

# Speech Translation Evaluation Report

## Summary, Conclusions & Next Steps:

### Data Analysis Key Findings:

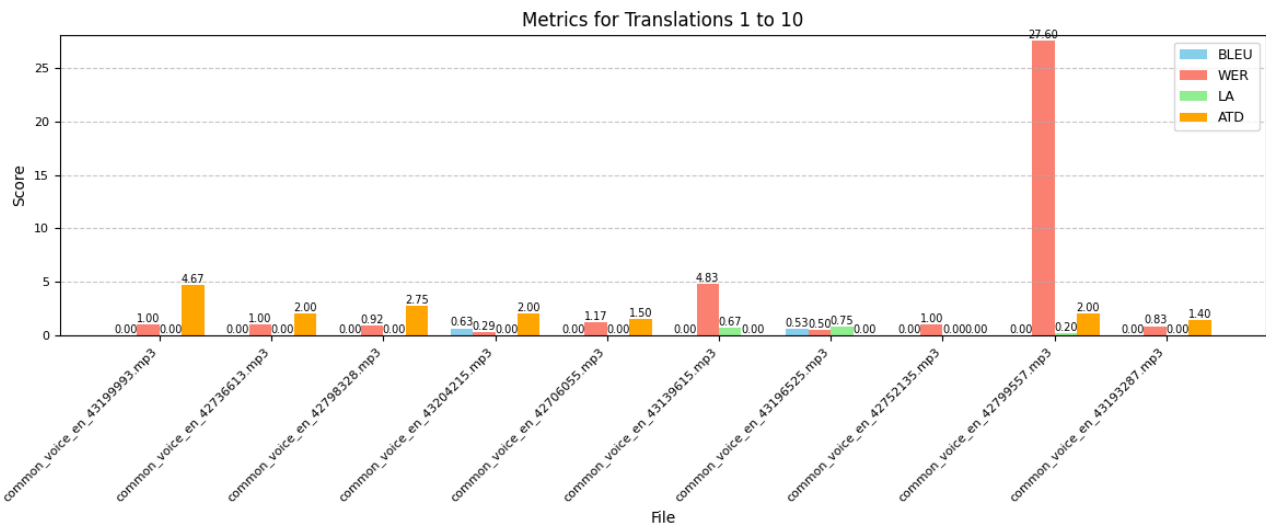
- The evaluation was performed on 170 translations.
- The average Sentence BLEU score is 0.1351.
- The average WER is 2.4794.
- The average Local Agreement (LA) is 0.1837.
- The average Average Token Delay (ATD) is 1.5504.
- Note: Overall Corpus BLEU requires all reference and predicted texts for calculation.

### Insights & Next Steps:

- Analyze batches with low BLEU/LA and high WER/ATD to identify potential issues with specific audio files or translation challenges.
- Investigate the specific sentences with poor metrics to understand the types of errors occurring.
- Consider using a different translation model or fine-tuning the current model if performance is unsatisfactory.
- Evaluate the impact of audio quality on translation performance.

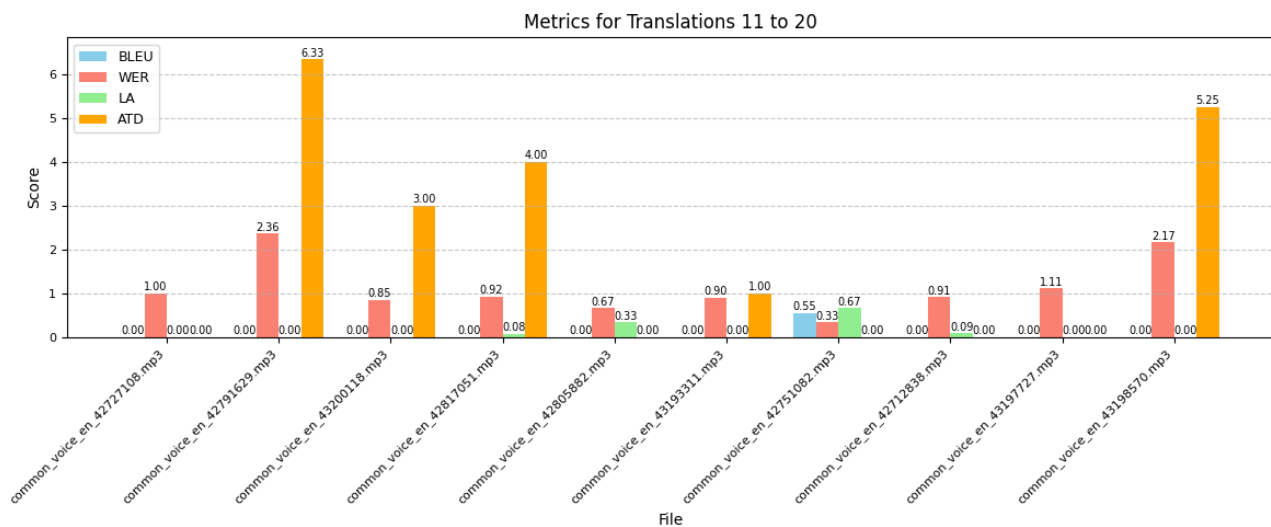
## Batch Metrics Plots:

