

# VÉSTEINN SNÆBJARNARSON

+45-50 300 434 / +354-66 33 727 | [vesteinnsnaebjarnarson@gmail.com](mailto:vesteinnsnaebjarnarson@gmail.com) | [vesteinn.is](http://vesteinn.is) | [github.com/vesteinn](https://github.com/vesteinn)



## EDUCATION

---

### University of Copenhagen

Copenhagen, Denmark

*PhD Fellow in Computer Science*

*Autumn 2022 -*

- Advised by Serge Belongie, co-advised by Ryan Cotterell at ETH Zürich.
- Focus on multimodal settings combining methods for Natural Language Processing and Computer Vision, compositionality of embedding spaces and generative models.
- ELLIS PhD student.

### University of Iceland

Reykjavík, Iceland

*MS in Computer Science - First Class with Distinction*

*October 2021*

- Thesis on Question Answering (QA) for Icelandic. Icelandic and Icelandic-English language models were trained and fine tuned for question answering. Cross lingual open QA system where questions are asked in Icelandic and answers given in English developed and adapted to an Icelandic only system.

### University of Iceland

Reykjavík, Iceland

*BA in Philosophy - First Class*

*June 2012*

### University of Iceland

Reykjavík, Iceland

*BS in Mathematics - First Class*

*October 2011*

### Uppsala University

Uppsala, Sweden

*Chemistry and Mathematics, one semester 45 ECTS*

*Autumn 2008*

### Reykjavík Junior College - Menntaskólinn í Reykjavík

Reykjavík, Iceland

*Matriculation Exam with focus on Physics and Mathematics - Eðlisfræðideild 1*

*May 2008*

## EXPERIENCE

---

### NLP Researcher / CTO

November 2019 –

*Miðeind ehf - Language technology*

*Reykjavík, Iceland*

- Promoted to CTO in autumn 2021, responsible for development and research of language technology solutions for Icelandic.
- Neural machine translation (NMT) systems between Icelandic and English, other languages in development. System in use by official websites of the Icelandic government and in collaboration with the Translation Centre at the Ministry for Foreign Affairs. Translation system available at <https://velthyding.is>.
- Training and analysis of neural language models, monolingual and cross-lingual, that can handle Icelandic.
- Development of various NLP-tools for Icelandic; named entity recognition, part-of-speech tagging, parsing, sentiment analysis, grammar and spelling correction.
- Large scale collection and cleaning of monolingual and parallel data.

### Senior Developer

January 2018 – October 2019

*Trackwell ehf – resource management business*

*Reykjavík, Iceland*

- Full-stack web development for fleet- and human-resource management solutions.
- Modernised systems. Front-ends written in react and containerisation of back-end systems.
- Designed multi tenancy solution for separated clients that had previously been running on multiple servers.

### Developer / PM / CTO

August 2012 – March 2017

*Cooori ehf – Icelandic / Japanese ed-tech company*

*Reykjavík, Tokyo*

- Started as a software developer after graduation, got promoted to product manager and moved to Tokyo to take on role of CTO in 2015.
- Cooori is an ed-tech company teaching languages online using AI based methods and spaced repetition. In 2015 the company pivoted from teaching Japanese to a global audience to teaching English, then adapting solutions directly for large companies in Japan, now focusing on standardised English exams.

### Research Assistant - Bioinformatics

February 2010 – August 2012

*University of Iceland, faculty of Life and Environmental Sciences*

*Reykjavík, Iceland*

- Assembled and analysed genomic, transcriptomic and epigenetic data, mainly in lichens. Got familiar with scientific packages in python, running computations on high performance computing clusters and operating Linux servers.
- Developed methods for handling assembly of metagenomic samples using multiple sequencing types.

## TEACHING

---

<b>Introduction to Language Technology</b>   <i>Sessional Teacher, University of Iceland</i>	Autumn 2021
<ul style="list-style-type: none"><li>• Head teacher in masters-level course in the Computer Science department.</li><li>• Focus on NLP using deep learning with comparison to classical methods when relevant.</li></ul>	
<b>AI lifecycle</b>   <i>Guest lecturer, University of Iceland</i>	Spring 2021
<ul style="list-style-type: none"><li>• Lecture on neural machine translation with a case study.</li></ul>	
<b>Introduction to Logic</b>   <i>TA, University of Iceland</i>	Spring 2012
<b>Mathematical Analysis</b>   <i>TA, University of Iceland</i>	Autumn 2009

## INVITED TALKS AND PRESENTATIONS

---

<b>IceBERT – language models for Icelandic</b>   <i>Icelandic Linguistics Society</i>	November 2021
<ul style="list-style-type: none"><li>• Invited talk on language models and NLP for Icelandic.</li></ul>	
<b>Artificial Intelligence and Question Answering</b>   <i>Icelandic Computer Science Society</i>	November 2021
<ul style="list-style-type: none"><li>• Lecture on AI and QA based on MSc thesis.</li></ul>	
<b>Language and Artificial Intelligence</b>   <i>Reykjavík City library</i>	September 2021
<ul style="list-style-type: none"><li>• An informal conversation for the public with a journalist interviewing. General discussion about AI and language.</li></ul>	
<b>Machine Translation, an Important Option</b>   <i>Reykjavík University</i>	May 2021
<ul style="list-style-type: none"><li>• Recorded talk about machine translation for Icelandic and its applications. Part of a conference about language technology, Máltækniþyltingin (e. <i>The Language Technology Revolution</i>)</li></ul>	
<b>ByteBERT - Masked Language Modeling for Morphologically rich Languages</b>   <i>ALPS, Grenoble</i>	January 2021
<ul style="list-style-type: none"><li>• Poster presentation at the Advanced Language Processing School.</li></ul>	

## PEER-REVIEWED PUBLICATIONS

---

Svanhvít Lilja Ingólfssdóttir, Petur Orri Ragnarsson, Haukur Páll Jónsson, Haukur Barri Simonarson, Vilhjalmur Þorsteinsson and **Vésteinn Snæbjarnarson**. *Byte-Level Grammatical Error Correction Using Synthetic and Curated Corpora*. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023. Toronto, Canada.

