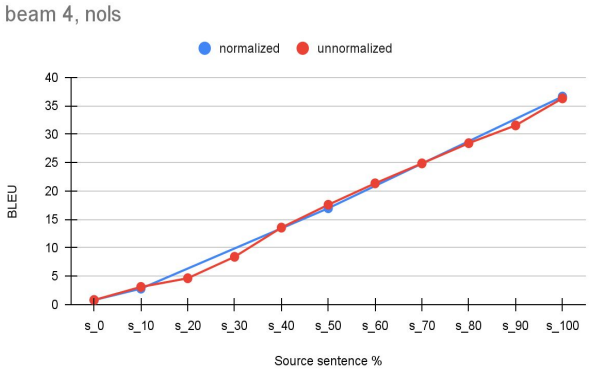
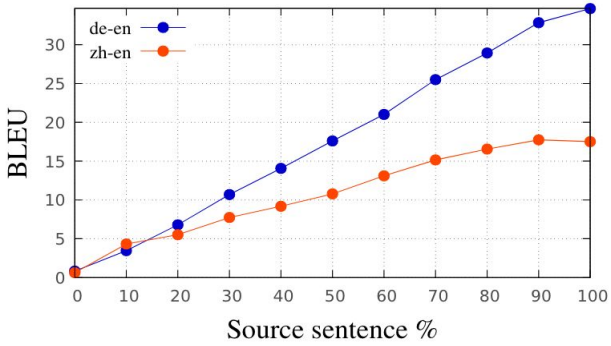
Baseline Riley & Chiang

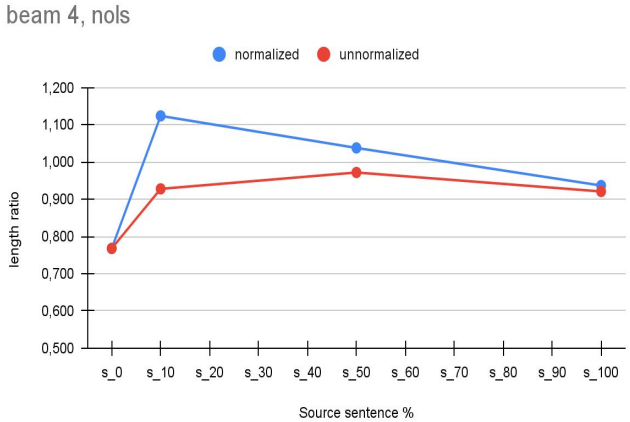
https://arxiv.org/abs/2210.10817

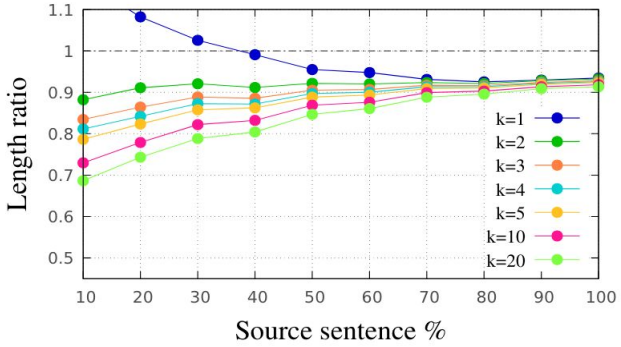
Detailed data at: Baseline Riley & Chiang

