Deliverable Task 6

ELRC Workshop Report for Iceland

Dissemination Level:	Public
Version No.:	V1.0
Date:	11-11-2016

Contents

1 ELRC Workshop in Iceland

The Iceland ELRC Workshop took place in Reykjavik, on the 11th of November, at the Culture House. Workshop was co-organized by ELRC Representative in Iceland – the Vigdís Finnbogadóttir Institute of Foreign Languages.

An event-dedicated web part was set up at the ELRC website before the event, at http://lr-coordination.eu/iceland. The page was populated with the agenda (http://lr-coordination.eu/iceland_agenda) and the online registration form (http://lr-coordination.eu/iceland_registration).

The opening addresses were delivered by Markus Foti (eTranslation/MT@EC) and Professor Auður Hauksdóttir (Vigdís Finnbogadóttir Institute). Next presentations were made by distinguished speakers from public administrations as the Permanent Secretary of the Ministry of Education, Science and Culture, Director of the Labor Directorate, Chief Officer of 112 emergency service, Project Leader of the Icelandic LT programme, representative from the Ministry of Foreign Affairs and others

From the very beginning a very positive tone for the workshop was set which showed how important is to collaborate for collecting of language resources to advance MT and language technology development for Icelandic. The speakers from academic sector provided a valuable overview of the developments of language resources and technologies in Iceland.

Both the audience and the speakers had a positive attitude towards the CEF platform and showed interest in how Iceland public administration can provide language resources to the European Commission in order to enchance quality of Icelandic language in MT@EC.

Picture 1 Tweet about opening of Icelandic workshop and opening of workshop

2   Workshop Agenda

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

8:30 – 09:00	Registration
09:00 – 09:20	Opening addresses **Markus Foti,** eTranslation/**MT@EC** Project Manager **Professor Auður Hauksdóttir**, Vigdís Finnbogadóttir Institute
09:20 – 9:30	Aims and Objectives **Andrejs Vasiljevs**, Tilde/ELRC **(Presentation)**
9:30 – 9:50	Europe and Multilingualism **Markus Foti**, eTranslation/**MT@EC** Project Manager
9:50 – 10:20	Automated Translation: How does it work? **Josef van Genabith** – DFKI/ELRC (**video**)
10:20 – 10:50	Coffee Break and Networking
10:50 – 11:20	Staða íslenskunnar í stafrænum heimi á tímum alþjóðavæðingar (Icelandic in the digital world in times of globalization) **Prof. Eiríkur Rögnvaldsson,** *University of Iceland* **(Presentation)**

11:20 – 12:00	Hvernig getur máltæknin gagnast samskiptum á íslenskri tungu? (What can language technology achieve for the Icelandic language?) Multilingual Public Services in Iceland *Tryggvi Axelsson*, *Director of Consumer Agency, Icelandic Language Council* **(Presentation)** Raddgreinar, talgerflar og hljóðstýring (speech recognition, synthetic voices…) **Jón Guðnason**, Reykjavík University **(Presentation)** Sjálfvirkar leiðréttingar (Automatic corrections) **Anna Björk Nikolásdóttir**, Reykjavík University **(Presentation)** Gagnasöfn (Corpora): Steinþór Steingrímsson **Árni Magnússon,** Institute of Icelandic Studies, University of Iceland **(Presentation)**
12:00 – 12:20	Panel discussion
12:20 – 13:20	Lunch Break
13:20 – 13:50	How can Public Institutions benefit from the CEF.AT Platform? **Markus Foti**, eTranslation/**MT@EC** Project Manager
13:50 – 14:10	Machine Translation in the Public Sector: Best practices **Rihards Kalniņš**, Tilde/ELRC **(Presentation)**
14:10 – 14:40	What Data is needed? Why? and Data and Language Resources: Technical and Practical Aspects *Andrejs Vasiljevs, Tilde/ELRC* **(Presentation)**
14:40 – 15:00	Vélræn samskipti á íslenskri tungu Snertifletir við höfundarétt **Erla S. Árnadóttir (Presentation)**
15:00 – 15:30	Coffee Break and Networking

