ELRC Workshop Report for Norway

Authors:	Sarah, Hails Gjelsten (Difi - Agency for Public Management and e-Government) Koenraad, De Smedt (Bergen University)
Version No.:	<V1.2>
Date:	2017-01-31

Contents

1   Executive Summary

The Norwegian ELRC Workshop took place on 8 June 2016 at Difi's course and conference venue in Oslo.

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

The workshop was also streamed online for virtual participation.

The program featured presentations from government, translation and language technology sectors.

It was commented upon by the audience and speakers that it was due time for the topic of automatic translation to be addressed.

The presentations were in Norwegian, except for the presentations by EC and ELRC participants which were in English.

All presentations are provided on ELRC Norway Workshop web page, including links to presentations in pdf format and video recordings of the presentations.

Information about the workshop was disseminated through Difi's communication channels. Invitations to the workshop were sent out to key persons in the area of public information, translation and language technology research organizations.

2   Workshop Agenda

09:00 – 09:10	**Opening and welcome** Koenraad De Smedt (Bergen University) Jan Hjelle (Ministry of Local Government and Modernisation)
09:10 – 09:15	**Welcome by the EC** Saila Rinne (EC Programme Officer EU Policies, Unit G3 Data Value Chain)
09:15 – 09:30	**Aims and objectives** Andrejs Vasiļjevs (Tilde/ELRC)
09:30 – 10:00	**Europe and Multilingualism – How can Public Institutions benefit from the CEF Automated Translation platform**? Saila Rinne (EC Programme Officer EU Policies, Unit G3 Data Value Chain)
10:00 – 10:45	**Multilingual public services in Norway** Facilitator: Sarah Hails Gjelsten (Difi) Presenters: Eva Tverberg (Ministry of Children and Equality) Mona Myran (Directorate of Diversity and Integration) Grete Borge (Ministry of Foreign Affairs) Siri Risnes (Directorate of Labour and Welfare) Sverre Hovland (Brønnøysund Register Centre) Peder Nicolai S. Omenaas (Directorate of Immigration)
10:45 – 11:15	*Coffee break*
11:15 – 11:40	**Norwegian in the Digital Age – Machine translation: how does it work?** Prof. Koenraad De Smedt (Bergen University)

11:40 – 12:00	**Machine Translation in the Public Sector: An Example** Rihards Kalniņš (Tilde/ELRC)
12:00 – 12:15	**Legal Framework for sharing data in Norway** Øystein Åsnes (Difi)
12:15 – 13:00	*Lunch*
13:00 – 13:30	**Development and use of machine translation in the Norwegian Public Sector** Gard Jenssen (Ministry of Local Government and Modernisation), Gro Mæle Liane (Oslo Economics)
13:30 – 14:15	**Sharing Norwegian Data** Facilitator: Koenraad De Smedt (University of Bergen) Participants: Åse Wetås (Language Council of Norway Arne Martinus Lindstad (Language Bank "Språkbanken", National Library) Gisle Andersen (Norwegian School of Economics)
14:15 – 14:35	**How can we engage? – Language Resource Coordination** Andrejs Vasiļjevs (Tilde/ELRC)
14:35 – 15:00	*Coffee break*
15:00 – 15:30	**Discussion: Future practice: Using shared data** Facilitator: Tor Alvik (Difi) Participants: everyone
15:30 – 15:40	**Summary and the road ahead:** Prof. Koenraad De Smedt (University of Bergen), Sarah Hails Gjelsten (Difi)

3   Summary of the workshop content

The workshop comprised a series of mini presentations with representatives from the following agencies:
- Ministry of Foreign Affairs,
- Ministry of Children and Equality,
- Directorate for Integration and Diversity,
- Directorate of Immigration,
- Directorate for Labour and Welfare,
- National Register Centre,
- Difi's department for Public Open Data,
- Norwegian Language Council,
- Norwegian Language Bank,
- Norwegian School of Economics.

The audience were encouraged to pose questions and make comments throughout the workshop.

3.1   Opening and welcome

The workshop began with Professor De Smedt from the University of Bergen explaining his involvement in organising the workshop and thanking the Agency for Public Management and e-Government (Difi) for their cooperation. Thereafter, Jan Hjelle, Director General at the Ministry of Local Government and Modernisation gave some words of welcome. Hjelle emphasized the importance of automatic translation for Norway's participation in the CEF and ISA² programmes. He mentioned that Norway faces particular challenges, but hopes that automated translation will be a prioritized area of work for public agencies in Norway. He and the Norwegian government expect results in this area. Saila Rinne from the EU Commission also wished the participants of the workshop welcome. She was pleased to see that so many had chosen to attend.