### Batch 1 Metrics:



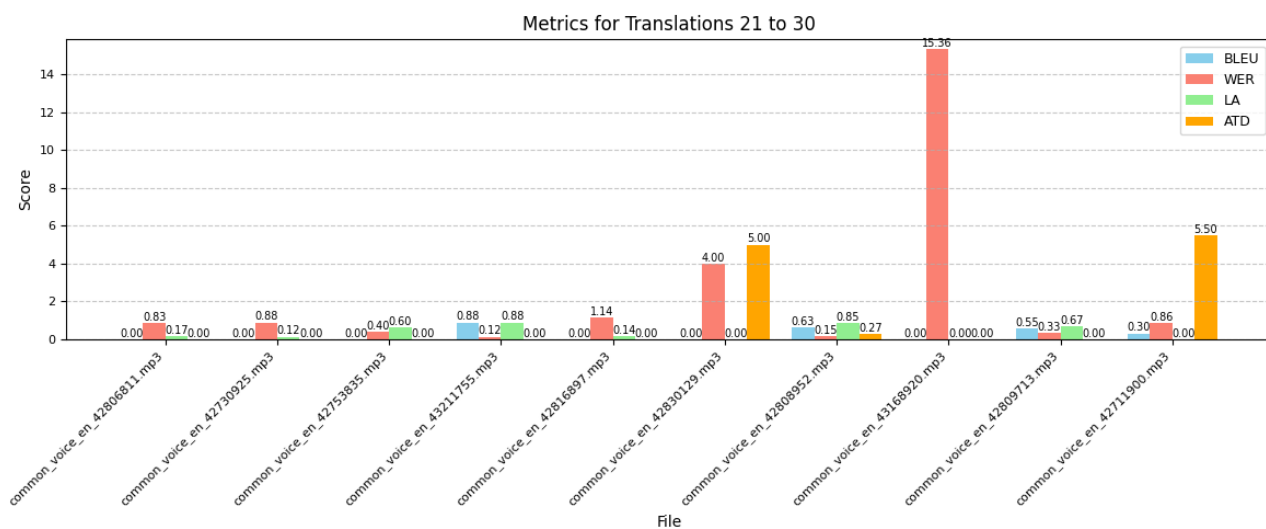
Batch 1 Avg: BLEU: 0.1164, WER: 3.9136, LA: 0.1617, ATD: 1.6317

### Batch 2 Metrics:



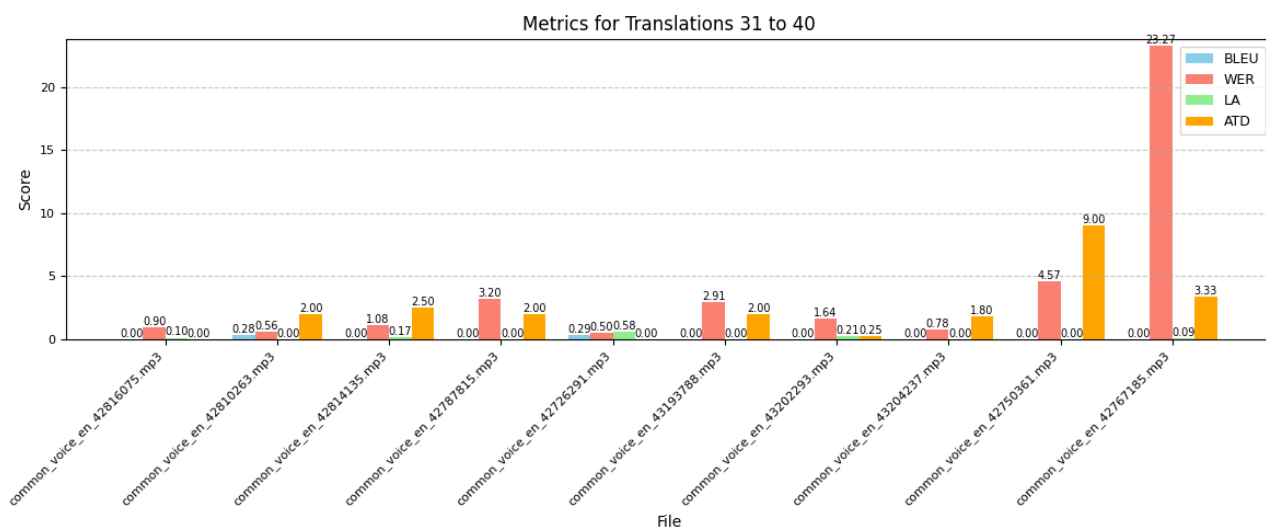
Batch 2 Avg: BLEU: 0.0546, WER: 1.1220, LA: 0.1168, ATD: 1.9583

### Batch 3 Metrics:



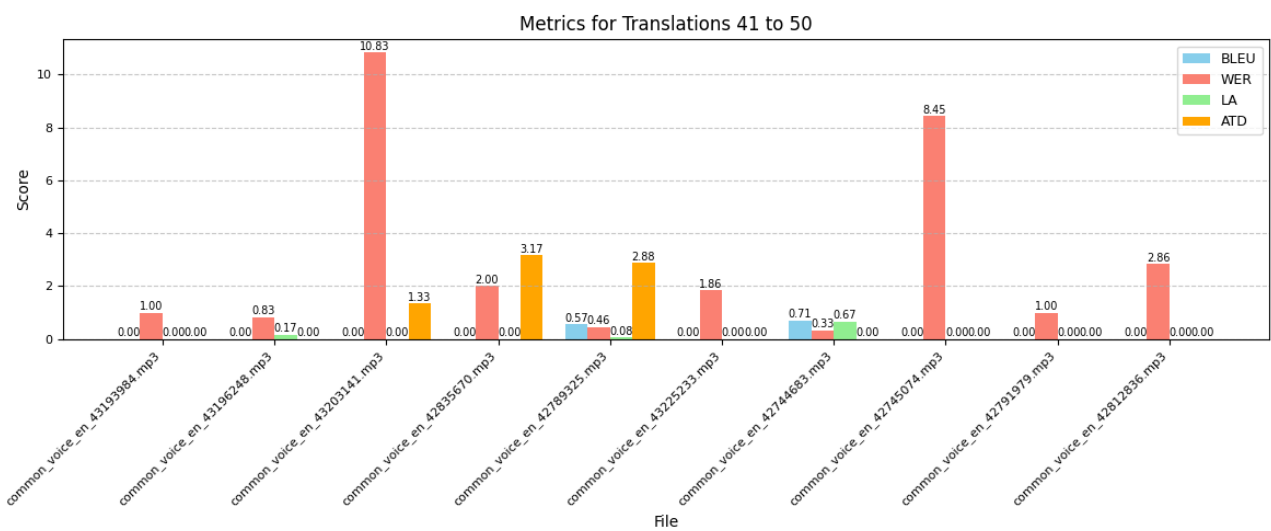
Batch 3 Avg: BLEU: 0.2357, WER: 2.4084, LA: 0.3422, ATD: 1.0773

### Batch 4 Metrics:



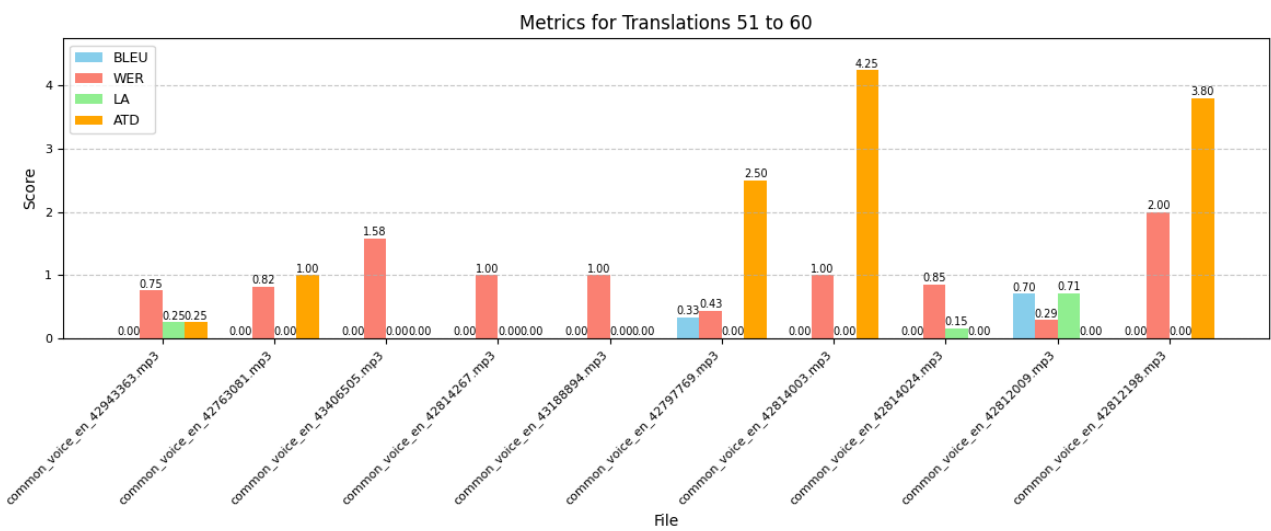
Batch 4 Avg: BLEU: 0.0573, WER: 3.9413, LA: 0.1155, ATD: 2.2883

Batch 5 Metrics:



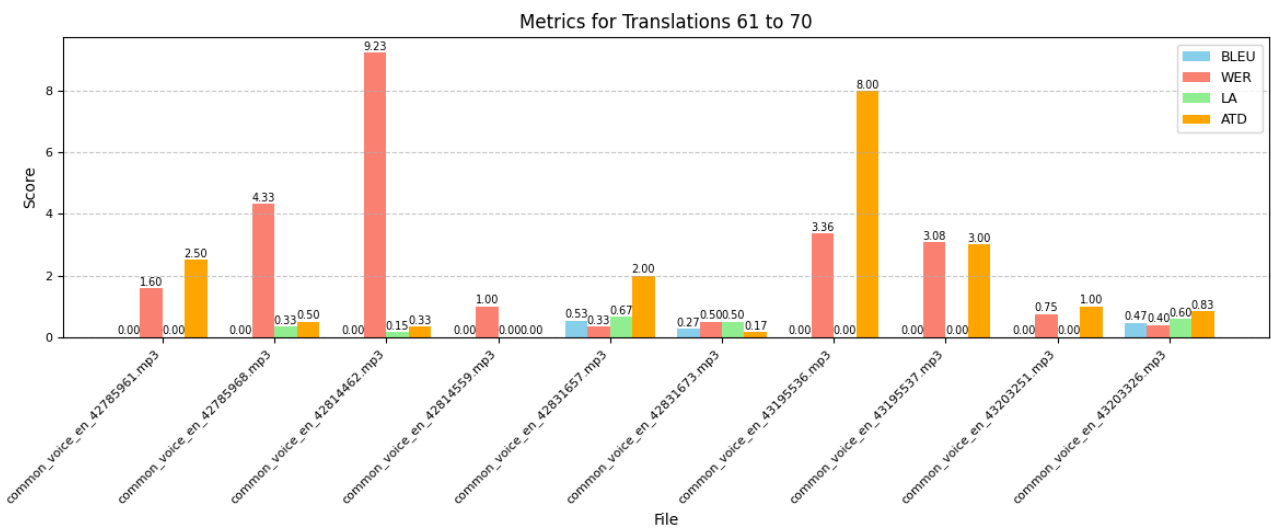
Batch 5 Avg: BLEU: 0.1274, WER: 2.9630, LA: 0.0910, ATD: 0.7375

Batch 6 Metrics:



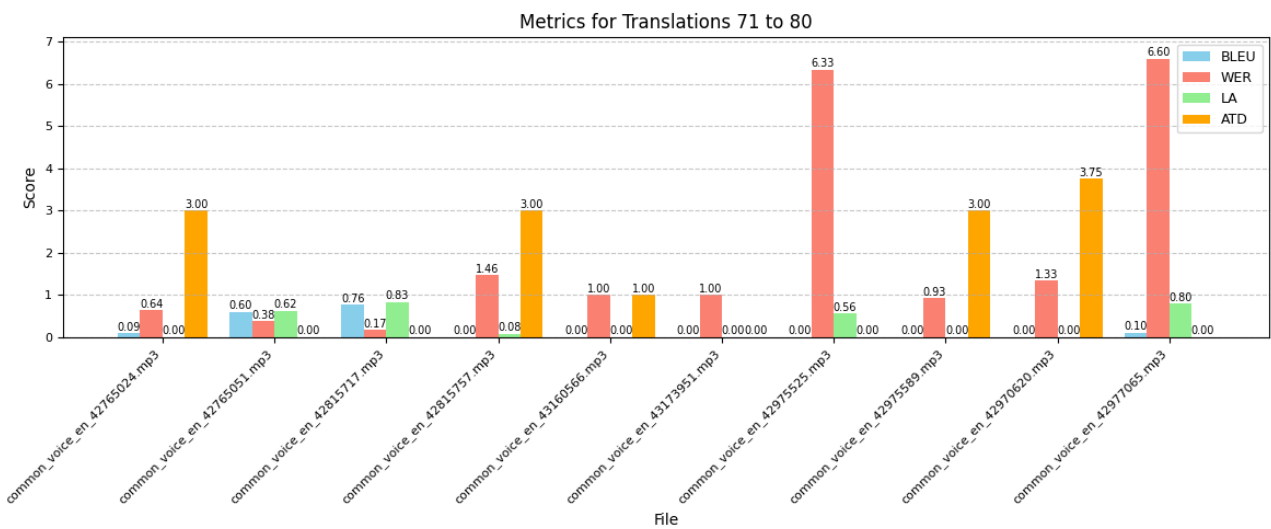
Batch 6 Avg: BLEU: 0.1027, WER: 0.9712, LA: 0.1118, ATD: 1.1800

Batch 7 Metrics:



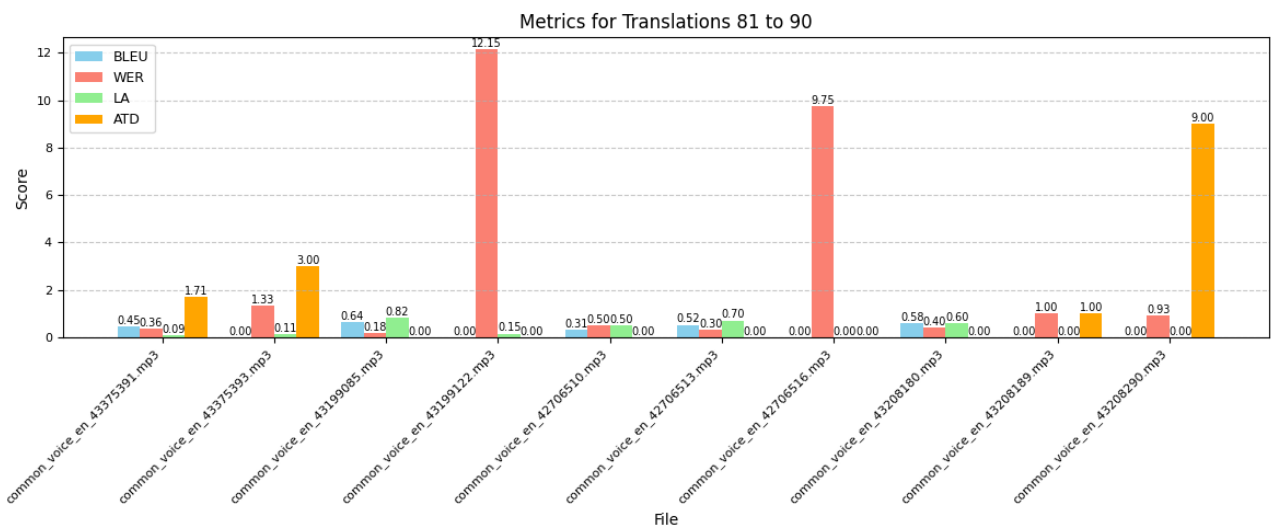
Batch 7 Avg: BLEU: 0.1272, WER: 2.4582, LA: 0.2254, ATD: 1.8333

Batch 8 Metrics:



