MRL Turkish Subset Submission

Kerem Zaman
UNC Chapel Hill

This dataset was fully manually constructed by a native Turkish speaker. While most sentences were created entirely by the author, some are modified versions of sentences sourced from online food recipes. The dataset contains 100 sentences in total. Considering all potential continuations (200 sentences), the average sentence length is 87.9 ± 60.8 characters or 11.8 ± 8.03 words (tokenized on whitespace).