

# Steinunn Rut Friðriksdóttir

Reykjavík, Iceland  
srf2@hi.is / +354-844-0693

A PhD student in computer science with an emphasis on Icelandic language technology (or Natural Language Processing). My areas of interest primarily lie in the interdisciplinary context of language technology, encompassing data science, artificial intelligence, and linguistics, while also extending into various other fields, such as medicine, philosophy, and sociology.

## Education

### University of Iceland (2021–present)

PhD studies in computer science with a focus on language technology. My current project involves researching biases in language models, with a two-fold goal. Firstly, to research the manifestations of prejudice in the output of these models, i.e. how the biases of society are transmitted to the decision making of AI models, and secondly, to explore whether these models can be used to investigate the presence of prejudice in Icelandic texts.

A previous project involved localizing an entity linking system for Icelandic, which could be used to enhance search engines, question-answering systems, and other technologies that rely on rapid information retrieval.

Supervisor: Hafsteinn Einarsson, Associate Professor.

### University of Iceland, Reykjavik University, and University of Gothenburg (2018–2020)

Master's program in language technology. I completed the first year at the University of Gothenburg in Sweden and the second year in the joint program of the University of Iceland and Reykjavik University.

Thesis: *Á næsta leyti að vissu leiti? Using Confusion Sets for Automated Spell-Checking.*

Supervisor: Anton Karl Ingason, professor.

### **University of Iceland and Concordia University (2013–2016)**

Bachelor's degree program totaling 240 ECTS in Icelandic and French studies. I completed one semester as an exchange student at Concordia University in Montreal, Canada.

Final project in Icelandic: *The Fight Against Language Errors: Is It Effective?*

Supervisor: Eiríkur Rögnvaldsson.

Final project in French studies: *The Little Prince: A Musical*. Translation of the musical *The Little Prince* based on the work of Antoine de Saint-Exupéry.

Supervisors: Ásdís Rósa Magnúsdóttir, Professor, and Ásta Ingibjartsdóttir, Adjunct Lecturer.

### **Egilsstaðir Upper Secondary School (2009–2012)**

Graduated with a matriculation examination in the language track.

## **Teaching Experience**

### **University of Iceland (2022–2026)**

#### **Programming for Humanities Students and Programming in Language Technology (Fall 2023, Fall 2024, Fall 2025) – ÍSL333G and MLT701F**

Joint courses designed for BA students in humanities and MA students in language technology. These courses aim to support students at the beginner level in programming, helping them grasp fundamental concepts and providing them with training to solve simple yet diverse language technology tasks in Python. Additionally, they provide a brief overview of core principles in language technology.

#### **Introduction to Language Technology (Fall 2022, Fall 2023, Spring 2024, Spring 2025, Spring 2026) – TÖL025M**

This course introduces key aspects of language technology, focusing on the analysis and processing of written language. Core language technology tasks and their applications are presented, along with essential tools and language resources, with a specific emphasis on Icelandic language technology. Methods involving embeddings, deep neural networks, and language models are emphasized for practical tasks, with an exploration of how AI methods have largely taken over human language processing.

### **Entity Linking (Spring 2023) – MLT805F**

This course covers automated entity linking, the process of identifying and linking proper nouns in text to their correct and real-world entities in a knowledge base such as WikiData. The linking process allows the computer to deduce, based on context, the specific reference, even for ambiguous entities (e.g., determining if “Anne Hathaway” refers to the actress or Shakespear’s wife). Entity linking enhances search engine effectiveness, simplifies human-computer interactions, and improves information flow between various systems.

### **Selásskóli (2017–2018)**

Classroom teacher for 3rd grade, with occasional substitute teaching for all subjects from 1st to 7th grade, excluding swimming and physical education.

### **Independent Tutoring (2015–2018)**

Private tutoring through the website kenna.is, primarily in Icelandic for primary school students and French for secondary school students.

## **Other Relevant Work Experience**

### **The Árni Magnússon Institute for Icelandic Studies (2020–2021)**

Research, data, and software development within the institute's language technology team.

### **University of Iceland Systems Engineering Lab (2019)**

Language technology consulting for a study on duplicated usability error reports. Part-time position.

## **Publications**

Steinunn Rut Friðriksdóttir, Hafsteinn Einarsson. Detecting Biased Language in Icelandic: A Named Entity Recognition Approach for Socially Responsible Text Analysis. Forthcoming publication at the Third Workshop on Socially Responsible Language Modelling Research (SoLaR) 2025.

