

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Thesis author** Bc. Václav Ryšlink

**Thesis title** Methods of Input Segmentation for Simultaneous Speech Translation

**Submitted** 2022

**Program** Computer Science **Specialization** Artificial Intelligence

**Review author** Peter Polák **Role** reviewer

**Position** Institute of Formal and Applied Linguistics, Charles University

## Review text:

The thesis is concerned with the topic of segmentation methods for simultaneous machine translation (SimulMT). SimulMT provides translations of unfinished sentences, maximizing the quality and minimizing the delay (latency) of the output. In SimulMT, segmentation methods split the input sentence into chunks whose translation should be independent of any forthcoming context, i.e., they can be translated without further revisions after new words arrive. At the same time, one would like the segments to be as short as possible to update the translation more frequently (i.e., to minimize the latency).

The thesis first collects segmentations of the ESIC dataset provided by human interpreters. The interpreters are given full sentences (contrary to the real-world setting where the words appear continuously) and were given infinite time to make the splits.

The thesis introduces 6 splitting methods: based on punctuation, silence, word alignment, phrase chunking, POS tags, and based on MT and two controls (after every word and sentence). In the second part of the chapter, the author categorizes the methods based on whether they can or cannot be directly applied in the online setting (i.e., with growing input). For the methods that are not directly applicable in the online setting (word alignment, MT, and golden data), the author proposes to use a binary classifier trained on the data produced by the splitting data.

The next chapter analyses the collected data and methods. Interesting is the observation that nearly 75% of the splits made by the interpreters were on the phrase boundaries, which radically limits the places where splits can occur. The analysis then continues with the comparison of annotators, and finally with an analysis of splitting methods. The author does first a pairwise comparison of common splits and proposes two new terms "granular and true unique splits" that help to find which splitting method is a subset of another method. Finally, the author evaluates the proposed methods in terms of quality and flicker when combined with the baseline and finetuned MT model. An important observation is that all methods with a look-ahead of one word are

Pareto optimal.

The thesis is divided into 6 chapters (plus an introduction and conclusion) and consists of 74 pages of the main text. The thesis is written in perfect English, the text is exceptionally clear and fluent. The author explains and argues every decision. The main text is accompanied by several graphs and figures. The amount of experiments is well above the expectation for a diploma thesis. The thesis contributes interesting observations, a dataset, and a visualization tool that can be used in future research.

Two small remarks on the form:

- Text figures (e.g., 5.9 or 5.14) are included as raster graphics rather than vector/LaTeX table.
- Moving the technical details (e.g., 3.3 Technical details of the annotation) to the Appendix.

Additional questions:

- Section 4.3.2: How many pairs/what's the ratio of words that have silence between them and are not at the sentence boundary?
- Would be more realistic to show the interpreters the words sequentially rather than the whole sentence? I.e., in the current setup, the interpreter "sees" the future and hence does decisions differently than in the real-world scenario when they translate without seeing the future context.

**I recommend the thesis for defense.**

**I suggest to not consider the thesis for the annual award.**

September 30, 2022

Signature: