Task 2 (23.07.), check-in on 25.07.

- Use cleaned Shakespeare file (will be uploaded) — Train-Test-Validation set will be uploaded so that everyone has the same split
    - Use Validation set mainly to optimise hyperparameters in interpolation, don‘t use it to optimise k
- Develop n-gram engine (based on BPE encoding) that can deal with different n
    - Unigram system first, then bigram system, then 3- and 4-gram; intrinsic evaluation for each:
        - Report Perplexity (on BPE subwords)
            - For bigram: look at how different k‘s affect perplexity
        - „Add-one“ normalisation (Laplace Smoothing)
        - Simple (not conditional) interpolation or Backoff
- Write a program for extrinsic evaluation (generate sentence from n-gram system)
    - Give context first
    - Generation to predict next word (for now: argmax (most likely), or sampling for more variance)
        - If word not present: assign average probability of all unigrams or assign most likely word of unigram
        - Use end-of-sequence tokens to determine stop generation