

# Experiment

## Using Finnish knowledge to understand Estonian

### 1 Introduction and background information

Welcome to this experiment, where I am exploring the applicability of Finnish knowledge to understanding texts in Estonian. This experiment is part of a project, where the performance of human native speakers of Finnish will be compared with the performance of a large language model trained on Finnish, on a small-scale task in Estonian. Large language models (LLMs) are abstract representations of languages trained on massive amounts of texts, and have been extensively researched in computational linguistics in the past few years. A hands-on, real-life application of LLMs is the development of chatbots, such as ChatGPT (OpenAI, 2021). ChatGPT is capable of producing human-like text output thanks to a language model built in the tool, called GPT-3 (Brown et al., 2020).

In this project, a similar language model will be evaluated and compared to human speakers. The model is called FinBERT (Virtanen et al., 2019), and it is a language model trained specifically for Finnish. In other words, we could say (with some exaggeration) that it is a native "speaker" of Finnish. With this project, I would like to explore the applicability of Finnish knowledge to understanding Estonian, when it comes to human speakers, as well as the Finnish language model mentioned above. Your contribution is highly appreciated!

### 2 Disclaimer

I would like to emphasize that no personal information (names, e-mail addresses, etc.) will be published or used publicly. Your participation is entirely voluntary, and you have the right to ask for the exclusion of your answers at any given time during the process. In case you wish to withdraw your participation, or have any other questions, please contact me at `kristof.fischer.3430@student.uu.se`. Thank you.

### 3 Instructions

First, you will be provided with short texts in **Finnish**, with a label for each text, indicating the *sentiment* (the attitude of the author towards the concepts/topics discussed) of the texts. The sentiment is either **negative** or **positive**. Please read the examples carefully. You do not need to do anything else other than reading the examples.

Then, you will be provided with some other, short texts in **Estonian** (18 in total). Your task here is to **assign a sentiment label** (negative or positive) to each of the texts, but only one per text. Please try to understand the texts **without help** and note that **it is perfectly fine** if you are unsure, or do not understand parts of the texts. This is **not a language exam** for a language you have never learned 😊. In case you really have no idea what label to choose, please make a guess based on your "instincts".

## 4 Preliminaries

First, please answer a few introductory questions about yourself. Please note that no personal information (name/e-mail/age) will be used publicly in the project.

1. Name:
2. Age:
3. E-mail:
4. Have you lived in Estonia before? (Yes/No)
5. Have you visited Estonia before? (Yes/No)
6. If you answered "yes" on the question above, how many times and for how long did you visit Estonia?
7. On a scale 1–5, approximately how much do you think you have been exposed to the Estonian language?  
*1 = I have never seen or heard Estonian / 5 = I have been exposed to Estonian on a regular basis*  
Please answer with a number between 1–5:

## 5 Tasks

### 5.1 Finnish task

As mentioned before, you will be provided with a few Finnish examples below (6 in total), where each short text is accompanied with a sentiment label (negative or positive). Your only task here is to **read the examples and their corresponding label**. You do not have to memorize anything.

#	Text	Sentiment
1.	Uuden mieheni omaisille suunnattu minun haukkuminen suututti todella.	negative
2.	Niskalapun ompelemiseen kyllä oli panostettu.	positive
3.	Kyllä sulla elämän puitteet on ainakin kunnossa.	positive
4.	16-22-vuotiailla naisilla on nykyään hirveän ruma ääni.	negative
5.	tuo vaan ihmiskunnan kurjan puolen julki!	negative
6.	Onneksi huumoria ei sentään oltu unohdettu ja tervettä itsekritiikkiä ja sarkasmia.	positive

## 5.2 Estonian task

You will now be provided with some short paragraphs in Estonian (18 in total). Your task here is to read the examples, try to understand them, and then fill out the columns called "Sentiment" and "Difficulty". For the "Sentiment" column, please write the label "negative" OR "positive" for each of the examples. In the column "Difficulty", please write a number between 1–5, to show how difficult you thought that specific example was to understand (1=Very easy, 5=Very hard).