Steinunn Rut Friðriksdóttir, Dan Saattrup Nielsen and Hafsteinn Einarsson. (2025). Hotter and Colder: A New Approach to Annotating Sentiment, Emotions, and Bias in Icelandic Blog

Comments. *Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025)*.

Steinunn Rut Friðriksdóttir. (2024, November). The GenderQueer Test Suite. In *Proceedings of the Ninth Conference on Machine Translation (WMT24)*, 327–340. Association for Computational Linguistics.

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2024. Gendered Grammar or Ingrained Bias? Exploring Gender Bias in Icelandic Language Models. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024, Torino, Ítalíu)*, 7596–7610. ELRA and ICCL.

Steinunn Rut Friðriksdóttir, Annika Simonsen, Atli Snær Ásmundsson, Guðrún Lilja Friðjónsdóttir, Anton Karl Ingason, Vésteinn Snæbjarnarson and Hafsteinn Einarsson. 2024. Ice and Fire: Dataset on Sentiment, Emotions, Toxicity, Sarcasm, Hate speech, Sympathy and More in Icelandic Blog Comments. *Proceedings of the Fourth Workshop on Threat, Aggression & Cyberbullying @ LREC-COLING-2024 (Torino, Ítalíu)*, 73–84. ELRA and ICCL.

Helgi Björn Hjartarson and Steinunn Rut Friðriksdóttir. 2024. Malmon: A Crowd-Sourcing Platform for Simple Language. *Proceedings of the 3rd Workshop on Tools and Resources for People with READING Difficulties (READI) @ LREC-COLING 2024 (Torino, Ítalíu)*, 15–21. ELRA and ICCL.

Steinunn Friðriksdóttir and Hafsteinn Einarsson. 2023. The Effect of Data Encoding on Relation Triplet Identification. *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa, Tórshavn, Færeyjar)*, 500–507. University of Tartu Library.

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. (2022). Bootstrapping Icelandic Knowledge Graph Data. *Proceedings of the ESSLLI 2022 Student Session*, 38–49.  
<https://doi.org/10.21942/uva.20367948>

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. (2022). Fictionary-Based Games for

Language Resource Creation. *Proceedings of the 2nd Workshop on Novel Incentives in Data Collection from People: models, implementations, challenges and results within LREC 2022*, 25–31. <https://aclanthology.org/2022.nidcp-1.5>

Steinunn Rut Friðriksdóttir, Valdimar Ágúst Eggertsson, Benedikt Geir Jóhannesson, Hjalti Daníelsson, Hrafn Loftsson, and Hafsteinn Einarsson. (2022). Building an Icelandic Entity Linking Corpus. *Proceedings of The Workshop on Dataset Creation for Lower-Resourced Languages within the 13th Language Resources and Evaluation Conference*, 27–35. <https://aclanthology.org/2022.dclrl-1.4/>

Steinunn Rut Friðriksdóttir, Hjalti Daníelsson, Steinþór Steingrímsson and Einar Freyr Sigurðsson. (2022). IceBATS: An Icelandic Adaptation of the Bigger Analogy Test Set. *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*, 4227-4234. <https://aclanthology.org/2022.lrec-1.449/>

Steinunn Rut Friðriksdóttir, Atli Jasonarson, Steinþór Steingrímsson and Einar Freyr Sigurðsson. (2021). ALEXIA: A Lexicon Acquisition Tool. *Proceedings of the CLARIN Annual Conference 2021*, 64–67.

Steinunn Rut Friðriksdóttir and Anton Karl Ingason. (2020). Disambiguating confusion sets as an aid for dyslexic spelling. *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding Difficulties (READI)*, 1–5. <https://aclanthology.org/2020.readi-1.1>

Steinunn Rut Friðriksdóttir and Anton Karl Ingason. (2020). Disambiguating Confusion Sets in a Language with Rich Morphology. *Proceedings of the 12th International Conference on Agents and Artificial Intelligence - Volume 1: NLPinAI*, 446–451. <https://doi.org/10.5220/0009371504460451>

Steinunn Rut Friðriksdóttir. (2020). *Á næsta leyti að vissu leyti? Notkun ruglingsmengja við sjálfvirka stafsetningarleiðréttingu* [MA thesis]. Skemman. <http://hdl.handle.net/1946/35368>