Length ratio normalized vs. unnormalized

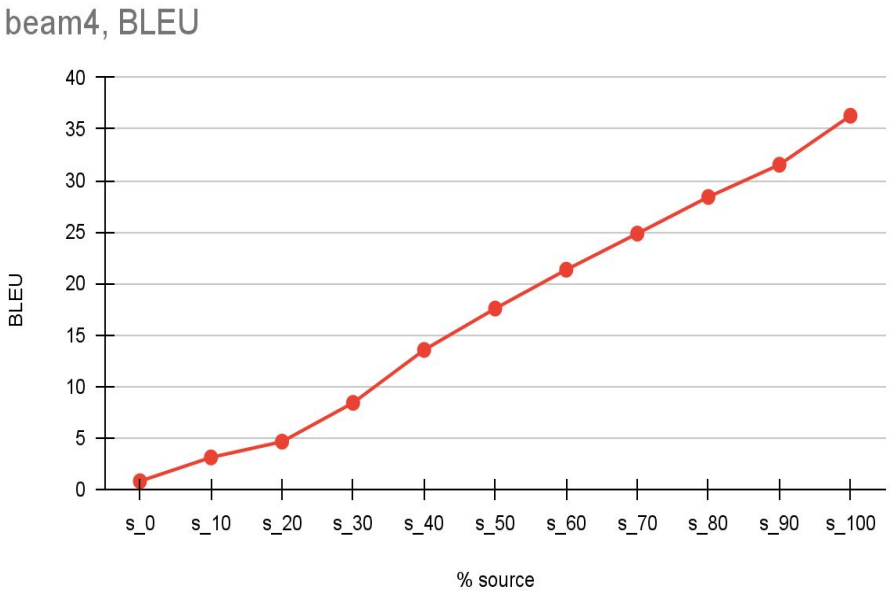


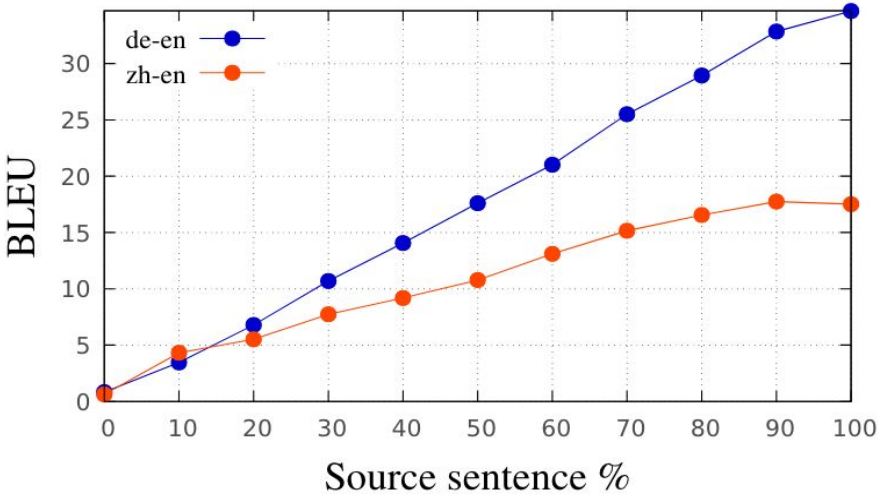






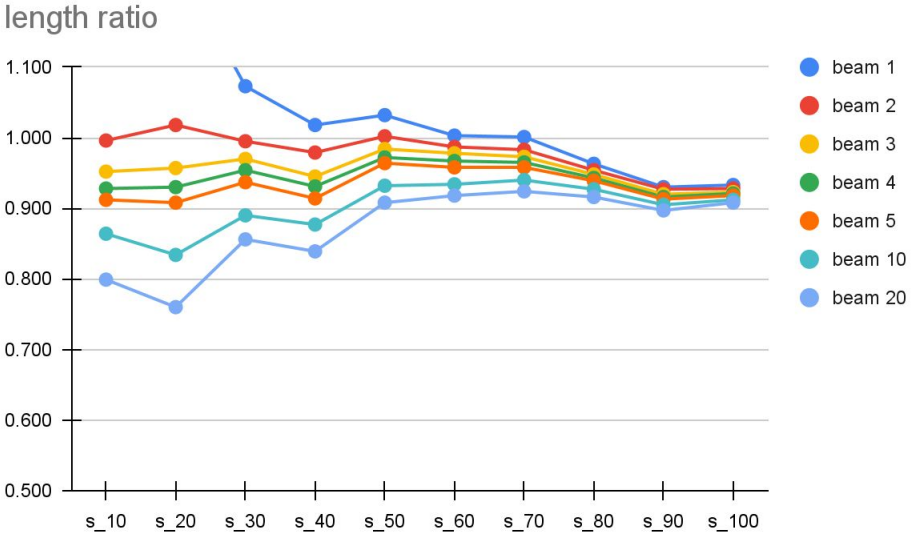
BLEU values (beam 4, de-en)