3.2   Aims and objectives

Andrejs Vasiljevs from Tilde/ELRC presented the aims of the workshop: to raise awareness and engage institutions on the identification and sharing of data. Thereafter Saila Rinne from the Commission presented in more detail the CEF.AT platform and the expected benefits for public agencies from the platform.

3.3   Panel discussion "Multilingual public services in Norway"

Senior advisor at Difi and CEF Digital coordinator for Norway, Sarah Jane Hails Gjelsten, led the session about "Multilingual public services in Norway". Public agencies from Norway were represented in this session. There was common agreement that it was about time that Norwegian public agencies gathered and discussed this topic.

Among the speakers were Eva Tverberg from the Ministry of Children and Equality, Mona Myran from the Directorate for Integration and Diversity, Grete Borge from the Ministry of Foreign Affairs, Siri Risnes from the Norwegian Labour and Welfare Administration (NAV), Sverre Hovland from the Brønnøysund Register Centre and Peder Nicolai S. Omenaas from the Directorate of Immigration.

Policy in Norway for providing multilingual information and services is based on the Norwegian Public Administration Act, which states that Norwegian public agencies are responsible for ensuring that information and guidance is provided for all groups and individuals within society. Thereby, responsibility for translating texts and adapting information lies with each public agency.

The speakers had a great deal to say about their needs in working with multilingual public services in Norway.

The Directorate of Integration and Diversity reported that 13 percent of the Norwegian population has immigrant backgrounds. In 2015, the 10 largest immigrant groups had backgrounds from Poland, Lithuania, Sweden, Somalia, Germany, Ireland, Denmark, the Philippines, Pakistan and Eritrea. Immigrants settle throughout Norway. Refugees are settled according to an immigration settlement program. 300 different languages are represented or spoken in Norway today, and there is a need for interpreters for more than 100 languages.

The representative from the Ministry of Children and Equality, Eva Tverberg, spoke about the coming connection of Norway to the European consumer complaints portal Online Dispute Resolution. As of February 2017, Norwegian consumers will be able to use the portal to submit complaints about goods and services they have bought abroad. They will be able to submit a complaint in Norwegian, and the written complaint will be received electronically, for example by a hotelier in Turkey. Via the CEF.AT tool, the hotel manager will be able to read and respond to the complaint in his or her mother tongue. The challenge is ensuring the quality of the translation between Norwegian and other languages – it needs to be improved so that comprehension is optimal.

Translation routines within the Norwegian Public Sector vary from agency to agency.

The Ministry for Foreign Affairs has the greatest translation activity among Norwegian public agencies. It has two translation units: one which works on EEA-related legal translations and another which works on the translation of speeches, documents and other texts for the Ministry itself.

It appears that those agencies which outsource their translation tasks, have a rather adhoc approach to ordering translations and do not have a system for gathering, structuring and archiving translated texts.

Agency workers are not aware of the value of translation memories. The possibility and value of obtaining translation memories from translation bureaus as yet is not known.

The participants at the workshop expressed a need for working smarter in aligning information in foreign languages for their respective user groups.

Several participants were positive towards automated translation and the pooling of datasets to work smarter. However, the question of how this should be organised was not fully discussed, only that coordination and cooperation is required. It was suggested that a follow-up meeting be organised to discuss this matter further.

3.4 Norwegian in the Digital Age – Machine Translation: how does it work?

Koenraad De Smedt from the University of Bergen briefly described the outcome and implications of the META-NORD study of the status of Norwegian in the Digital Age, which was published as a META-NET White Paper. The status of Norwegian in Information and Communication Technologies is still much worse than that of English. However, some

resources and tools for R&D are available through national infrastructures such as Språkbanken (Language Bank) and CLARINO.

Professor De Smedt then explained main challenges, approaches and technologies behind Machine Translation, with an emphasis upon Statistical Machine Translation (SMT).