**Vésteinn Snæbjarnarson**, Annika Simonsen, Goran Glavaš, Ivan Vulić. Transfer to a Low-Resource Language via Close Relatives: The Case Study on Faroese. The 24th Nordic Conference on Computational Linguistics, 2023. Tórshavn, Faroe Islands.

**Vésteinn Snæbjarnarson** and Hafsteinn Einarsson. 2022. Cross-Lingual QA as a Stepping Stone for Monolingual Open QA in Icelandic. In Proceedings of the Workshop on Multilingual Information Access (MIA) at NAACL 2022. Seattle, WA, USA.

**Vésteinn Snæbjarnarson** and Hafsteinn Einarsson. 2022. Natural Questions in Icelandic. In Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022). Marseille, France.

**Vésteinn Snæbjarnarson\***, Haukur Barri Símonarson\*, Pétur Orri Ragnarsson, Svanhvít Lilja Ingólfssdóttir, Haukur Páll Jónsson, Vilhjálmur Þorsteinsson and Hafsteinn Einarsson. 2022. A Warm Start and a Clean Crawled Corpus – A Recipe for Good Language Models. In Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022). Marseille, France.

Anna Björk Nikulásdóttir, Þórunn Arnardóttir, Jón Guðnason, Þorsteinn Daði Gunnarsson, Anton Karl Ingason, Haukur Páll Jónsson, Hrafn Loftsson, Hulda Óladóttir, Einar Freyr Sigurðsson, Atli Þór Sigurgeirsson, **Vésteinn Snæbjarnarson** and Steinþór Steingrímsson. 2021. *Help Yourself from the Buffet: National Language Technology Infrastructure Initiative on CLARIN-IS*. In Proceedings of the CLARIN Annual Conference.

Haukur Barri Símonarson, **Vésteinn Snæbjarnarson**, Pétur Orri Ragnarsson, Haukur Páll Jónsson and Vilhjálmur Þorsteinsson. 2021. *Miðeind's WMT 2021 Submission*. In Proceedings of the Sixth Conference on Machine Translation (WMT21, EMNLP), Dominican Republic.

Haukur Páll Jónsson\*, Haukur Barri Símonarson\*, **Vésteinn Snæbjarnason\***, Steinþór Steingrímsson\* and Hrafn Loftsson. 2020. *Experimenting with Different Machine Translation Models in Medium-Resource Settings*. In Proceedings of the 23rd International Conference on Text, Speech and Dialogue (TSD 2020). Brno, Czech Republic (Online). Lecture Notes in Computer Science, vol 12284. Springer, Cham.

Vivian P. W. Miao, Sheeba S Manoharan-Basil, **Vésteinn Snæbjarnarson** and Ólafur Andr sson. 2012. *Expression of *lec-1*, a mycobiont gene encoding a galectin-like protein in the lichen *Peltigera membranacea**. Symbiosis 57, 23-31.

## AWARDS AND GRANTS

---

- Icelandic Language Technology Transfer Fund (Mark a tlun   tungu og t kni) | 29.8M ISK** 2022-2023
- For developing a system capable of automated captioning of English videos in Icelandic. Collaboration with the three largest local distributors of TV material, including the national provider.
- Icelandic Language Technology Transfer Fund (Mark a tlun   tungu og t kni) | 29.3M ISK** 2022-2023
- For developing an Icelandic–Polish translation system for initial use in collaboration with the City of Reykjav k.
- The University of Iceland Science and Innovation Motivational Award | 500k ISK** Summer 2021
- For the cross-lingual question answering part of MSc thesis, with focus on the ability of being able to get an answer in English when one can not be found in Icelandic. Awarded annually to a project affiliated with the University in some way.
- Icelandic Student Innovation fund | 2.7M ISK** Spring 2020
- For creating an Icelandic QA corpus - Natural Questions in Icelandic.
  - Five students were hired in summer 2020 to create the corpus.
- Icelandic Student Innovation fund | 900k ISK** Summer 2010
- For developing methods for combining different genome assembly methods in a post-processing step.
- Participant at International Chemistry Olympiads | Moscow and Budapest** Summer 2008 and 2009
- Competed on behalf of Iceland after qualifying in domestic competitions.

## ACADEMIC SERVICE

---

**Thesis examiner:** Automatic Abstractive Summarization for Icelandic Text, 60 ECTS, Þ r Sverisson, University of Iceland, 2023.

**Reviewing:** ACL (2023), COLING (2022), Bioinformatics (Journal).

**Chairing:** Co-chair at *Machine Learning and Statistical Methods* session at LREC 2022.

## SUPERVISION

---

**BS:** Christoffer Ringgaard - Fine Grained Visual Categorization of Sailboats, University of Copenhagen, 2023.