Length ratio experiments

Length ratio from fairseq



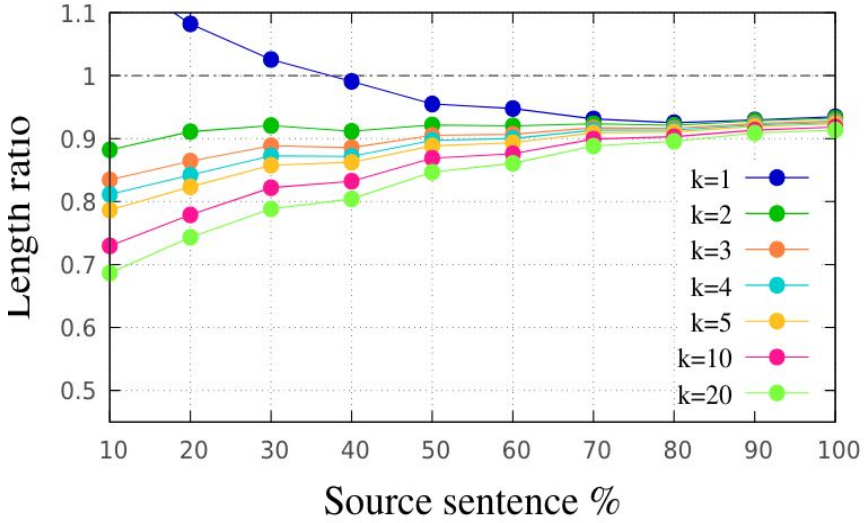
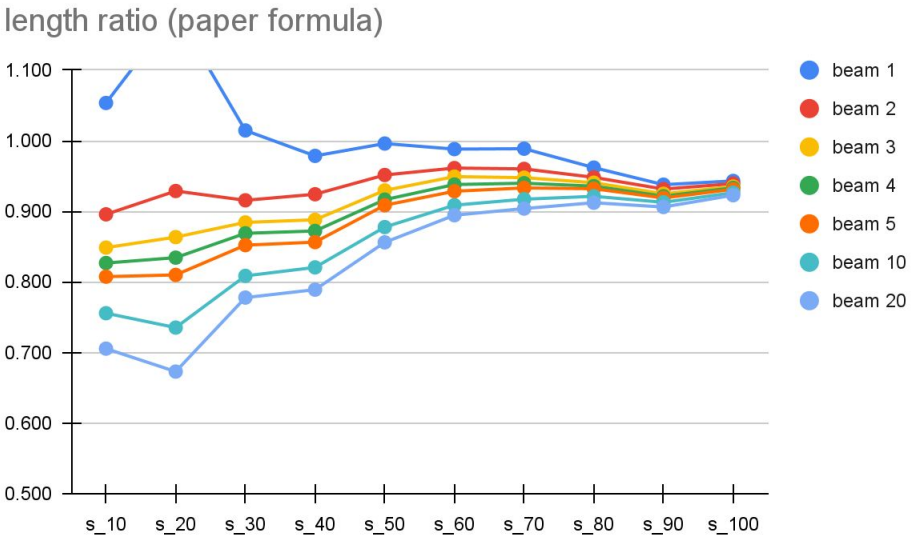


Figure 3: Length ratio versus source sentence percentage (s), for various beam sizes (k). For high s, there is a slight bias towards shorter outputs that increases mildly with k, whereas for low s, we see extreme bias, towards longer or shorter outputs depending on k.