15:30 – 16:00	Hagnýting máltækni og þörf fyrir vélræn samskipti (Applied language technology and the need for machine translations) Þýðingamiðstöð utanríkisráðuneytisins **Katrín Einarsdóttir,** Director of Ministry for Foreign Affairs - Translation Centre **(Presentation)** Þjónusta við erlenda ferðamenn (Service for foreign tourists) **Þórhallur Ólafsson,** Neyðarlínan 112, **(Presentation)** Vinnumarkaður og erlent starfsfólk (Labour market and foreign workers) Vinnumálastofnun (Directorate of labour)
16:00 – 16:30	Panel discussion
16:30 – 16:50	How can we engage? How organizations can be involved in ELRC activities, The work ahead: Best Practice for the Future - Capitalizing on your Valuable Data Kolbeinn Björnsson, Project Leader
16:50 – 17:00	Wrap-up, onsite conclusions and commitments

3   Workshop Participants

Iceland ELRC workshop received 41 registrations spanning a wide range of ministries, state and public organizations – potential contributors of language data. Workhop organisers were honored that Ms. Vigdís Finnbogadóttir, the former President of Iceland and UNESCO Goodwill Ambassador participated at the worshop.

The workshop was well-attended by 32 participants.

Picture 2 Workshop audience represented various public institutions

4   Summary of Content of Sessions

All presentations are provided on ELRC Iceland Workshop website. This website contains presentations in .pdf format.

4.1   Opening addresses

Markus Foti, eTranslation/**MT@EC** Project Manager

Professor Auður Hauksdóttir, Vigdís Finnbogadóttir Institute

4.2   Aims and Objectives

Andrejs Vasiļjevs, Tilde/ELRC **(Presentation)**

Andrejs Vasiļjevs from Tilde/ELRC presented the aims of the workshop: to raise awareness and engage institutions on the identification and sharing of data.

4.3   Europe and Multilingualism

Markus Foti, eTranslation/**MT@EC** Project Manager

In his talk, Mr. Foti highlighted that language diversity is anchored both in the European Charter of Fundamental Rights and in the Treaty on the European Union.

The Charter of Fundamental rights has three notions of protection for multilingualism:

1. Protection of linguistic and cultural diversity.
2. Protection against discrimination based on (race, ethnicity, religion and) language.
3. Right of the citizen to address the EU in their own (official EU) language, and the right to receive a response in the same language.

The Commission has been supporting **the use of language technologies in public administrations**. In particular, the Directorate-General for translation (DGT) has been over the last years increasingly investing in LT to support the human translation of EU legislative texts and to support national public administrations in their communication tasks. The machine translation system developed by DGT is currently available to staff of the EU institutions and bodies and to national administrations of EU Member States.

The European Commission is providing support to the **deployment of mature language technologies through the Connecting Europe Facility (CEF) Programme**. DG CNECT implements focused deployment actions under the CEF programme to make pan-European digital public services such as Europeana, Open Data Portal or Online Dispute Resolution platform multilingual and accessible across all EU languages. For this purpose, 12 MEUR funding has been allocated to Automated Translation in the work programmes 2014 and 2015 of CEF.

The Commission Communication on the Digital Single Market strategy was published by vice president Andrus Ansip and commissioner Guenther Oettinger on 6 May 2015. It sets up an ambitious agenda to fill the "missing link" of the single market and respond to the rapidly growing importance of the digital market where more and more transactions and economic

activity takes place. There are a number of barriers (legal, institutional, cultural) that hamper the achievement of the Digital Single Market. European Digital Market is multilingual:

- 90% of European consumers prefer to browse websites in their own language;
- 82% of the 4000+ web shops are mono-lingual;
- 42% of multilingual internet users never purchase online products and services in languages other than their own.

CEF programme supports the emergence of the Digital Single Market. The DSM has to be multilingual because pan-European public services address the whole EU, as opposed to national online public services.

