

BART

byte-level Byte Pair Encoding (BPE) ([\[1508.07909\] Neural Machine Translation of Rare Words with Subword Units](#))

- Better to keep the text nice and tidy
- Does consider new lines

Pegasus

SentencePiece ([google/sentencepiece: Unsupervised text tokenizer for Neural Network-based text generation.](#))

- Handles spaces around full stops
- Does not consider new lines

T5

SentencePiece ([google/sentencepiece: Unsupervised text tokenizer for Neural Network-based text generation.](#))

ProphetNet

WordPiece ([JAPANESE AND KOREAN VOICE SEARCH](#))