Length ratio experiments

Length ratio as described in the paper



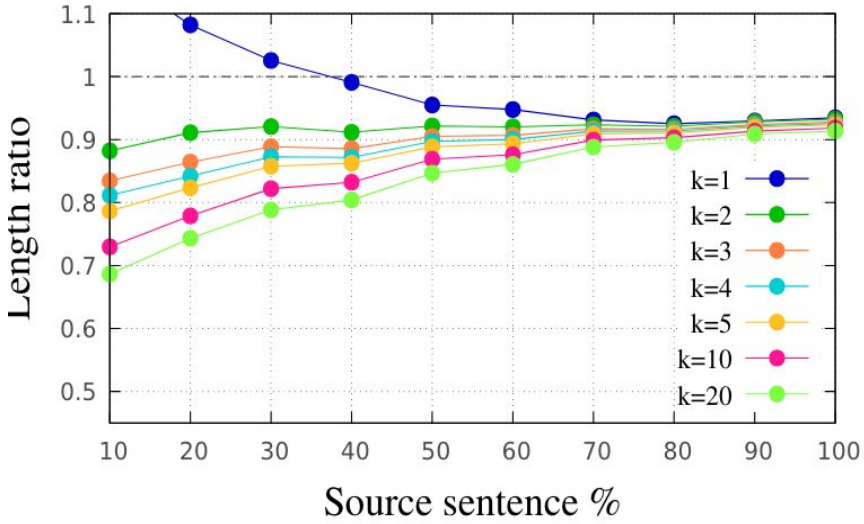
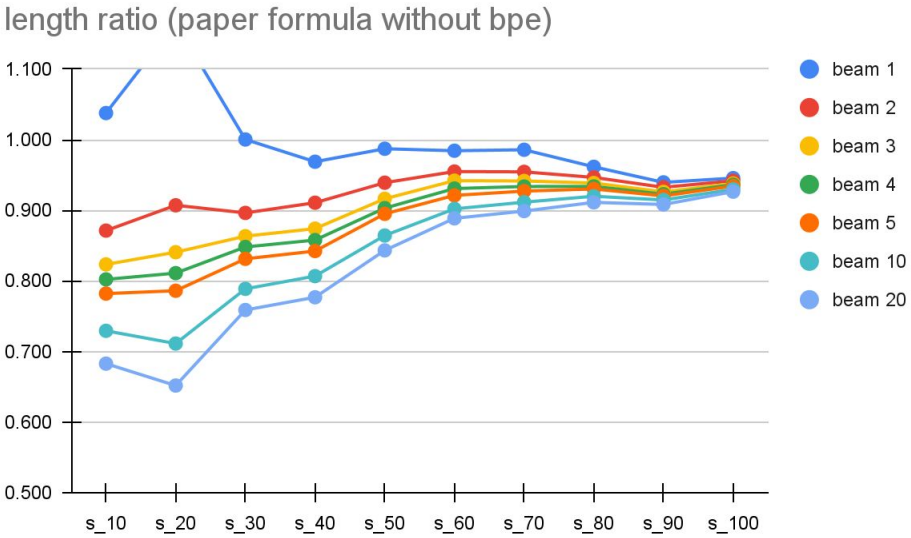


Figure 3: Length ratio versus source sentence percentage (s), for various beam sizes (k). For high s, there is a slight bias towards shorter outputs that increases mildly with k, whereas for low s, we see extreme bias, towards longer or shorter outputs depending on k.

Length ratio experiments

Length ratio as described in the paper but without bpe



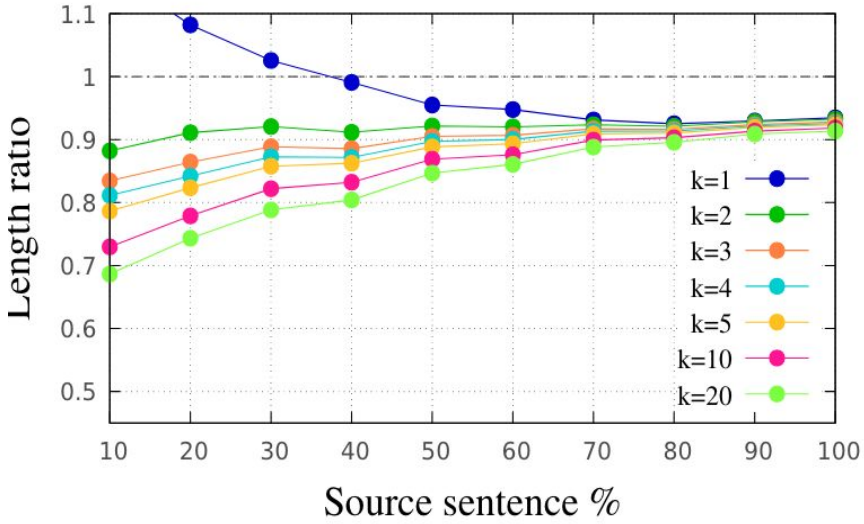
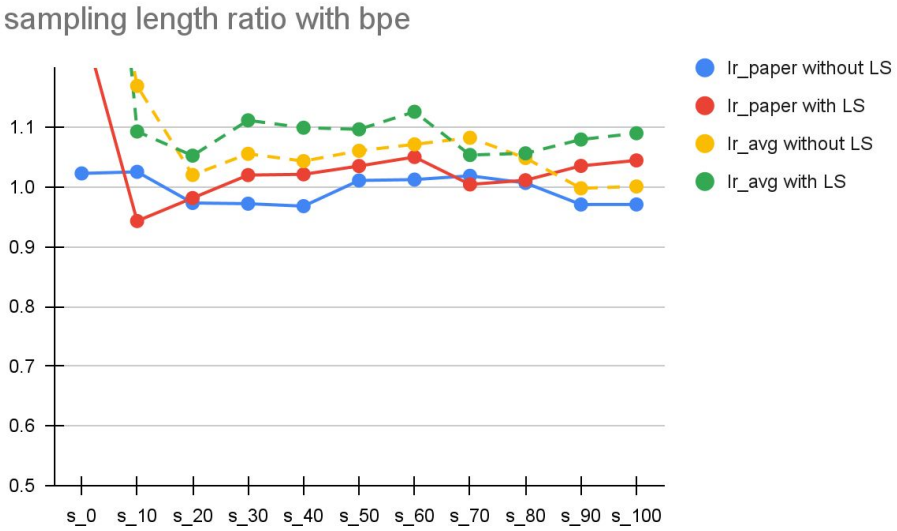


Figure 3: Length ratio versus source sentence percentage (s), for various beam sizes (k). For high s, there is a slight bias towards shorter outputs that increases mildly with k, whereas for low s, we see extreme bias, towards longer or shorter outputs depending on k.

Sampling experiments



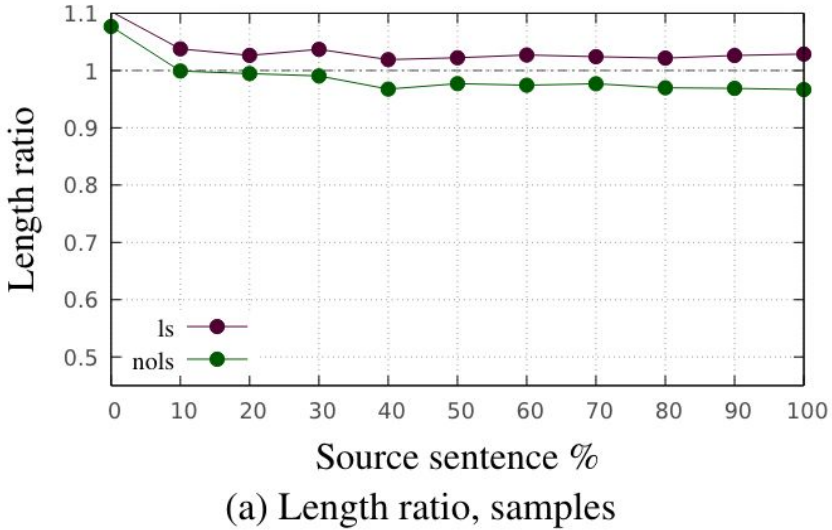
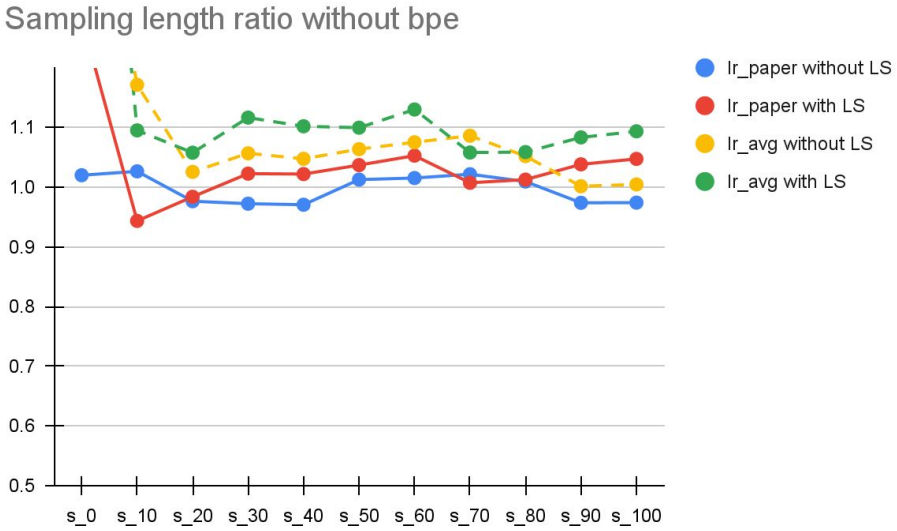