Citizens want to use their own language when using public services online - pan-European online services serve users who speak 24 different languages, and who do not share any single common language.

Available online translators (e.g. Google Translate, Microsoft Bing) have gaps in language coverage (especially: the small EU languages as target language) and are not secure (e.g. patient data in eHealth systems cannot be sent to Google Translate).

The time has come to put the results of research and innovation actions in visible use at European level. EU support for multilingualism in Digital Europe:

- Over 200 MEUR invested since 2007 – this has led to advancement in language technologies
- important results of EC funding programmes: Machine translation toolkit Moses, META-SHARE language resources, MT domain adaptation pilots, standards and workflows for language processing, tools for terminology work …
- Language technologies can deliver **effective multilingual support to public online services** to complement the valuable work of translators
- Now it is time to put language technologies into use for public services!

CEF an opportunity to show that language technologies can deliver effective multilingual support to public online services.

4.4 Automated Translation: How does it work?

Josef van Genabith – DFKI/ELRC (**video**)

Prof. Josef van Genabith was not participated in the meeing in his person, during Iceland wokshop pre-recorded video presentation was provided to audience. In his presentation, prof. Josef van Genabith provided overview of the MT with example of the transationnof the Chinies food meanue. He in his presentation emphasized importance of the data to get better MT systems.

4.5 Staða íslenskunnar í stafrænum heimi á tímum alþjóðavæðingar

Prof. Eiríkur Rögnvaldsson, *University of Iceland* **(Presentation)**

Eiríkur Rögnvaldsson from University of Iceland provided oveview of Situation of the Icelandic language in a digitalized world in the era of globalization. He give insights into different opinions about the future of the Icelandic language:

- Gunnar Smári Egilsson, *Fréttatíminn*, 7.3. 2015: „When you think about it, it´s obvious that the Icelandic language will disappear in the next 50-70 years. At least as a complete language in peoples´ diversified communication.“
- Tryggvi Gíslason, *Eyjan*, 10.12. 2015: „[…] The Icelandic language has never been stronger than now as a living vernacular in a diversified society […] The Icelandic language […] will live for unforseen future.“

He expressed that Icelandic language will have big changes in society in few years due to influence of moder technology and change of habits that were influenced by - Smart phone revolution, Interactive videogames, YouTube- and Netflix-revolution, Increase in tourist-stream to Iceland, Increase in number of citizens with another mother language and Speech controlled devices.

Eirikur presented table according to UNESCO (2009) classification which provide Scale for language situation and expressed his opinion that Icelandic language is on safe side:

Degree of endangerment	Intergenerational Language Transmission
Safe	The language is spoken by all generations; intergenerational transmission is uninterrupted
Vulnerable	Most children speak the language, but it may be restricted to certain domains (e.g., home)
Definitely endangered	Children no longer learn the language as mother tongue in the home
Severely endangered	The language is spoken by grandparents and older generations; while the parent generation may understand it, they do not speak it to children or among themselves
Critically endangered	The youngest speakers are grandparents and older, and they speak the language partially and infrequently
Extinct	There are no speakers left

Profesor Eirikur presented main points of action which has been taken to safeguaer Icelandic language:

- Conclusion from the Parliament, May 2014, Usage of Icelandic in digital information technology - The Parliament concludes that the Minister of education and culture should appoint a committee of specialists in languistic and information technology which makes a plan of actions to make the Icelandic viable in the digital world and contribute to usage of Icelandic in that field. The plan should include a timeplan of actions and milestones, estimate of costs and funding. Committee was appointed in September 2014
- Parliament, on the day of the Icelandic language, 16.11. 2015, Statement from the Minister of education and culture:" It has to be and it will be, that in the coming times and years enough funding will be secured to do the work necessary to make the

Icelandic language viable in a digital world. The cost of doing this will be a small portion of the price we need to pay if the language disappears."

