they can be easily reused. It is also important to raise awareness of the existence and value of translation memories created when translation work is outsourced to private language services providers.

3.9 Collecting language resources in Norway for ELRC – challenges and possibilities

Jon Arild Olsen, senior adviser at the National Library of Norway and technological national anchor point, presented recent results from the collection of language resources for eTranslation. Unlike the situation in EU-countries, collecting language resources in Norway is not a supplementary activity, but constitutes the only source for Norwegian language data in eTranslation. The reason is that the Commission does not translate to or from Norwegian. Today, most Norwegian language data in eTranslation has been provided by the law department in the Norwegian Ministry of Foreign Affairs. These data have been supplemented with translations of anonymised complaints provided by the European Consumer Center Norway (cf. 3.6). The National Library has recently concluded an agreement with Amesto (now Semantix), Norway’s most important commercial language services provider. Because of the agreement, Amesto has freely transferred translation memories from public contracts. The National Library has also made an agreement with Norway’s largest standard organization, Standards Norway. The agreement permits the reuse of translation memories and multilingual terminology lists on the condition that the memories are submitted to random scrambling in order to prevent automatic reconstitution of the translated texts. Jon Arild finally presented two measures taken to ensure that eTranslation can handle Nynorsk. The first measure consists in digitizing and aligning sentences from textbooks for Norwegian primary and secondary schools. In accordance with the Norwegian Language Act, these textbooks are published in both official written forms. The National Library has applied for CEF funding to realize this measure. The second measure is the
production of a parallel corpus based on translation memories provided by two Norwegian news agencies, Norsk Telegrambyrå (NTB) and Nynorsk Pressekontor. According to an agreement recently concluded with the National Library, the two agencies will provide randomly scrambled translation memories from approximately 40,000 news reports translated from Bokmål into Nynorsk.

3.10 Translating for public sector clients


When the National Library approached Semantix in the autumn 2017, the company welcomed the proposal that they transfer translation memories from translations made for public customers. An importance factor behind Semantix’ acceptance of the proposal, was that the National Library acted on behalf of the Norwegian government with a clear mandate, and that the agreement made it clear that the National Library assumed responsibility for all questions relating to privacy, confidentiality and intellectual property rights in the further use of the data.

3.11 Coordinating public procurement of translation services in Sweden

Klas Ericsson from the Swedish Legal, Financial and Administrative Services Agency (Kammarkollegiet) gave a thorough presentation of the process and the results of a joint public tender for the acquisition of translation and interpretation services for public administrations at the state level. As a result of the tender, a framework agreement has been concluded with seven language services providers for a duration of two years. Klas gave a thorough review of the entire process, which started with an evaluation of the current LSP market in Sweden and the various needs of public administrations. Klas stressed the fact that a common public tender have many advantages. It is efficient and cost effective, and it increases the quality of the tender process. It also alleviates the need for each public administration to follow up the agreement. Of particular interest in the context of eTranslation, is that the framework agreement clearly stipulates the client’s ownership, not only of the translations but also of the translation memories. This stipulation is often forgotten when contracts are made between public administrations and LSPs on an individual basis. A coordinated national tender process can therefore be of great assistance in the collection of translation memories.

3.12 Translation memories, privacy, and confidentiality

Øyind Grinde, head of information security at the Norwegian Agency for Public Management and eGovernment, discussed challenges related to the management of privacy and security issues in public digital services in general. He specifically focused on the challenges created by cloud-based technology and outsourcing of ICT services. He stressed that both are considered a necessary and partly desired development by Norwegian authorities, but that they require increased awareness in public administrations of the kind of data they produce as well as of how different kinds of data must be treated (risk assessment).

Very few Norwegian public administrations have an ongoing need for translation that is large enough for them to have an internal translation service. The consequence is that more than half of all translations in the public sector are carried out by private language services providers. What can be done to make these translations useful in eTranslation? The agreement between the National Library and Amesto/Semantix should be followed up with similar agreements with other LSPs. Hopefully, it will make it easier to conclude new agreements. Such agreements would be made easier if the public clients’ ownership to translation memories were clearly stipulated in all contracts made with private LSPs. The Norwegian Agency for Public Management and eGovernment could assist in this by developing recommendations and templates for agreements with LPSs.