Figure 8: Length ratio of translations and percentage of unique 1-grams versus source sentence percentage (s), both with label smoothing (ls) and without (nols). Results for samples are computed based on 1000 samples for each test sentence; results for beam search vary across beam sizes (k). For samples, label smoothing increases the length ratio from slightly below the reference length to slightly above it; otherwise it has no discernible effect. (These results are for German-to-English; see Appendix C for Chinese-to-English.)

Sampling experiments



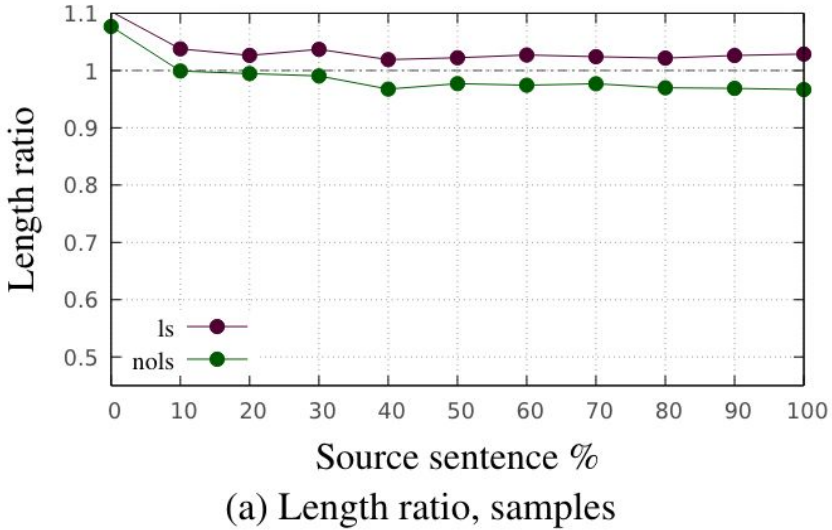
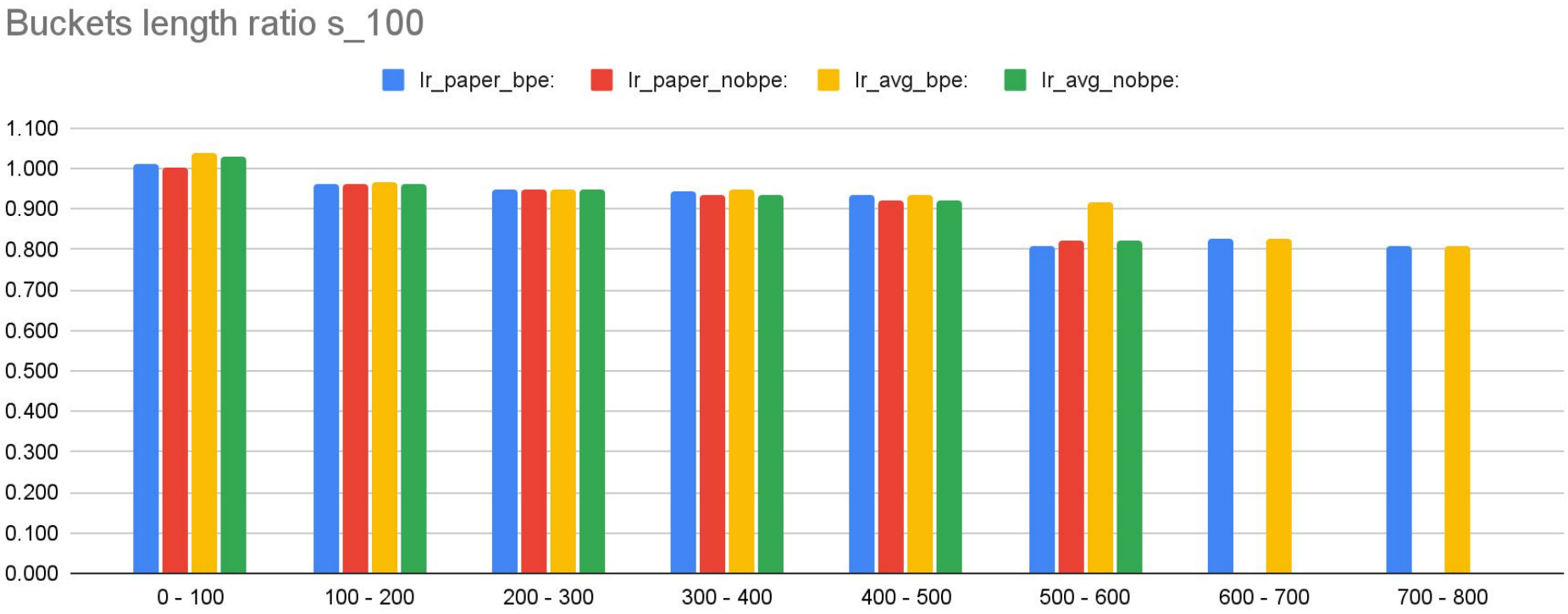