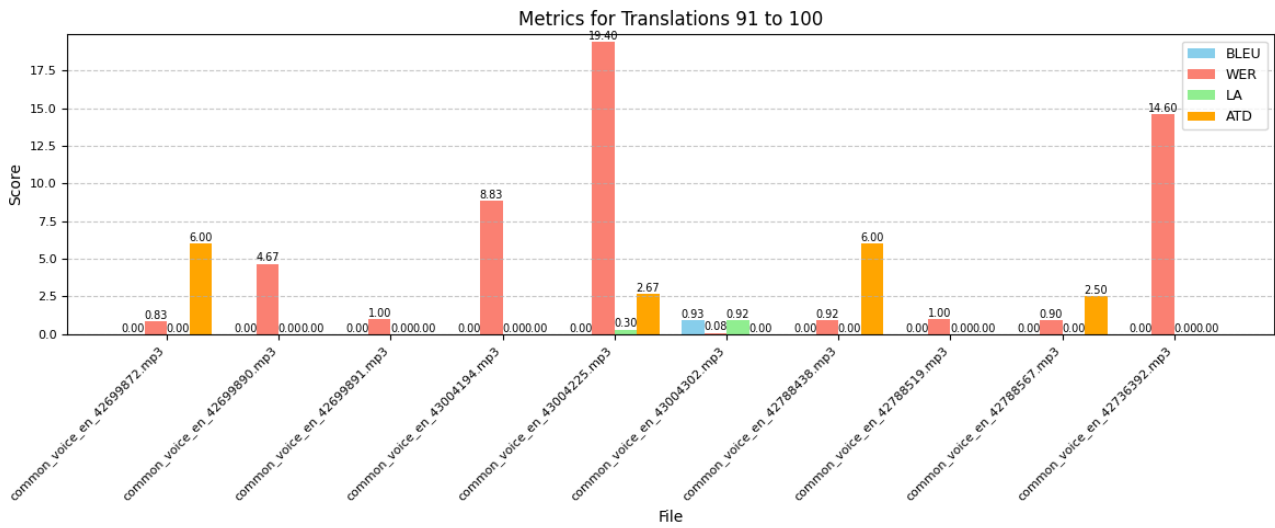
Batch 8 Avg: BLEU: 0.1550, WER: 1.9841, LA: 0.2891, ATD: 1.3750

Batch 9 Metrics:



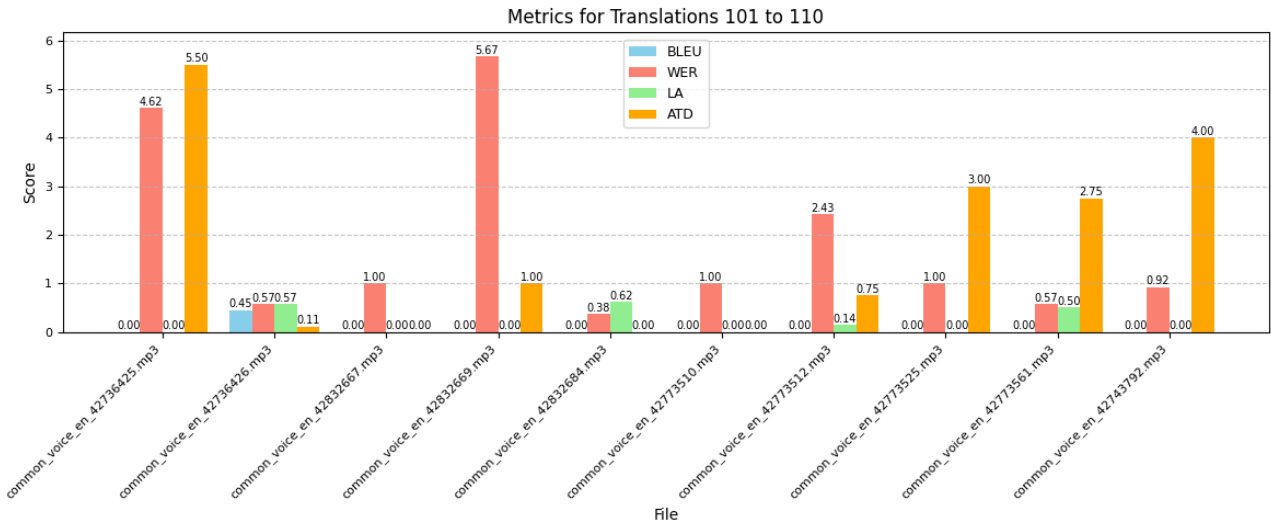
Batch 9 Avg: BLEU: 0.2498, WER: 2.6911, LA: 0.2974, ATD: 1.4714

Batch 10 Metrics:



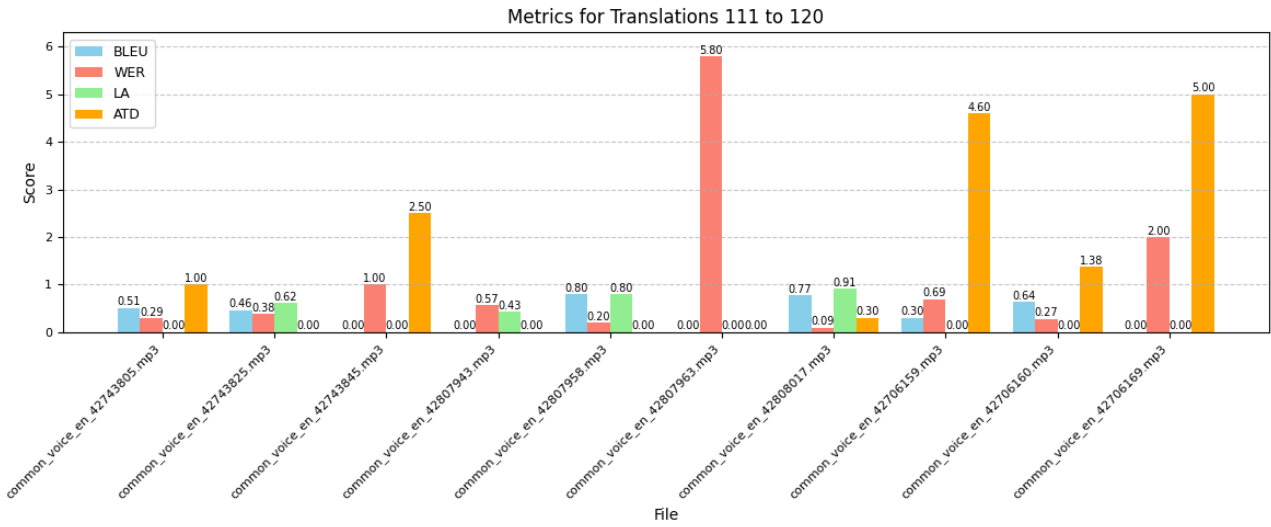
Batch 10 Avg: BLEU: 0.0926, WER: 5.2233, LA: 0.1223, ATD: 1.7167