- Máltæknisjóður (a fund that funds projects related to Technology and Icelandic language) founded in spring 2015:
  - with 15 million ISK
    - But 40 million ISK where planned for the fund in the plan for 2015
  - Contribution 30 million ISK in the budget bill 2016
    - but 90 million ISK where planned for the fund in the plan for 2016
  - 50 million ISK as extra budget for 2016
    - for 8 months as a preparation project
    - map, strategy, technical implementation
    - Financial and project plan for 5 years

4.6   Hvernig getur máltæknin gagnast samskiptum á íslenskri tungu?

4.6.1   Multilingual Public Services in **Iceland**

***Tryggvi Axelsson**, Director of Consumer Agency, Icelandic Language Council* **(Presentation)**

In his presentation Mr. Axelsson presented some examples of the where multilingualism has high importance to reach it aims. Law nr. 2/1993 – validation of the EEA agreement: **ÍSLAND:** A part of the inner market of EEA/EU countries – a market of 500 million people. Under this law Iceland enjoy four parts of Freedom:

- Products and services
- Free flow of people within the area
- Free establishment and residency
- Financial funds

According to the statistic consumers from Iceland do the most electronic shopping within EEA-countries. This give ground for new challenges and importance for different solutions to provide efficient protection for consumers over borders – and within borders.

Consumer's protection in rapidly changing e-commerce environment has high importance to ensure that all can get best out of this e-commerce opportunities. And in case if sales go wrong – flaw in product, service not received, Consumers´ has the right and want to resolve issues without the court of law– fast and easy procedure.

He provided examples of the ODR- On line dispute resolution system. He expressed that future are in more sophisticated solutions to protect customers and any solution must be Electronic, simple and efficient and furthermore he mentioned that Machine translations are prerequisite for the future.

4.6.2   Raddgreinar, talgerflar og hljóðstýring

Jón Guðnason, Reykjavík University **(Presentation)**

Mr. Guðnason presented general ideas about speech analyzer and speech processor for Icelandic. He provided visual block-schemes to best describe communication and

human/computer interactions. He provided some of the examples where this technology can be applied and will serve great benefit for the users:

- Direct use
  - Communication with and controlling smartphones (Siri and company)
  - Doctors recording patients
  - Speeches at Parliament
  - Automated telephone answering
  - Assistance for blind and deaf people
- Indirect usage
  - Automated information procession from colloquial/vernacular
  - Research of evolution of colloquial/vernacular
  - Pronounciation and dialects
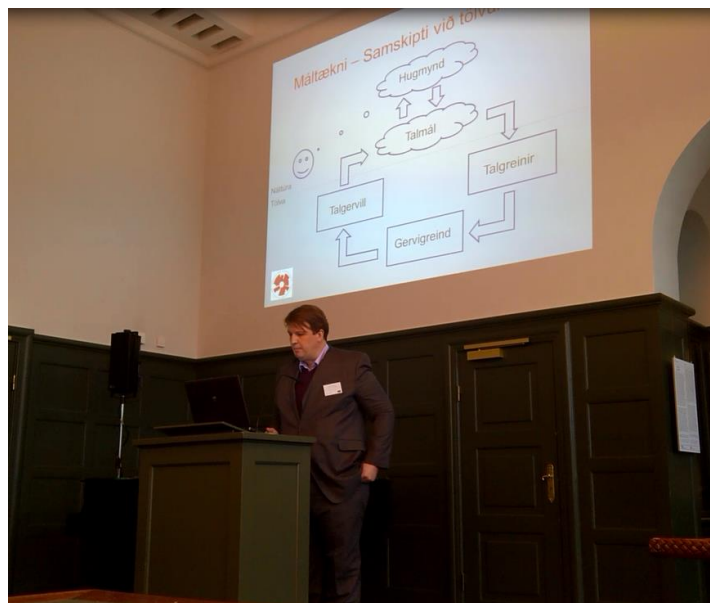
Picture 3 Jón Guðnason

