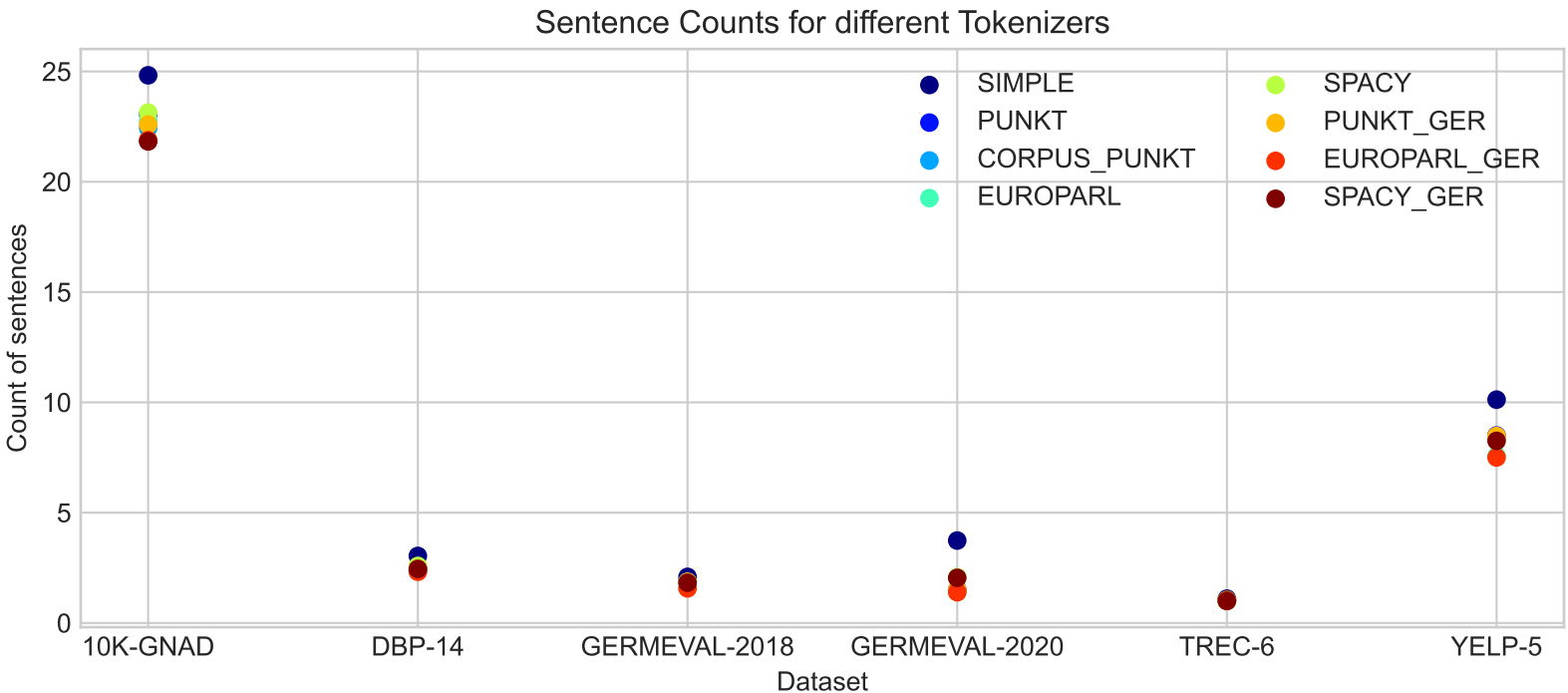
Sentence Counts for different Tokenizers
Count of sentences
Dataset
0
5
10
15
20
25
10K-GNAD
DBP-14
GERMEVAL-2018
GERMEVAL-2020
TREC-6
YELP-5
SIMPLE
PUNKT
CORPUS_PUNKT
EUROPARL
SPACY
PUNKT_GER
EUROPARL_GER
SPACY_GER