

Investigating incidental vocabulary acquisition capabilities of pretrained language models

- **Supervisor:** Lütfi Kerem Şenel
- **Examiner:** Prof. Schuetze
- **BSc, MSc, Open:** M.Sc.
- **General Topic Area:** Language Model Analysis, lexical semantics.
- **Prerequisites:** Moderate experience in Python. Familiarity with the Huggingface library is a plus.
- **Details:** Vocabulary acquisition is the process of learning new words. It is widely accepted that much second language vocabulary learning occurs incidentally while the learner is engaged in extensive reading. It is shown that most vocabulary knowledge could be acquired after two encounters in a reading task. Although the current state of the art models show impressive language understanding and vocabulary knowledge, their understanding of word meaning, contrary to humans, greatly depends on a large number of observations of the target word in various contexts.

The aim of this study is to investigate incidental vocabulary acquisition capabilities of various pretrained language models. In the language learning literature, several studies have been performed on humans to investigate human vocabulary acquisition capabilities. Most of these studies can be adapted and applied to language models (i.e., make a language model forget a word and test how well it can relearn it based on some given contexts). A model can try to learn the meaning of a new word through i) finetuning, ii) prompting (in context learning) using the examples that contain the target word.
- **Project Takeaways:** Hands-on experience with state-of-the-art pretrained Language Models; gaining insights about the strengths and weaknesses of these models.