Confidentiality and privacy are two important limitations for open use of public data. What are the consequences for assignments that private translation companies perform for the public sector? Unlike eTranslation, most automated translation services available online (such as Google and Bing) cause data to be shared with the service provider. This fact does not seem to be widely known among employees in public administrations. By raising their awareness of this fact, eTranslation should be much more widely used as a stand-alone service. In contrast to private MT engines, you can choose to have your data deleted within 24 hours, so it is safe to use for confidential documents.
4 Synthesis of Workshop Discussions

In Norway, the organization of translation services in central public government offices has not changed since the previous workshop in June 2016. There is no central repository or translation office shared by all ministries as is the case in Finland; nor have the ministries themselves specifically committed themselves to managing and performing translations, with the exception of the Ministry of Foreign Affairs. The rest simply contract LSPs, whenever a need arises. However, a recent agreement between the National Library and Norway’s most import LSP company Semantix raises hope that similar agreements can be made with other companies in the LSP sector. This would greatly increase the access to translation memories of public origin.

A second important topic was measures to increase the amount of data in Nynorsk, the second official written language in Norway. The addition of Nynorsk to eTranslation has been decided by the Norwegian Ministry of Culture, which is responsible for Norwegian language policy. The main challenge for the inclusion of Nynorsk is the fact that very few parallel texts are available in Nynorsk and English. Existing parallel corpora are made up almost exclusively of Norwegian Bokmål and English. In order to meet this challenge, the National Library, in dialogue with the developers of eTranslation, has decided to collect a substantial parallel corpus of Bokmål and Nynorsk for the purpose of using Bokmål as a pivot language in eTranslation. The most important source of parallel texts in Norwegian Nynorsk and Norwegian Bokmål are textbooks and teacher manuals for Norwegian schools, which must be published in both official forms of written Norwegian according to Norwegian law. On the basis of these publications, the National Library of Norway aims to create a parallel corpus. Another important source of parallel texts in Nynorsk and Bokmål is the Nynorsk Press Office, which is a member of the NTB (the Norwegian News Agency group). Their contribution will consist of a corpus of news texts translated from Norwegian Bokmål into Norwegian Nynorsk and covering various subject areas.

A third important topic was the need of raising awareness about the value of textual data produced by public administrations, whether published online or for internal use, and more specifically the value of translations done internally or outsourced. The need for awareness raising also applies to the privacy and confidentiality of documents sent out for external treatment, for example when such information is kept in the form of translation memories by external executives without the client being aware of this. Either the documents must be anonymised before dispatch, or the contract with external executives must ensure that translation memories are returned and/or deleted.

4.1 Success stories and lessons learnt

- Agreements with commercial companies (Amesto/Semantix, Standards Norway and two news agencies).
- Plans for the Norwegian Language Council and the Norwegian Agency for Public Management and eGovernment to develop recommendations and templates for agreements between public administrations and LSPs
- Contacts established during the workshop will hopefully lead to delivery of parallel texts from the Supreme Court of Norway and the Norwegian Institute of Public Health.
5 Workshop Participants

The Norwegian ELRC workshop received many registrations spanning a wide range of ministries and public organizations, and from LSPs and academia.

The workshop was well-attended – there were 47 participants. Table below illustrate distribution of participants per sectors.

<table>
<thead>
<tr>
<th>Number of participants</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public sector</td>
<td>22</td>
</tr>
<tr>
<td>Technology eco-system</td>
<td>13</td>
</tr>
<tr>
<td>(=research, academia, private companies, developers of digital services)</td>
<td></td>
</tr>
<tr>
<td>LSPs (=translators, interpreters, LSPs)</td>
<td>5</td>
</tr>
<tr>
<td>Other (EC and ELRC consortium representatives, secretariat support staff, and other participants that cannot be grouped into any of the above)</td>
<td>7</td>
</tr>
<tr>
<td>Total:</td>
<td>47</td>
</tr>
</tbody>
</table>
6 Appendix

6.1 Localised invitation letter

I samarbeid med European Language Resource Coordination arrangerer Nasjonalbiblioteket

seminar om eTranslation - EUs automatiske oversettelsestjeneste

Flerspråkighet er en europeisk kjerneverdi, men skaper også barrierer i en verden hvor stadig mer informasjon utveksles over landegrensene. Automatisk oversettelse kan bidra til å overvinne disse barrierene og muliggjøre et virkelig flerspråklig Europa.

Norge deltar i EUs infrastrukturprogram Connecting Europe Facility. Et av tilbudene i programmet er eTranslation, en automatisk oversettelsestjeneste som kan brukes av offentlig administrasjon i EØS-området.