4.6.3 Sjálfvirkar leiðréttingar (Automatic corrections)

Anna Björk Nikolásdóttir, Reykjavík University **(Presentation)**

Picture 4 Anna Björk Nikolásdóttir

4.6.4   Gagnasöfn (Corpora): Steinþór Steingrímsson

Árni Magnússon, Institute of Icelandic Studies, University of Iceland **(Presentation)**

In his presentation he provided overview of the The Databases of Árni Magnússon Institute. As he mentioned we have all used dictionaries. When we need to understand or express ourselves about texts about some specialties, specific encyclopedias can be useful to us. Those who create the dictionaries and the encyclopedias need to have access to textbooks with texts. In recent years, access to such data for the public has improved and one of such database is the databases that the Árnastofnun hosts and gives access to. He provided overview of the database and provided examples of it content. At Árnastofnun, they have the latest, Marked Icelandic Language, 25 million words from all sources. First version will be ready next year.

4.7   How can Public Institutions benefit from the CEF.AT Platform?

Markus Foti, eTranslation/**MT@EC** Project Manager

His presentation focused on the public services and machine translation. He generally described the role of MT and users who could use this technology. He describe that current MT@EC will transform into CEF.AT platform for use of MT in public institutions. He described several benefits of MT usage - usage of MT provide quick and cheap access to information in foreign languages, provide opportunity to understanding information received in a foreign language that otherwise could not be used or would require substantial time and costs to translate, as well as making multilingual use of websites possible.

The purpose of the Automated Translation building block is, on one hand, to provide multilingual support to other pan-European DSIs so that citizens, administrations and companies in all EU countries can use them in their own language, and, on the other, to offer

automated translation services to public administrations to facilitate cross-border information exchange.

CEF will support all Member States (plus NO and IS) to improve the quality of translation of their languages.

Member States can offer valuable raw material for developing automated solutions – texts and their translations, multilingual terminology lists etc. With the material that public administration can provide, technological solutions can make the automated translation of your language better.

- CEF AT helps to create accessible and multilingual **pan-European** online public services
- Technology underlying CEF.AT needs linguistic data (**language resources**) in EU languages
- Member States should **take ownership** of their own language and make sure their language is adequately supported in CEF AT
- Member States' involvement and help are essential for CEF AT efforts to succeed!
- **European Language Resource Coordination** (ELRC) is a project, funded through CEF Digital, that facilitates this process

He also provided some detailed information on CEF.AT (number of language pairs, corpora used to train MT systems, formats maintained, document translation features, feedback mechanism, etc.). He pointed out security as one of the most important issues. CEF.AT has already solved this problem and therefore users can feel safe when using this service. He also shared some information about the online services connected to CEF.AT. These services include the Internal Market Information System (IMI) Internal Market Information System – an intranet for more than 7000 public administrations in the EU Member States, SOLVIT an on-line problem solving network concerning misapplication of Internal Market law by public authorities, Tenders Electronic Daily (TED), Communication and Information Resource Centre for Administrations, Businesses and Citizens (CircaBC).

4.8 Machine Translation in the Public Sector: Best practices

Rihards Kalniņš, Tilde/ELRC **(Presentation)**

Rihards Kalniņš from Tilde/ELRC presented the example of 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, comprehensive, large-scale, machine translation service for the public sector, fully integrated into government platforms. Basic characteristics of the Hugo:

- o Novel methods of corpora collection
- o Balanced language models to get optimal results
- o Over 100M sentences of training data in multiple languages
- o Addition of public sector specific terminology
- o Integrated into e-services and gov't websites
- o Languages (currently):
  - Latvian-English

- English-Latvian
- Latvian-Russian

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.

4.9 What Data is needed? Why? and Data and Language Resources: Technical and Practical Aspects

Andrejs Vasiļjevs, Tilde/ELRC (Presentation)

Andrejs Vasiļjevs of Tilde/ELRC began by illustrating how the volume of data impacts machine translation quality. Then, he listed what is data in the framework of ELRC: anything that contains "words", preferences for "sentences", even for sentences expressed in multiple languages, good examples are reports, speeches, documents, web pages, brochures, etc. 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. Finally, he encouraged users to share data and recommended the first steps on that road, including the ELRC helpdesk service.