3.5   Machine translation: Enabling multilingual communication in e-government

Rihards Kalniņš from ELRC/Tilde presented the importance of MT in society, such as e-Commerce, for which it has the power to engage consumers across Europe, and e-Government, and how it can engage citizen participation and provide better access to services. Kalniņš presented the example of Canadian MT-policy and the example of MT for e-Government in Latvia. Since 36% of the Latvian population speaks Russian at home, the challenge was to enable residents to use e-Services in their native language. Latvia built up an innovative MT-service called Hugo, adapted it to relevant domains including EU Presidency texts. It was made available on various user-friendly platforms such as apps and kiosks, and it played an important role for visitors and the media at the EU Presidency headquarters.

3.6   Legal framework for Open data sharing

Senior Advisor at Difi and DSI coordinator in Norway for Open Public Data, Øystein Åsnes, gave a very interesting presentation about the framework for sharing data in Norway. Åsnes leads a team of advisors who work to encourage Norwegian public agencies in sharing their public data. Difi is also responsible for running a data hotel where people can find and share Norwegian public datasets.

The Norwegian government promotes the opening up of and sharing of public data according to legal directives and the Public Administration Act. Agencies should make their open data available in machine.readable format and register it according to the Norwegian Licence for Public Open Data (NLOD).

During his presentation, Åsnes said that 72 % of public agencies in Norway know of other agencies that have information that they could use, yet only 55 % of public agencies know that they have information which could be share with others.

Difi advocates that the gains are stronger than the risks when it comes to sharing data.

3.7   Automated translation in the Norwegian Public Sector

The timing of the workshop coincided with the publishing of a report carried out by the Ministry of Modernisation and Local Government (KMD), on the needs for automated translation in the Norwegian state sector and foreseen costs for collecting necessary data. Gard Jenssen from KMD and Gro Mæle Liane from Oslo Economics presented the main findings from the report.

In the report, it is estimated that Ministry of Foreign Affairs spends 0,5 million EURO per year on translation, the Ministry of Justice carries out a great deal of translation work too, translating legal texts and spending 120 000 EURO per annum.

Other agencies such as the Public Roads Administration, Marine Authority and Finance Inspectorate spend approximately 40 000 EURO per year.

Several agencies use external translation bureaus. Fees for translation from Norwegian to English is approximately 50 EURO per page, as much as 125 EURO per page (300 words) for Norwegian to Sami translations.

The report indicates that there are mixed opinions about the need for CEF.AT, as leading agencies with in-house translation units already have their own automated translation tools.

Two areas identified in the report where automated translation may be beneficial are 1) as a web tool for translation to cut government translation costs and increase translation efficiency, and 2) as a component to develop new online, cross-border services.

The crux is ensuring that the CEF.AT tool translates well enough for the Norwegian language. That is not the case today.

Nevertheless, the report recommends an initial collection of existing data from public sector agencies for the CEF.AT system, and then creating a system for a continued delivery of corpora segments by public agencies to the CEF.AT in the future.

3.8   Panel discussion: Sharing Norwegian Language Data

Koenraad De Smedt led a panel that discussed language policy, resource sharing and the role of language resources with an emphasis on terminology.

Åse Wetås from Språkrådet (Language Council of Norway) explained official Norwegian language policy, which rests on political support for maintaining both written norms (Bokmål and Nynorsk).

Arne Martinus Lindstad from Språkbanken at the National Library of Norway described what Språkbanken has to offer: not only an infrastructure for sharing resources and tools, available to R&D, but also a growing set of services for analyzing language use in the library collections.

Gisle Andersen from the Norwegian School of Economics gave an overview of monolingual and multilingual language resources and focused on terminology, which is important for handling MT in specialized domains. He presented Termportalen (a part of CLARINO) which provides access to terminology resources in Norway.

3.9   How can we engage?

Andrejs Vasiļjevs from ELRC/Tilde explained the importance of collecting a sufficient amount of language data, especially multilingual data, in general and specialized domains. Data can come from many sources and can be in different formats. In addition to mining the web, the vast number of documents not available on the web should be considered for inclusion; these are often found in public service administrations. The ELRC can help the national collection of sufficient materials by providing assistance with legal advice, privacy processing (anonymization), model licensing agreements, format conversion and other data preparation, description and delivery. There are options for submitting data directly to the ELRC repository.

3.10 Panel discussion – Best practice for the future – using shared data

Tor Alvik from Difi led the final general discussion on the way forward. There was general consensus that public administrations in Norway should cooperate better on MT and that the ELRC initiative looks promising. There was, however, also a general feeling that this matter has not yet been sufficiently politically anchored and must find its proper place in the operations of the government bodies before real work with data collection can start. There is still an open question as to whether data should be sent directly to the ELRC, or first collected in Norway, for instance, at Difi or the National Library's "Språkbanken". A follow-up meeting after the summer seems to be in order.

