Do you sprichst français?

Text-to-speech (TTS) technologies have become an integral part of our daily lives, powering applications like GPS systems and voice assistants. However, a significant limitation is their difficulty in handling multilingual sentences, resulting in inaccurate pronunciation. This thesis aims to explore potential ways to address the challenge of word-level language detection which would allow to build TTS systems capable of handling these situations.

Initial situation

Over the past few decades, Text-to-Speech (TTS) technologies have significantly evolved, becoming increasingly realistic and finding applications in a wide range of contexts. However, a significant limitation of these technologies is their struggle with pronouncing multilingual sentences. Since Text-to-Speech for a single language at a time is a solved task, one potential solution to address this problem is the implementation of word-level language detection. Unfortunately, this area remains understudied, with existing solutions proving effective only at the sentence level. This refinement could significantly improve user experience, particularly in applications such as GPS navigation, where correct pronunciation is crucial.

Concept

No satisfactory solution currently exists for the problem at hand. Instead of aiming for a single optimal solution for all use cases, this thesis explores different approaches and evaluates their merits and drawbacks. The implemented approaches include a naïve rule-based algorithm using word lookup in dictionaries, a method based on Named Entity Recognition (NER) and language detection, a statistical approach using word and sentence language predictions, and a machine learning approach based on word vectorization using transformers. These approaches were evaluated on nine categories of cross-lingual texts, providing an accurate analysis of their strengths and weaknesses.

Results

An approach using classical named entity recognition was not conclusive at all, even for a navigation use case, due to current NLP technologies being designed for monolingual texts only and performing less well than the rule-based approach, which is logically not the preferred method due to its inability to capture context. However, the approach based on vectorization via transformers performed best in eight out of nine categories and for each of the languages tested. It achieved an average accuracy per sentence of nearly 93% over the test set, despite an insufficient volume of training data, suggesting that using transformers may be a promising way to solve this NLP problem that millions of people encounter every day.

Diego Fabrice Fraile Felipe
079 332 49 35
diegofraile@bluewin.ch

Hugo Perotto
078 825 75 77
hugoperotto@gmail.com
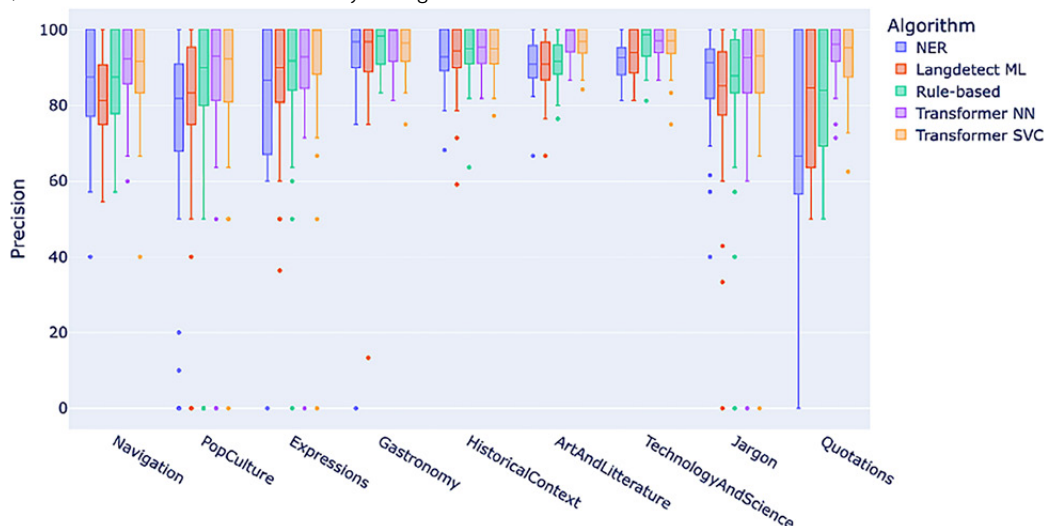
Approaches comparison over categories