### Batch 11 Metrics:



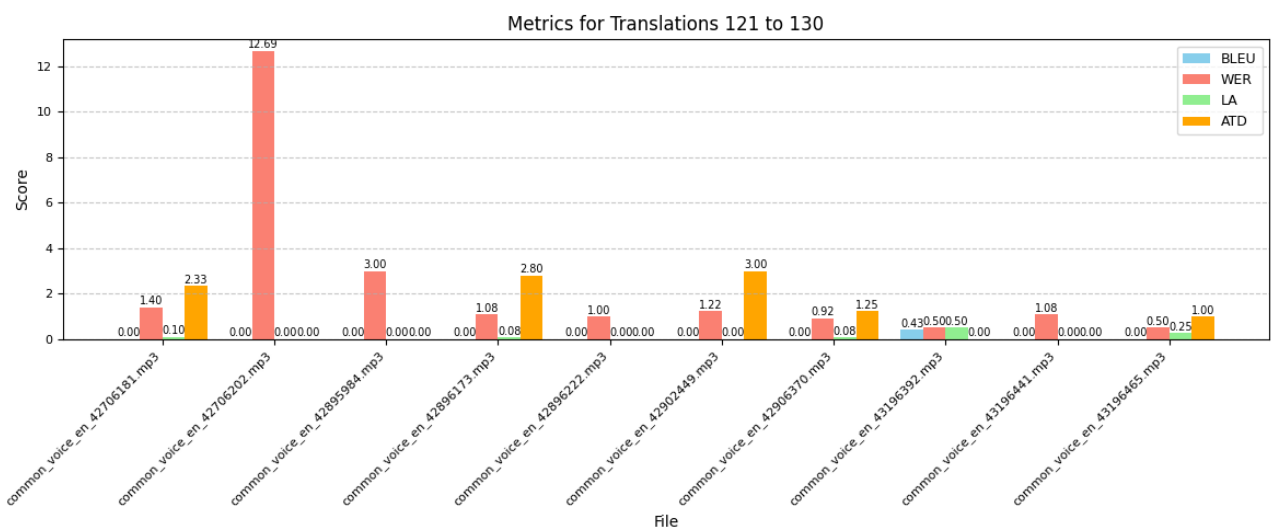
Batch 11 Avg: BLEU: 0.0446, WER: 1.8152, LA: 0.1839, ATD: 1.7111

### Batch 12 Metrics:



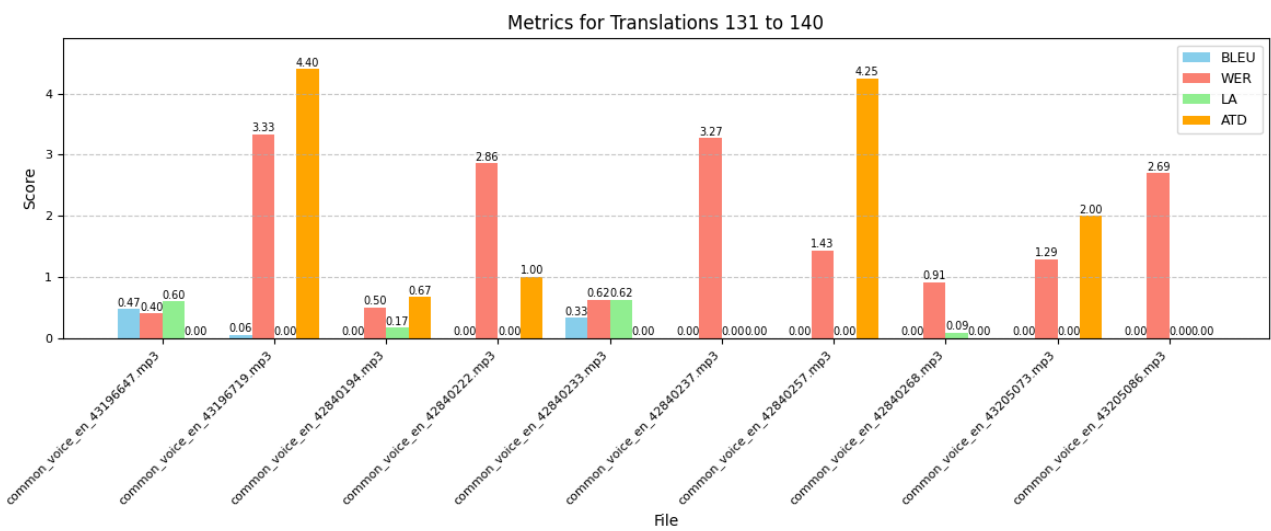
Batch 12 Avg: BLEU: 0.3482, WER: 1.1298, LA: 0.2753, ATD: 1.4775