3.11 Workshop Presentations

All presentations and video recordings are available via the links below:

09:00 09:10	–	Velkommen og åpning (video) Koenraad De Smedt (Universitetet i Bergen) Jan Hjelle (Kommunal- og moderniseringsdepartementet) Saila Rinne (EC Programme Officer EU Policies, Unit G3 Data Value Chain)
09:15 09:30	–	Aims and objectives Andrejs Vasiljevs (Tilde/ELRC) (presentation) (video)
09:30 10:00	–	Europe and Multilingualism – How can Public Institutions benefit from the CEF Automated Translation platform? Saila Rinne (EC Programme Officer EU Policies, Unit G3 Data Value Chain) (presentation) (video)
10:00 10:45	–	Panel: Flerspråklige offentlige tjenester i Norge (video) Ordstyrer: Sarah Hails Gjelsten (Difi)
11:15 11:40	–	Norsk i den digitale tidsalderen – Maskinoversettelse: hvordan fungerer det? Prof. Koenraad De Smedt (Universitetet i Bergen) (presentasjon)
11:40 12:00	–	Machine Translation in the Public Sector: An Example Rihards Kalniņš (Tilde/ELRC) (presentation) (video)
12:00 12:15	–	Juridiske rammer for åpen datadeling Øystein Åsnes (Difi) (presentation) (video)
13:00 13:30	–	Utvikling og bruk av maskinoversettelse i regi av det offentlige i Norge (video) Gard Jenssen (KMD) (presentation) Gro Mæle Liane (Oslo Economics) (presentation)
13:30	–	Panel: Deling av norske språkdata

14:15	Ordstyrer: Koenraad De Smedt (University of Bergen) Panedeltakere: Åse Wetås (Språkrådet) (video) Arne Martinus Lindstad (Språkbanken, Nasjonalbiblioteket) (presentation) (video) Gisle Andersen (Norges Handelshøyskole) (presentation) (video)
14:15 14:35	–	How can we engage? – Language Resource Coordination Andrejs Vasiļjevs (Tilde/ELRC) (presentation) (video)
15:00 15:30	–	Diskusjon: Beste praksis for fremtiden – Bruk av delte data (video) Ordstyrer: Tor Alvik (Difi)

3.12 E-mail invitation

Meld deg på workshop om automatisert oversettelse, 8. juni 2016

Mange virksomheter bruker mye tid og store ressurser på oversettelse av dokumenter. På seminaret 8. juni får du høre mer om hvordan automatisert oversettelse kan gi lettere og raskere kommunikasjon på tvers av landegrenser.

Kvaliteten på automatisk oversettelsesverktøy er avhengig av datagrunnlaget i verktøyet. Hvordan kan din virksomhet bidra til å forbedre og bruke den nye teknologien?

Norge skal delta i EUs digitale indre marked*. Folk skal ha bedre tilgang til digitale varer og tjenester på tvers av landegrenser i Norden og Europa, men språk kan være en barriere. Connecting Europe Facility-programmet (CEF) er EU-kommisjonens satsing som Norge deltar i som skal legge bedre til rette for grenseoverskridende digital infrastruktur og digitale tjenester. En byggekloss i arbeidet er en automatisk oversettelsesplattform (CEF.AT) som skal gjøre infrastrukturen og digitale tjenester flerspråklig.

På onsdag 8. juni 2016 skal representanter fra EU kommisjonen og prosjektet ELRC (European Language Resource Coordination) besøke Norge. I samarbeid med Direktoratet for forvaltning og IKT (Difi) og Universitetet i Bergen skal de holde en workshop om oversettelsesplattformen, CEF.AT. Under workshopen skal det være presentasjoner og diskusjoner om behov, rollen og muligheter som automatisert oversettelse kan ha for offentlig forvaltning i Norge, og verdien datadeling har i utvikling av digitale tjenester på tvers av landegrenser.

MELD DEG PÅ I DAG: For mer informasjon om workshopen, program og påmelding: http://www.lr-coordination.eu/no/norway

Har du spørsmål, ta kontakt med Sarah Hails Gjelsten sarah@difi.no, tlf. 906 609 44, eller Koenraad De Smedt desmedt@uib.no, tlf. 55 58 20 52

    *Meld.St. 27 Digital agenda for Norge (2015-2016)