Seminaret vil vise hvordan eTranslation brukes i daglig arbeid, og diskutere hvordan offentlige virksomheter i Norge kan styrke det norske språkets funksjonalitet i tjensten gjennom deling av språkdata.

Se programmet her: [http://www.lr-coordination.eu/no/node/859](http://www.lr-coordination.eu/no/node/859)

Seminaret er gratis, men påmelding er nødvendig.

**Tid:** 20. november, kl. 10-16.  
**Sted:** Nasjonalbiblioteket, Solli plass  
**Påmelding:** [http://www.lr-coordination.eu/no/l2norway_registration](http://www.lr-coordination.eu/no/l2norway_registration)

Vennlig hilsen,

Jon Arild Olsen  
Seniorrådgiver  
Nasjonalbiblioteket  
Tlf. 97006578
6.2 Dissemination of the workshop

Automatisk oversettelse i Europa
Seminar om EUs automatiske oversettelsesjeneste eTranslation


Arr: European Language Resource Coordination (ELRC) og Nasjonalbiblioteket.

Morgenrøde
Filmvisning

Fantastiske fremtider

Vi er i et fremtidig Norge. Apokalypten har blitt realitet, og sivilisasjonen har brott sammen. Gjennom disse dystre landskapet vandrer to skapninger som er like mistenksame i hverandre som de er i havet til det virkelige omgivelsene.

Morgenrøde (70 min. Anders Eltid Fjærlien, 2014). Visninga er delt av Nasjonalbibliotekets filmprogram "Siktøkk på sivilisasjonen - science fiction i norsk film".

ld: 18:00
6.3 Localized agenda

Når: 20. november 2018
Hvor: National Library of Norway, Henrik Ibsens gate 110, 0255 Oslo

Organiserende:
- The European Language Resource Coordination (ELRC) consortium
- Nasjonalbiblioteket

09:30 – 10:00  Registrering og kaffe
Velkommen
Taler: Hege Høsøien
(Nasjonalbiblioteket)

Del 1. Tilkoble et flerspråklig Europa: Europeisk kontekst og lokale behov

10:10 – 10:30  Norges deltakelse i CEF Digital
Taler: Sarah Hails Gjelsten
(Difi)

10:30 – 10:50  Hvordan virker eTranslation?
Taler: Koenraad De Smedt
(Universitetet i Bergen)

10:50 – 11:10  The European Language Resource Coordination – tiltak og tjenester
Taler: Andrejs Vasiljevs
(Tilde, ELRC representant)

11:10 – 11:30  Kaffepause

11:30 – 11:50  TermBank og andre tiltak for eTranslation i Sverige
Taler: Simon Dahlberg
(Det svekse Språkrådet)

11:50 – 12:10  Bruk av eTranslation i Klageportalen
Taler: Ragnar Wiik
(Forbruker Europa)

eTranslation og norsk språkpolitikk
Taler: Åse Wetås
(Språkrådet)

Paneldebatt 1

Hvordan sikre representasjon av begge målformer i eTranslation og andre språkteknologiske tjenester?
Ordstyrer: Jon Arild Olsen

12:30 – 13:00  Innsamling av språkressurser i Norge for ELRC – utfordringer og muligheter
Taler: Jon Arild Olsen
(Nasjonalbiblioteket)

13:00 – 13:40  Lunsj

Del 2. Språkdata i praksis

13:40 – 14:00  Oversetter for kunder i offentlige sektor
Taler: Bjørn Ove Asprem
(Semantix)

14:00 – 14:20  Samordning av offentlig anskaffelser av oversettelsesomnестer i Sverige
Taler: Klas Ericsson
(Kammarkollegiet)

14:40 – 15:00  Kaffepause

15:00 – 15:20  Oversettelsesomnester, personvern og konfidensialitet
Taler: Øyind Grinde
(Difi)

Paneldebatt 2

Hvordan vokter vi personvern og konfidensialitet i eTranslation?
Ordstyrer: Hege Høsøien
(Nasjonalbiblioteket)
Participants: Jon Arild Olsen (National Library of Norway), Bjørn Ove Asprem (Semantix), Klas Ericsson (Kammarkollegiet), Øyind Grinde (Difi)
Konklusjoner
Talere: Hege Høsøien og Jon Arild Olsen
(Nasjonalbiblioteket)