CODECHECK certificate 2025-002 for Evaluating Subtitle Segmentation for End-to-end Generation systems

Item	Value
Title	Evaluating Subtitle Segmentation for End-to-end Generation systems
Authors	Alina Karakanta, François Buet, Mauro Cettolo, François Yvon
Ref. paper	https://aclanthology.org/2022.lrec-1.328/
Codecheckers	Alex Brandsen (0000-0003-1623-1340), Matthew Sung, Matthijs Westera (0000-0001-7777-1864)
Date of Check	2025-02-14
Summary	Full reproduction of code which evaluates subtitle segmentation
Repository	https://github.com/fyvo/EvalSubtitle
Ref. certificate	10.5281/zenodo.15173758

Table 1: CODECHECK summary

Output	Comment
results_exp1.csv	Output of experiment 1
results_exp2.csv	Output of experiment 2
output/projected.fr	Output of experiment 3

Table 2: Summary of output files generated

Summary

Overall, this is documented well and runs well, except for the third example in the lrec folder README, there the external library mwerSegmenter causes issues; on Windows it doesn't install at all, but it works on Linux systems.

The README steps produces 1 image (see below), so doesn't fully reproduce the 4 images in the paper.

Installation prerequisites and computational environment

- Fresh environment in anaconda, python 3.9.18
- pip install -r EvalSubtitle/lrec/requirements.txt

Data preparation

None

Running the code

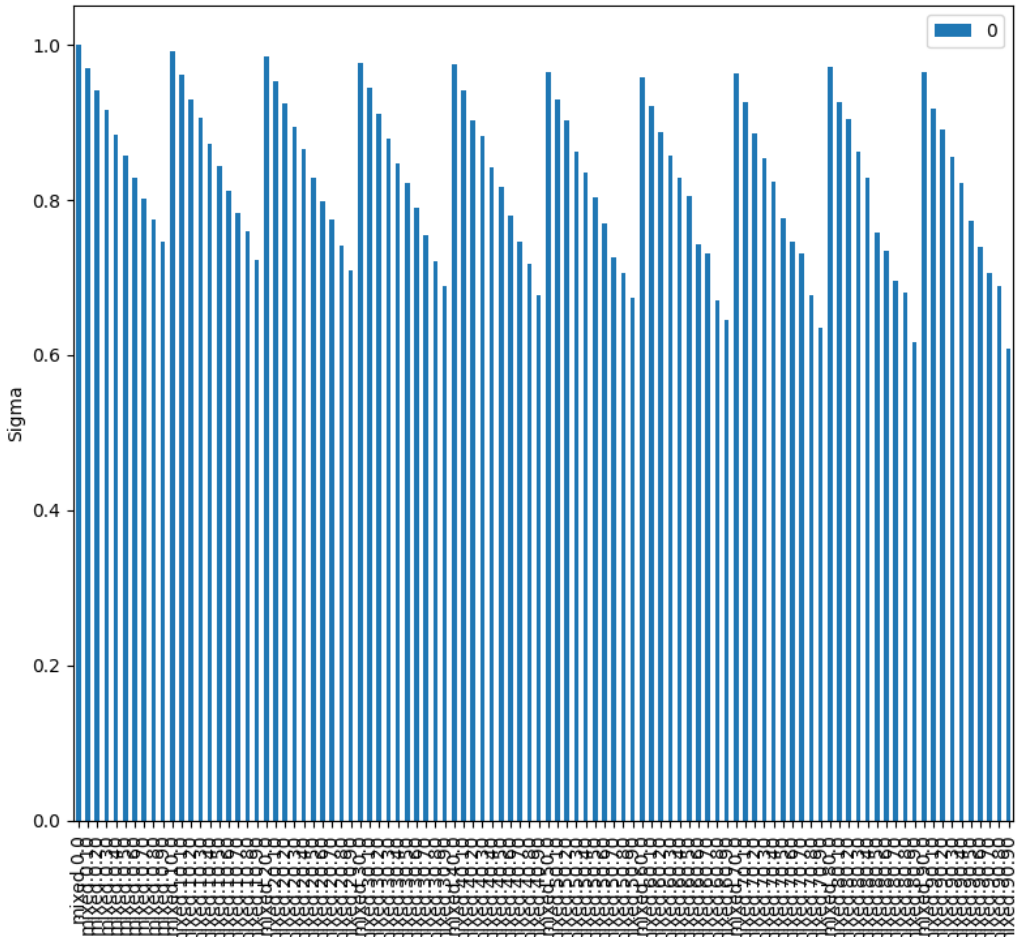
Ran the following commands, as documented in lrec/README.md

1. python degrade_and_eval.py --output_dir . --results_file results_exp1.csv
2. python bleu-br_upper_bound.py --output_dir . --results_file results_exp2.csv
3. bash bound_proj.sh ../data/nmt.fr ../data/amara.fr nmt fr

Outputs

The scripts output a lot of temporary files, which are incorporated into the main result files. This includes many .txt files called 'amara.mixed.0.0.txt', with the numbers in range of 0-90, and folders called add, delete, replace and shift, also containing many text files.

The first script outputs the file results_exp1.csv, this contains metrics for different systems and modes. It also opens the following matplotlib plot, which I think corresponds to fig. 4 in the paper.


0
1.0
0.8
0.6
0.4
0.2
0.0
Sigma

The second script outputs the file results_exp2.csv, this contains metrics for different p_txt and p_tags combinations.

The third script generates output/project.fr, a text file containing French text <eob> and <eol> tags

Acknowledgements

This codecheck is an output of the CodeCHECK-NL workshop held in Leiden, on 14-02-2025

Citing this document

Brandsen, A., Sung, M. & Westera, M. (2025). CODECHECK certificate 2025-02 for Evaluating Subtitle Segmentation for End-to-end Generation systems. Zenodo. https://doi.org/10.5281/zenodo.15173758