A3. Coursework on speech translation (3.75 points)

The goal of this coursework is to carry out different experiments of speech translation on the Covost2 database. To this purpose a series of Jupyter Notebooks as examples have been provided on the speech translation task Europarl-ST.

Recommended steps

1. Select one dataset for a speech translation task (source-english) from Covost2 setup to be used in the coursework and inform the lecturers on your choice.

2. Run the following experiments:

- Based on L4.1_ASR_Whisper_Baseline jupyter notebooks, run a similar experiment using your assigned source language. Report WER figures using different whisper models
- Based on L4.2_ST_Whisper_Baseline_Europarl-ST jupyter notebooks, run a similar experiment using your assigned (source-english) Covost2 setup. Report BLEU and COMET figures using different whisper models
- Based on L4.3_ST_Cascade_Whisper_NLLB_Baseline_Europarl-ST jupyter notebooks, run a similar experiment using your assigned (source-english) Covost2 setup. Report BLEU and COMET figures using the best whisper model obtained in the first experiment
- Based on L4.4_ST_Cascade_Whisper_NLLB_Finetuning_Europarl-ST jupyter notebooks, run a similar experiment using your assigned (source-english) Covost2 setup. Report BLEU and COMET figures using the best whisper model obtained in the first experiment. Evaluate improvements on BLEU and COMET with respect the previous baseline speech translation experiment

4. Work on extensions of the previous notebooks

- Run cascade experiments on one additional alternative translation model

- Explore different fine-tuning strategies: different sizes of fine tuning data, different kind of sentences (output ASR transcriptions, transcription references, not normalize/normalize transcriptions and translations, etc.)
- Explore training and inference parameters

Submission (deadline: 27/01/2026)

1. This coursework is intended to be carried out individually

2. A report of maximum 4 pages must be submitted including:

   - Brief description of the selected dataset and limitations adopted to run the experiments on your dataset according to your computing capabilities (maximum number of tokens, number of samples in training, validation and test, etc.)
   - BLEU and COMET scores for provided notebooks
   - BLEU and COMET scores on your own extensions
   - Conclusions

3. Notebooks:

   - Those four to run baseline experiments
   - Those you need to run your own extensions