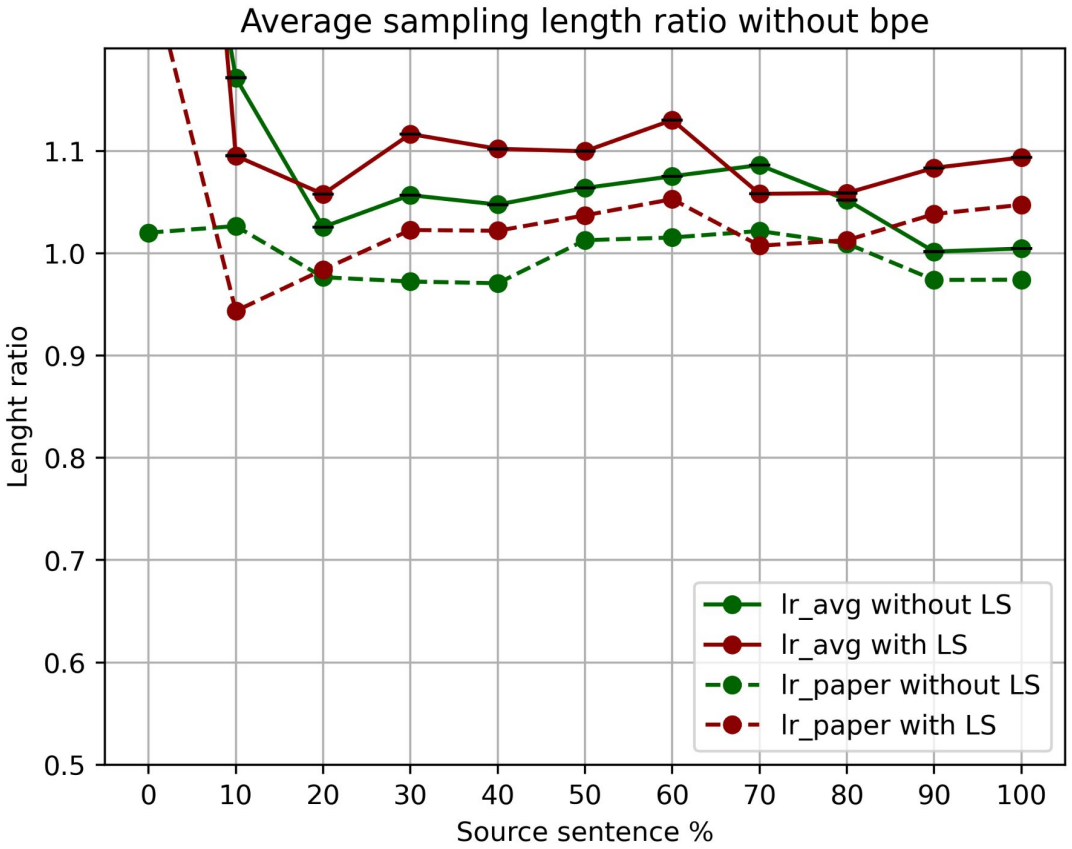
Figure 8: Length ratio of translations and percentage of unique 1-grams versus source sentence percentage (s), both with label smoothing (ls) and without (nols). Results for samples are computed based on 1000 samples for each test sentence; results for beam search vary across beam sizes (k). For samples, label smoothing increases the length ratio from slightly below the reference length to slightly above it; otherwise it has no discernible effect. (These results are for German-to-English; see Appendix C for Chinese-to-English.)

