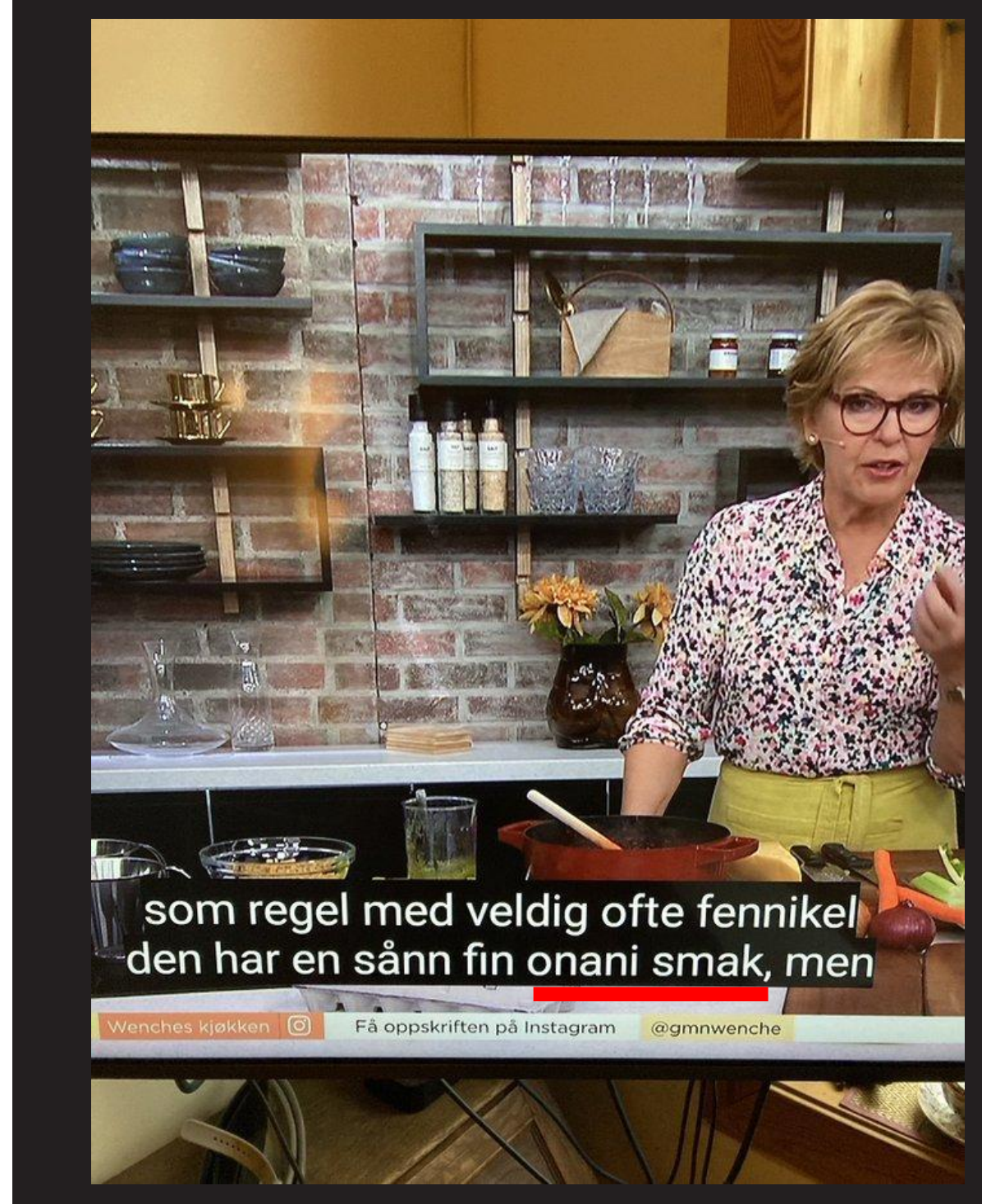
Automated Identification of Severe Errors in Speech to Text Transcripts

