**Models:**

Intros: <https://huggingface.co/jesperjmb/parlaBERT/tree/main>

Compounded intros: <https://huggingface.co/jesperjmb/CompundedIntros/tree/main>

Split intros: <https://huggingface.co/jesperjmb/MergeIntrosNSP>

Base training data: <https://github.com/welfare-state-analytics/riksdagen-annotations/tree/csv>

**Training\_iteration\_x**: Training after employing active learning.

**nsp\_train\_x**: Training data for the split intros: Randomly sampled intros and the subsequent text block and their respective independent classification “tag” as well as an annotation on if they have been split or not.

**compounded\_intros\_2**: Training data for if an intro has been compounded with the speech. “compounded\_intros\_1” is not annotated but is included to showcase which rows were included as a part of the active learning process.

**regexvsbert**: Comparing regex and BERT differences

**evaldata**: Test data (should be manually annotated)

**df\_compound\_eval**: Predictions after using active learning on the data. Text 1 is the relevant column. It is solely the predictions and not annotated in the file itself.

**df\_splits\_eval**: Predictions for split intros except final iteration. It is solely the predictions and not annotated in the file itself.