Motivation for End-to-End SLT

Benefits:
- Uncertainty directly handled.
  - Target-language considerations influence speech recognition.
- Potentially fewer NN parameters.

Drawbacks:
- Insufficient training data.
  - Speech + transcript and parallel texts much more common than speech + translation.
- 20–40x longer input sequences (sound timeframes vs. subwords).
- Difficult alignment problem within sentences/utterances.
- Non-golden utterance segmentation not yet much considered.