Assignment 1

Notes

* Corpus:
  + Tutto in una riga (if line.startswith(“CHAPTER”), prima del Tokenizer)
    - Però si perde il cambio di contesto del paragrafo o del capitolo (dobbiamo pensarci/provarci)
  + Tieni apice in filtro, se si ha tempo togliere tutti i rimanenti delle abbreviazioni (I’m, togliere m).
* Dataset:
  + Padding su window\_length
* CBOW:
  + Batch\_size da sperimentare (1? o normale)

Entro 13 sera:

* Mattia -> implementazione 1.3
* Tommaso -> implementazione modelli