Length ratio buckets



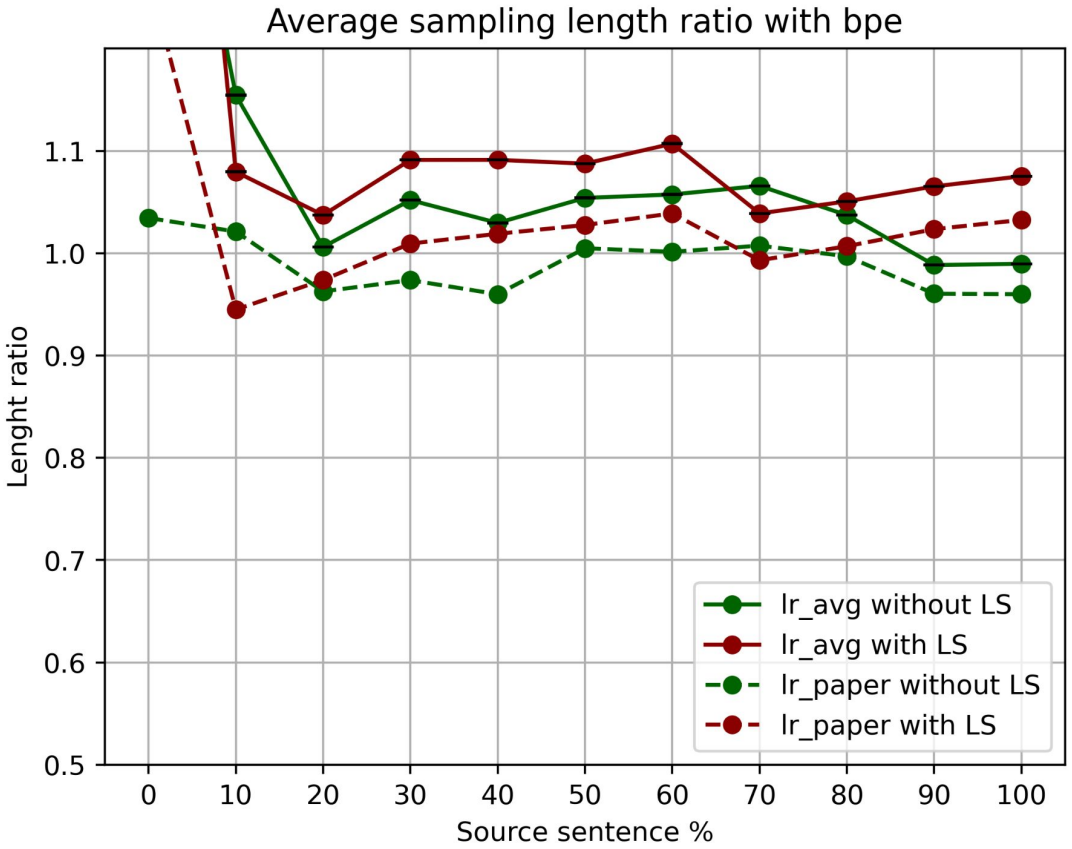
Standard error

- Sampling average length ratio without bpe



Standard error

- Sampling average length ratio with bpe



Standard error

- Length buckets for s_100 without bpe