Steinunn Rut Friðriksdóttir. (2017). Le petit prince: spectacle musical. Traduction du spectacle musical Le petit prince d'après l'œuvre d'Antoine de Saint-Exupéry [BA thesis]. *Skemman*. <http://hdl.handle.net/1946/24347>

Steinunn Rut Friðriksdóttir. (2016). *Baráttan gegn málvillum. Skilar hún árangri?* [BA thesis]. *Skemman*. <http://hdl.handle.net/1946/23596>

## Data and software

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2025. IceBERTBias.  
<https://huggingface.co/steinunnfridriks/IceBERTBias>

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2025. mBERTBiasIce.  
<https://huggingface.co/steinunnfridriks/mBERTBiasIce>

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2025. ScandiBERTBias.  
<https://huggingface.co/steinunnfridriks/ScandiBERTBias>

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2025. IceBiasNER.  
<https://huggingface.co/datasets/steinunnfridriks/IceBiasNER>

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2025. IceBias.  
<https://huggingface.co/datasets/steinunnfridriks/IceBias>

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2025. Inappropriate Word Highlighter.  
<https://github.com/steinunnfridriks/InappropriateWordHighlighter>

Steinunn Rut Friðriksdóttir and Hafsteinn Einarsson. 2025. Named Entity Recognition for Bias Detection in Icelandic (model training code).  
<https://github.com/steinunnfridriks/NERBiasIce>

Steinunn Rut Friðriksdóttir. 2024. The GenderQueer Test Suite.  
<https://github.com/steinunnfridriks/TheGenderQueerTestSuite>

Steinunn Rut Friðriksdóttir, Annika Simonsen, Atli Snær Ásmundsson, Guðrún Lilja Friðjónsdóttir, Anton Karl Ingason, Vésteinn Snæbjarnarson and Hafsteinn Einarsson. 2024. Ice and Fire: Dataset on Sentiment, Emotions, Toxicity, Sarcasm, Hate speech, Sympathy and More in Icelandic Blog Comments.  
[https://huggingface.co/datasets/hafsteinn/ice\\_and\\_fire](https://huggingface.co/datasets/hafsteinn/ice_and_fire)

Steinunn Rut Friðriksdóttir. 2024. CEFR classifier for Icelandic (prototype).  
[https://github.com/steinunnfridriks/CEFR\\_firstlanguage\\_classification](https://github.com/steinunnfridriks/CEFR_firstlanguage_classification)

Hjalti Daníelsson, Steinunn Rut Friðriksdóttir and Steinnþór Steingrímsson. (2021). Icelandic Multi-SimLex. *CLARIN-IS*. <http://hdl.handle.net/20.500.12537/121>.

Steinunn Rut Friðriksdóttir, Valdimar Ágúst Eggertsson, Benedikt Geir Jóhannesson, Hjalti Daníelsson, Hrafn Loftsson, and Hafsteinn Einarsson. (2022). MIM-GOLD-EL - entity linking corpus for Icelandic. *CLARIN-IS*. <http://hdl.handle.net/20.500.12537/168>.

Steinunn Rut Friðriksdóttir, Hjalti Daníelsson and Steinnþór Steingrímsson. (2022). Word Embeddings - FastText optimized for IceBATS 22.04. *CLARIN-IS*.  
<http://hdl.handle.net/20.500.12537/211>.

Steinunn Rut Friðriksdóttir, Hjalti Daníelsson and Steinnþór Steingrímsson. (2021). IceBATS - The Icelandic Bigger Analogy Test Set. *CLARIN-IS*.  
<http://hdl.handle.net/20.500.12537/120>.

Steinunn Rut Friðriksdóttir and Atli Jasonarson. (2021). Alexia: Lexicon Acquisition Tool for Icelandic (Orðtökutól) 2.0. *CLARIN-IS*. <http://hdl.handle.net/20.500.12537/95>.

Steinunn Rut Friðriksdóttir and Atli Jasonarson. (2021). Stopporðalisti fyrir Risamálheildina / Stop-words for the Icelandic Gigaword Corpus (21.08). *CLARIN-IS*.  
<http://hdl.handle.net/20.500.12537/124>.

Steinunn Rut Friðriksdóttir and Anton Karl Ingason. (2020). The Icelandic Confusion Set Corpus (ICoSC) 2.0 (2020-05-06). *CLARIN-IS*. <http://hdl.handle.net/20.500.12537/19>.