



EUROPEAN LANGUAGE EQUALITY

META  NET
META  FORUM 2022



The Icelandic Language in the Digital Age

Eiríkur Rögnvaldsson (Stofnun Árna Magnússonar í íslenskum fræðum)
eirikur@hi.is

08/09-06-2022 META-FORUM 2022 – Joining the European Language Grid (hybrid conference)
<http://www.european-language-grid.eu> – <http://european-language-equality.eu>

Achievements of the past few years

- **Language Technology Programme for Icelandic (LTPI), 2019-2022**
 - Budget ≈20 M€ – Mostly government funded
- **Iceland became a member of CLARIN and an Icelandic CLARIN centre was established**
 - All deliverables of the LTPI are available for free from CLARIN-IS under standard open licenses
- **The Government published an AI Strategy document in 2021**
 - The importance of developing LT resources and tools for Icelandic is explicitly mentioned
- **An LT consortium comprising academic institutions, NGOs and private companies was established**
 - The members of the consortium have cooperated closely on developing LT resources and tools
- **LT education and research has been strengthened considerably**
 - The number of researchers, students and developers working on LT has multiplied

Strengths and highlights

- **A number of new resources have been created within the LTPI, such as:**
 - *Icelandic Gigaword Corpus*, a general national corpus, 1.9 billion tokens
 - *Parlce*, aligned parallel corpus between Icelandic and English, 3.5 million translation units
 - *Samrómur*, audio corpus, 2.85 million sentences from 28,000 speakers, 247,800 minutes
- **A number of new tools have been created within the LTPI, such as:**
 - *Greynir*, a package that parses text and generates constituency trees, finds lemmas, inflects NPs, assigns PoS tags, and more
 - *GreynirTranslate*, IS-EN and EN-IS translation models based on a multilingual BART model
- **A number of tools are under development and will be finished by the end of the LTPI, such as:**
 - A new text-to-speech system
 - A new speech recognition system
 - A new machine translation system between Icelandic and English

Threats and shortcomings

- **The LTPI will end later this year and its future is uncertain**
 - The Government has indicated that the LTPI will be prolonged
 - However, no firm promises have been made and future funding is not secured
- **A great number of important resources and tools is still lacking, such as:**
 - spoken language corpora; parallel corpora between Icelandic and other languages than English; corpora for different purposes (sentiment analysis, question answering, summarisation), annotated multimodal corpora; term lists
 - sentiment analysis; summarization; question answering; natural language understanding; natural language generation; dialogue management; disambiguation; translation between Icelandic and other languages than English; speech translation; automatic subtitling; advanced speech synthesis (intonation, empathy); specialised grammar checking
- **More cooperation between academia and industry and more investment from industry is needed**

Recommendations for the SRIA

- Small languages need special dedicated LT programmes
- Special support for low-resourced languages where the LT market is small and not sustainable
- Products of the major international IT companies must be compatible with small languages



A delegation headed by the President of Iceland and the Minister of Culture visited five major international IT companies in May



European Language Grid
European Language Equality



Thank you!



The European Language Grid has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 825627 (ELG).

The European Language Equality project has received funding from the European Union under grant agreement № LC-01641480 – 101018166 (ELE),

Eiríkur Rögnvaldsson (Stofnun Árna Magnússonar í íslenskum fræðum)
eirikur@hi.is

08/09-06-2022 META-FORUM 2022 – Joining the European Language Grid (hybrid conference)
<http://www.european-language-grid.eu> – <http://european-language-equality.eu>

Session 8: Technology Support of Europe's Languages in 2022

- 11:20 Language Report Irish – Teresa Lynn (Dublin City University, Ireland)
- 11:25 Language Report Dutch – Frieda Steurs (Instituut voor de Nederlandse Taal, Netherlands)
- 11:30 Language Report Bulgarian – Svetla Koeva (Bulgarian Academy of Sciences, Bulgaria)
- 11:35 Language Report Welsh – Gareth Watkins (Bangor University, UK)
- 11:40 Language Report Icelandic – Eiríkur Rögnvaldsson (Institute for Icelandic Studies, Iceland)
- 11:45 Discussion