Preprocessing steps:

* each sentence is converted to lowercase and the punctuation is completely removed (following McMahan, Natural Language Processing with PyTorch)
* Other non-alpha characters are also removed

Note:

We include the classes which we used to load, clean and vectorize the data in the appendix at the end of this report. Each class in well commented and to avoid redundancy, we do not again describe our preprocessing process in detail. We just provide a brief summary of the performed steps.