Flavours of Subword Units

- Byte Pair Encoding (BPE, Sennrich et al. (2016))
  `http://github.com/rsennrich/subword-nmt/`
- Google Wordpieces (Wu et al., 2016)
  Code probably unavailable, used in speech.
- SubwordTextEncoder in Tensor2tensor (Vaswani et al., 2017)
  `https://github.com/tensorflow/tensor2tensor`

STE	Blíží_ se_ k_ tobě_ **tramvaj** _ ._ Z_ **tramvaj** e_ nevysto upil i_ ._
BPE	Blíží se k tobě **tramvaj** . Z **tramva@@ je** nevy@@ stoupili .
BPE underscore	Blíží_ se_ k_ tobě_ **tramvaj@@ _** ._ Z_ **tramvaj@@ e_** nevy@@ stoupili_ ._

The best now is SentencePiece: `https://github.com/google/sentencepiece`