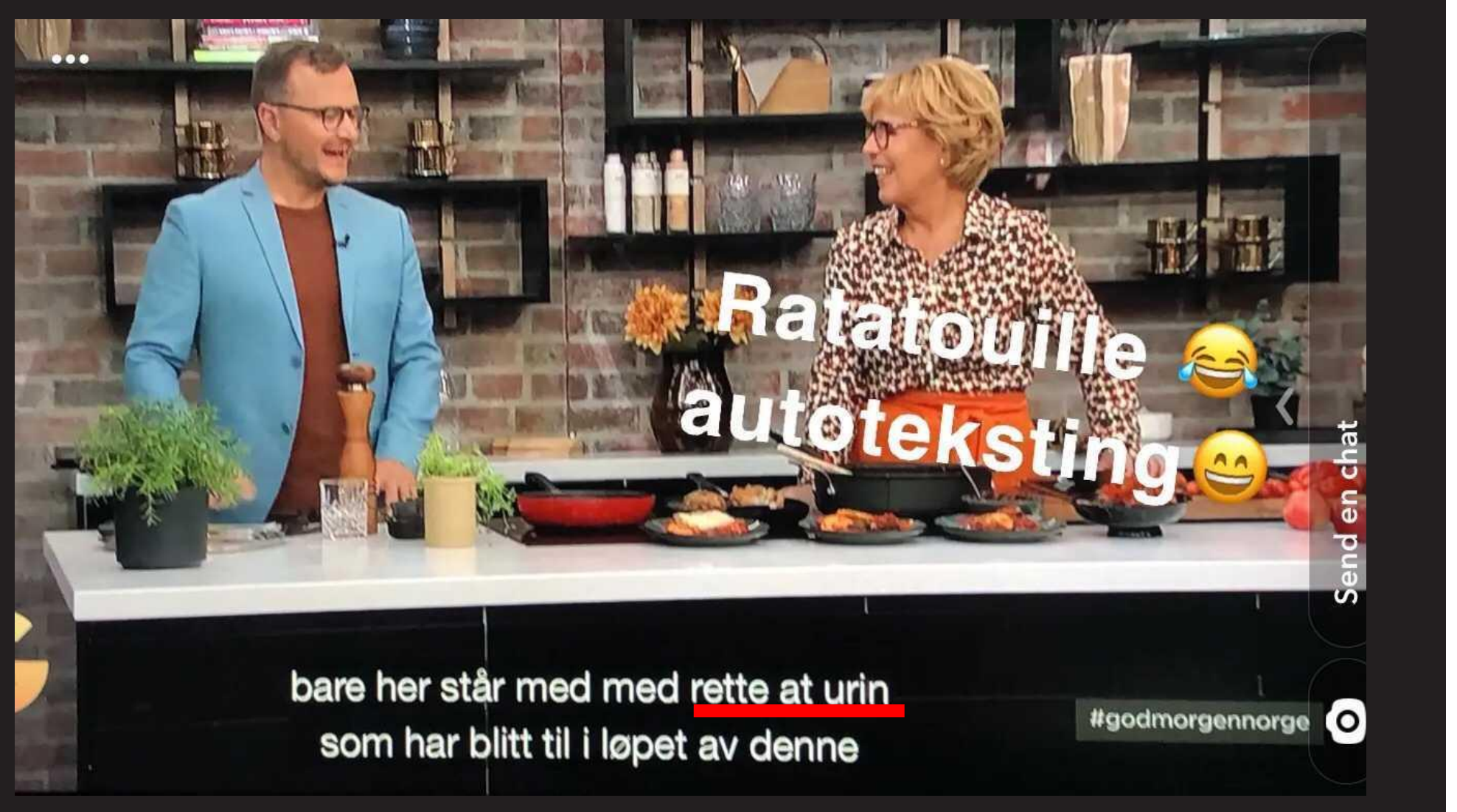
Media Futures

Frederik Hjelde Rosenvinge

Supervisors: Samia Touileb & Lubos Steskal


som regel med veldig ofte fennikel
den har en sånn fin onani smak, men
Wenches kjøkken
Få oppskriften på Instagram
@gmnwenche

A single word error can be catastrophic


Ratatouille 😂
autoteksting 😄
bare her står med med rette at urin
som har blitt til i løpet av denne
#godmorgennorge
Send en chat

A sequence of wrong words is a unique challenge

Abstract

Automatic Speech Recognition systems (**ASR**) are used to automatically generate text from speech. ASR can be done quickly, but often at the cost of precision, meaning that the speaker's intended word was wrongly transcribed. The purpose of this thesis is to firstly identify such errors, secondly to correct them. We will be working with Norwegian texts, with data from **TV2**. The plan is to use pre-trained language models for Norwegian based on the BERT architecture.

Research question

1. RQ 1: To what extent can we develop a system, using pre-trained language models, that can **find** severe errors in speech-to-text transcripts given their context?
2. RQ 2: To what extent can we develop a system that can **correct** severe errors in speech-to-text after being identified as errors?
3. RQ 3: How **fast** can we make the system, and will it be fast enough to be used online?

Data

- Provided by TV2.
- Transcripts from shows like God Morgen Norge, News, and political debates.