4.10 Vélræn samskipti á íslenskri tungu Snertifletir við höfundarétt

Erla S. Árnadóttir (Presentation)

In her presentation "Electronic communication in the Icelandic language associated with copyright" Erla S. Árnadóttir provided overviewof main issues related to IPR questions in terms of usng and re-using data. She described important explanations of terminology concepts behing translation and literature adjustments. Translating work from one language to another consist of copying and creating adjustments. This means it is neccessary to have authors´ permission to translate work and register it into a database. A database is a protected phenomena according to copyright law. The database is protected if the requirement of significant investment in providing, verifying or putting together information is fulfilled. A database protection consists of creating summary of the database or reusing its information. The one creating an adjustment, has the same right over it as the   author of the original work has over it, but respecting the original authors´ rights is necessary.

4.11 Hagnýting máltækni og þörf fyrir vélræn samskipti

4.11.1 Þýðingamiðstöð utanríkisráðuneytisins

Katrín Einarsdóttir, Director   of Ministry   for   Foreign   Affairs   -   Translation Centre (Presentation)

Katrín Einarsdóttir presented experience from her work at The Translation Centre of the Ministry for Foreign Affairs. The Translation Centre was founded in 1990. The Main tasks are to translate the Agreement on the EEA and the regulations, directives, decisions and

recommendations adopted under that agreement as well as translation of other international agreements and legal acts. The EEA Agreement has approx. 10,000 acts which have been incorporated into the EEA Agreement - it can be estimated that it corresponds to 70,000 OJ pages and even more (standard page is more than double the A4-page 12-point font with line spacing 1.5 (900-1300 words)).

The team working in Translation Centre of the Ministry for Foreign Affairs is built of 35 employees, who works under specific assignments in separate groups: (a) Society(5), (b) Finance(9), (c) Science (9), (d) Technology (6), and (q) Pre-processing information and back office processing, terminologies, other projects such as treaties (7). Groupmanagers manage setup within groups, both the project domain and how tasks are performed. Terminology team coordinates use of terms. Handles guidance for usage of terms. In theyre work they are using vairous tools that increase efficiency of tasks to be acomplished (e.g.Machine Translation vs Computer Assisted Translation (MT vs TM), CAT - Computer Assisted Translation, Trados, MultiTerm, Termbase). One of the biggest challanges in the translation work is the definition of new concepts and a search for Icelandic translations of these special concepts. New words are often required when the word search is not successful. All terms and phrases are collected in the Terminal of Translation Center. The concept library is published on the website of the Ministry of Foreign Affairs (http://www.hugtakasafn.utn.stjr.is/). The translation center's terminology contains about 70,000 entries and is growing. Displays the expertise that has accumulated in the translation center.

4.11.2 Þjónusta við erlenda ferðamenn (Service for foreign tourists)

Magnús Hauksson, Neyðarlínan 112, **(Presentation)**

Magnús Hauksson from The Icelandic Rescue center presented theyre experiance with working in multilingual enviroment. On January 1st 1996, the 146 emergency numbers at that time routed to 112. Primary tasks are receive emergency notifications, evaluate the need for assistance, dispatch the appropriate response parties, provide the response parties and the caller with service and information, prepare a report about the incidents. Magnús Hauksson provided general statistic of service Around 150 000 cases pr year (200 000 calls), over 50% for the Police, 13% for various health related inquiries (non emergencies), around 30% medical transport, about 3% fire and 4% other, SAR, Civil protection, Flight incidents etc,

 Telephone calls from foreign mobile phones 6.9% of total calls (ca. 10.500 of all cases), need for translate usually less than one per month as operators are fluent in English, Icelandic guides usually present for non-English speaking tourists. If Emergency Operator (EO) fails to understand caller, predefined steps must be taken by EO. EO must examines origins of phone number, attempts to confirm language, contacts translation services, translator either conferenced in or arrives on site, conference usually on within two minutes.