Batch 13 Metrics:



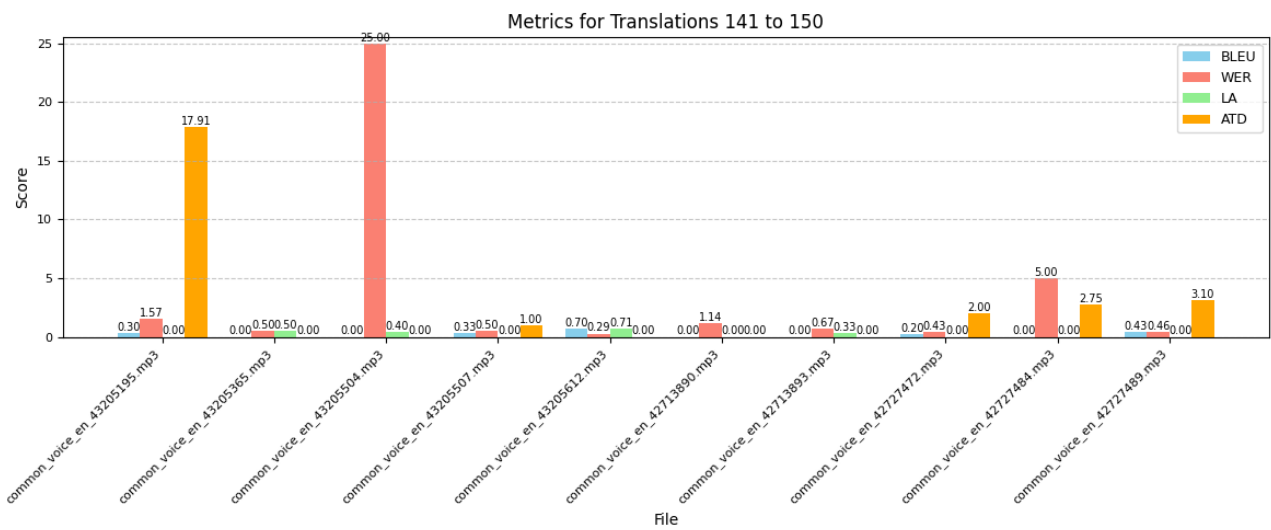
Batch 13 Avg: BLEU: 0.0430, WER: 2.3404, LA: 0.1010, ATD: 1.0383

Batch 14 Metrics:



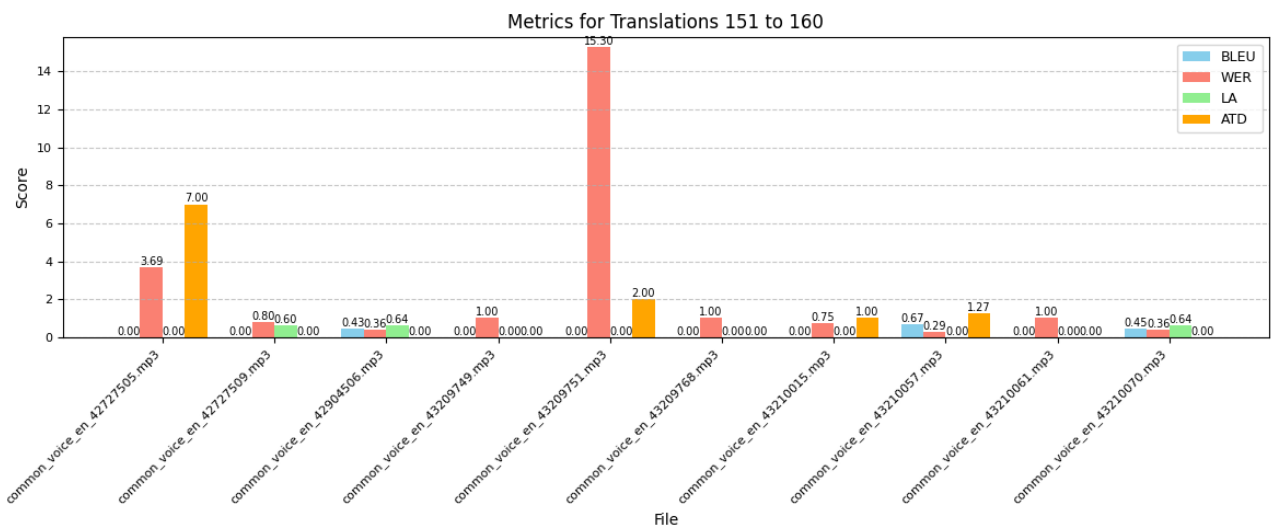
Batch 14 Avg: BLEU: 0.0861, WER: 1.7304, LA: 0.1483, ATD: 1.2317

Batch 15 Metrics:



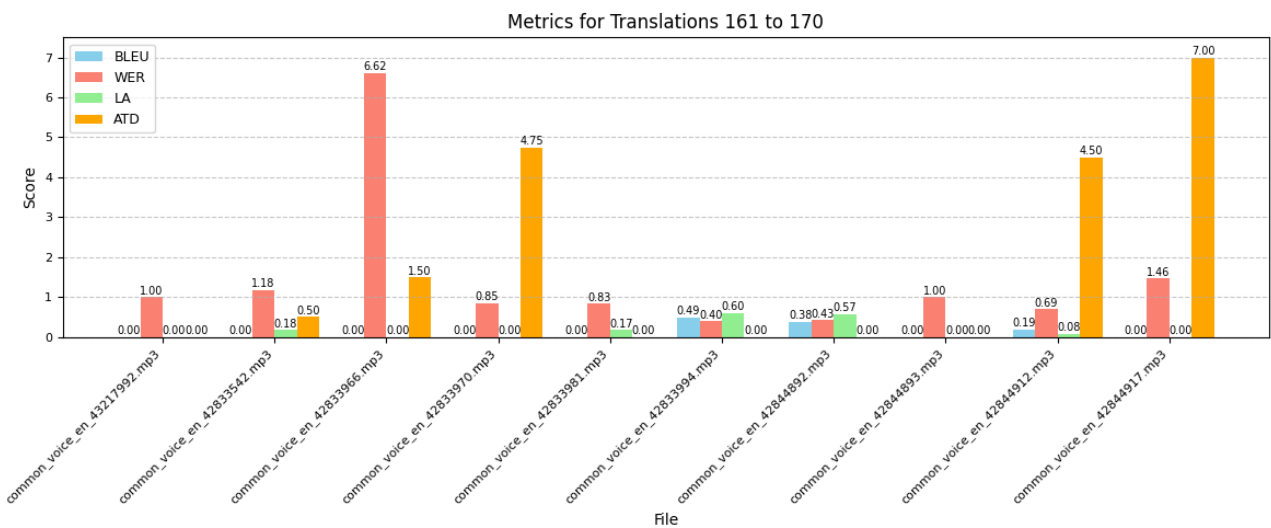
Batch 15 Avg: BLEU: 0.1965, WER: 3.5557, LA: 0.1948, ATD: 2.6759

Batch 16 Metrics:



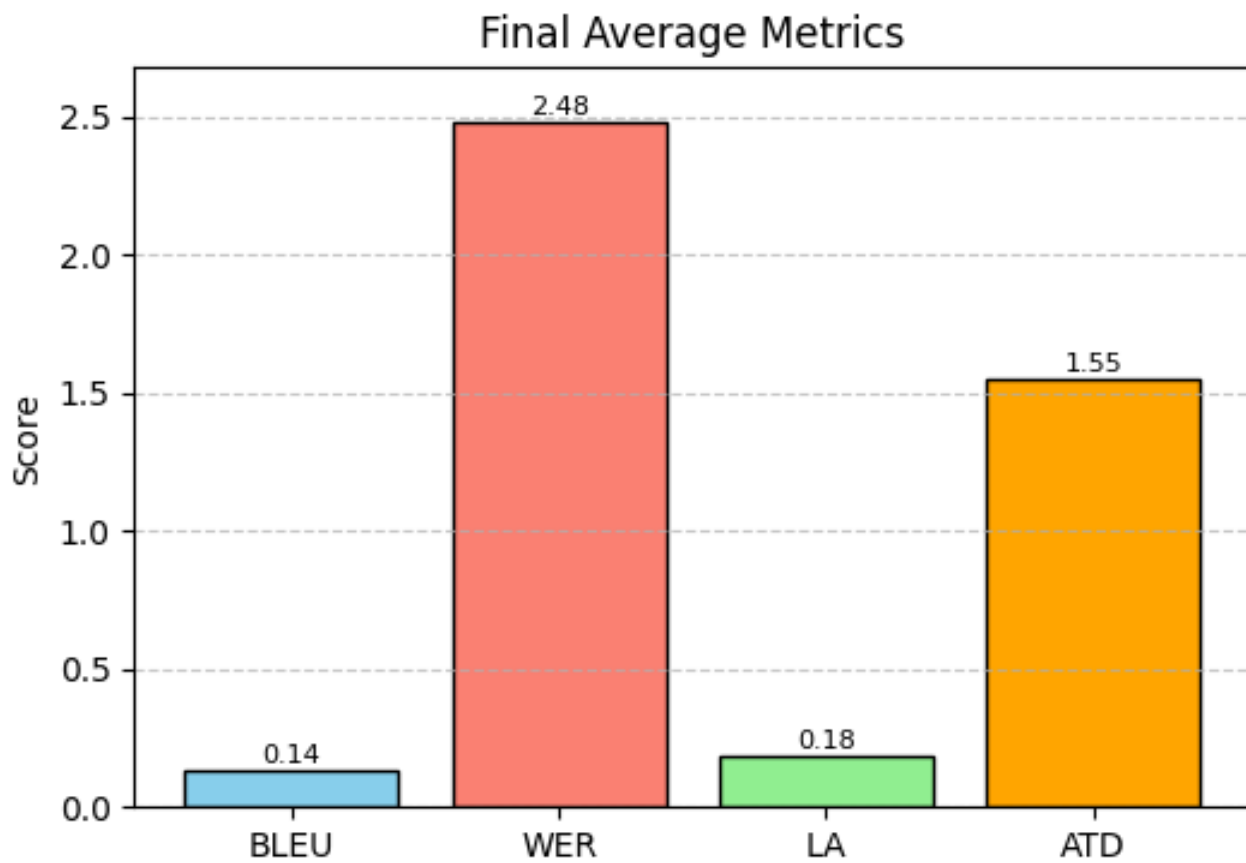
Batch 16 Avg: BLEU: 0.1548, WER: 2.4555, LA: 0.1873, ATD: 1.1273

Batch 17 Metrics:



Batch 17 Avg: BLEU: 0.1049, WER: 1.4459, LA: 0.1597, ATD: 1.8250

Final Average Metrics Plot:



#### Overall Average Metrics:

Avg Sentence BLEU: 0.1351

Avg WER: 2.4794

Avg LA: 0.1837

Avg ATD (proxy for Latency): 1.5504 seconds