Post-processing of the ASR system's output text involves both **finding** and **correcting** errors, meaning finding misplaced words and replacing them with the correct ones. Looking at the **context** surrounding a word is an approach that is feasible.

Challenges

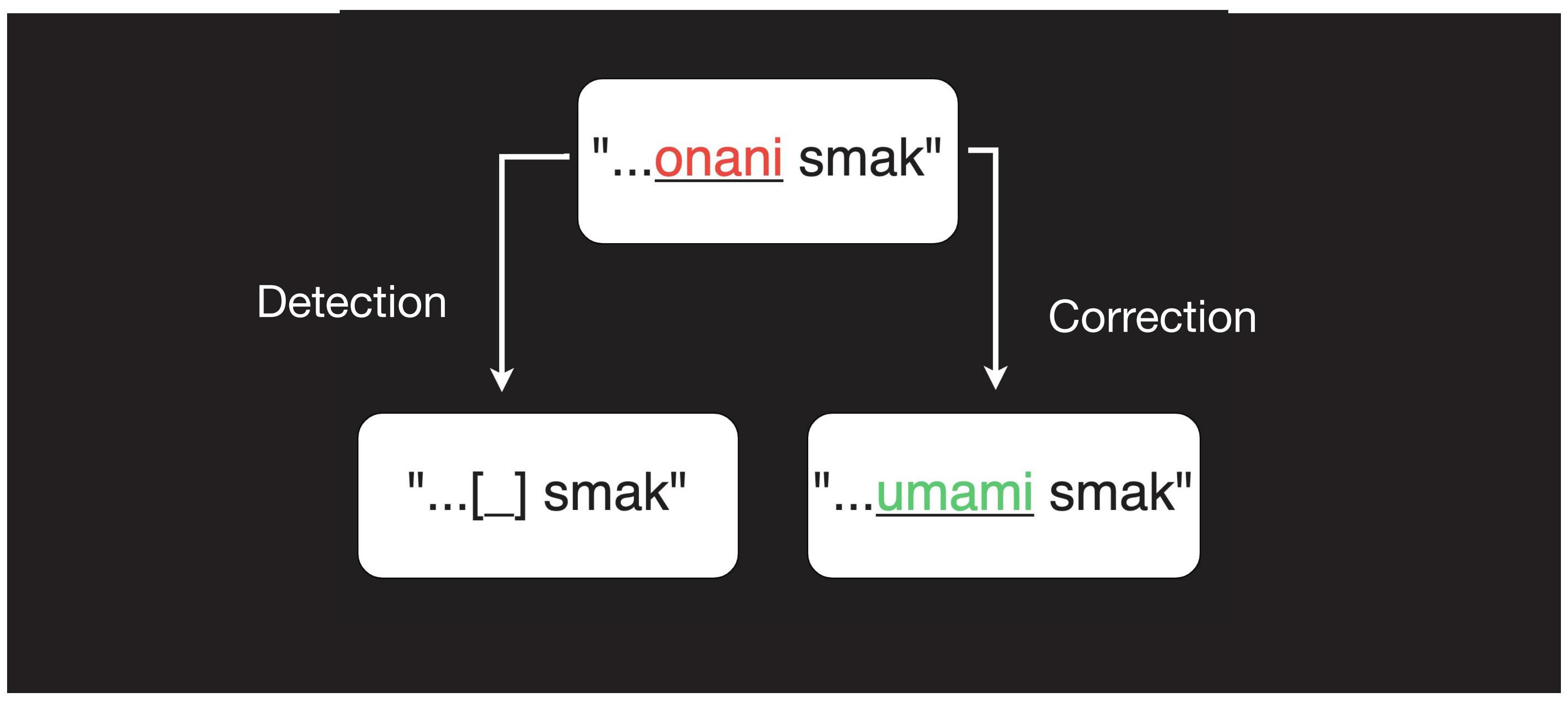
- Defining a threshold for when a word is classified as catastrophic.
- Finding these catastrophic errors.
- Correcting errors.

Types of errors

- Single word errors.
- Sequences of wrongly transcribed words (See Ratatouille picture) pose a unique challenge.
- Non-words.

Methods

The plan is to use pre-trained language models for Norwegian, such as NorBERT, to see how likely or surprising a word is given its context. BERT-based language models are trained on large amounts of text, and can learn how a word fits into a given context. If a word is surprising enough, i.e it doesn't fit the context, and is *also* classified as catastrophic, then it needs to be removed or corrected.



PARTNERS

HOST

UNIVERSITY OF BERGEN

FUNDER

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).