#	Text	Sentiment	Difficulty
1.	Et odavamalt läbi ajada, ööbisid bändiliikmed budget-hotellides. "Hotell nagu hotell ikka, aga üks ta natuke räpane kipub olema," räägib muusik India standardist.		
2.	maxima on väga normaalne pood. korralik ja puhas, toidud on head. saage üle juba.		
3.	Iraan ja suurriigid jõudsid ööl vastu pühapäeva tuumaprogrammi osas kokkuleppele, teatas Euroopa Liidu välispoliitika juht Catherine Ashton.		
4.	kui loll peab inimene olema, et seda poliitvangerdust uskuma jääda...		
5.	Mullu uppus Eestis terve aasta jooksul 53 inimest, 2011. aastal 62 inimest ja 2010. aastal 97 inimest.		
6.	Mary õppis koolipõlves klaverit, flööti ja klarnetit, kuid oli ka spordipiiga, kes mängis korvpalli ja hokit. Ta armastas hobuseid ning osales oma hobuse Dianaga ratsavõistlustel.		
7.	USA inimõigusorganisatsioon Freedom House paigutas Eesti internetivabaduste uuringus käsitatud riikide hulgas esimesele positsioonile USA ja Saksamaa ees.		
8.	Morosini suri pärast 14. aprillil toimunud Serie B mängu Livorno ja Pescara vahel. Ta oli kõigest 25-aastane.		
9.	Harjumaal aitas naaber täna öösel põlema läinud korterist välja ema ja kaks last		
10.	Euroopa Komisjon on lõpuks ometi hakanud probleemi lahendama, tulles välja uue initsiatiiviga, kus pearõhk on pandud lennunduslepingute sõlmimisele ja muutmisele kolmandate riikidega, ka omandisuhete reeglistiku ülevaatamine.		
11.	Neljapäeval rääkis Rahvusvahelise Valuutafondi (IMF) juht Christine Lagarde, et liiga palju kasinust liiga kiiresti võib kaasa tuua lisaprobleeme.		
12.	Ajakirjanikuna ei saa ma politseile soovitada peksasaamise mahavaikimist, aga maksumaksja ja kodanikuna on tunne, et vähemalt piinlikud eneseõigustused oleksid võinud ära jääda.		
13.	Sa tunnend ennast nähtamatuna. Sinu e-kirjadele vastata, sinu projektidele ei tule tagasisidet ega tunnustust. . . . Need on üsna tõenäolised ohumärgid vallandamisest.		
14.	Gerd Kanteri seni ainsast maailmameistriitlist on möödas kuus aastat ja vanameister soovib veel kord kogeda toonast tunnet.		
15.	Täna Peterburis Ermitaaži Teatris toimunud X Balti Tähe auhindade üleandmise tseremoonial anti Balti Tähe auhind Rahvusooper Estonia peadirektorile Aivar Mäele.		
16.	Lõpuks pean aga tunnistama, et pea suvaline kontsert on natuke parem kui plaadikuulamine. Kas see peab lõpuni paika ka Nokias kitsas ja hapnikuvaeses saalis, pole juba Alice Cooperi probleem.		
17.	Nii jube oli siinset juttu lugeda.. Kuidagi väga kahtlane see kõik.		
18.	Tubli, need imelised vaated, mis tipust avanevad, tasuvadki kogu ettevõtmise.		

Kiitos paljon avusta! ☺

## References

- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.
- OpenAI. ChatGPT (Version 3.5). OpenAI Website, 2021. URL <https://chat.openai.com/>.
- Antti Virtanen, Jenna Kanerva, Rami Ilo, Jouni Luoma, Juhani Luotolahti, Tapio Salakoski, Filip Ginter, and Sampo Pyysalo. Multilingual is not enough: BERT for Finnish. *arXiv preprint arXiv:1912.07076*, 2019.