No applied language technology is in use, Google voice translation is under review. As one of the channels that can be used in emergency situations are social media (e.g. Messenger, Whatsapp, Lync, Skype etc.), to progress in social media the machine translation technology will be necesary.

5 Dissemination of ELRC activities in Iceland

Information about the workshop was disseminated through different communication channels.

Picture 5 Vigdis homepage: http://vigdis.hi.is/frettir/2016_10_31/malthing_um_velraenar_thydingar_og_velraen_samskipti_a_islenskri_tungu

Information about Workshop was shared through social network.

Picture 6 Information on Facebook

After the Iceland d workshop summary of outcomes was provided on Vidgdis homepage.

Picture 7 Vigdis homepage: http://vigdis.hi.is/frettir/2016_11_17/elrc_workshop_in_reykjavik_on_11_november_2016

An interview with Icelandic ELRC workshop speaker Professor Eirikur Rögnvaldsson, filmed during the recent **ELRC workshop in Reykjavik**, was broadcast on the **Icelandic National Broadcasting Service** (RÚV) evening news. In the interview, Professor Rögnvaldsson spoke about Icelandic in the digital age. The segment aired on November 16 to mark the celebration of the Day of the Icelandic Language.

Professor Eirikur Rögnvaldsson teaches the Icelandic language at the University of Iceland. He is also an expert in the field of natural language processing. At the ELRC seminar in Reykjavik, Professor Rögnvaldsson gave a lecture entitled "Icelandic in the Digital World in the Era of Globalization."

Click here to watch the segment on the Icelandic National Broadcast Service (begins at 17:34).
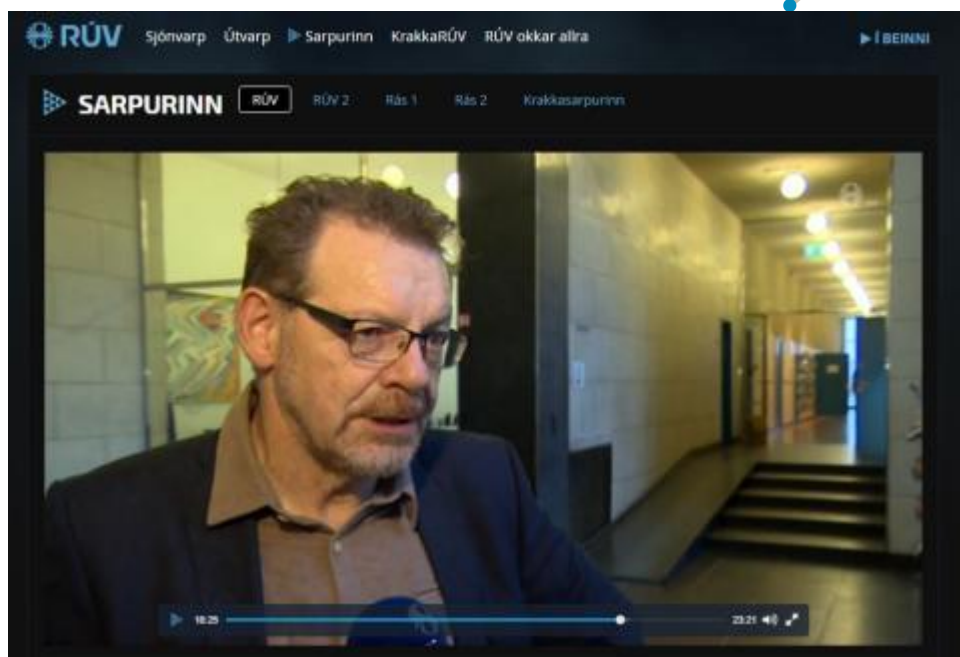
Picture 8 Professor Eirikur Rögnvaldsson