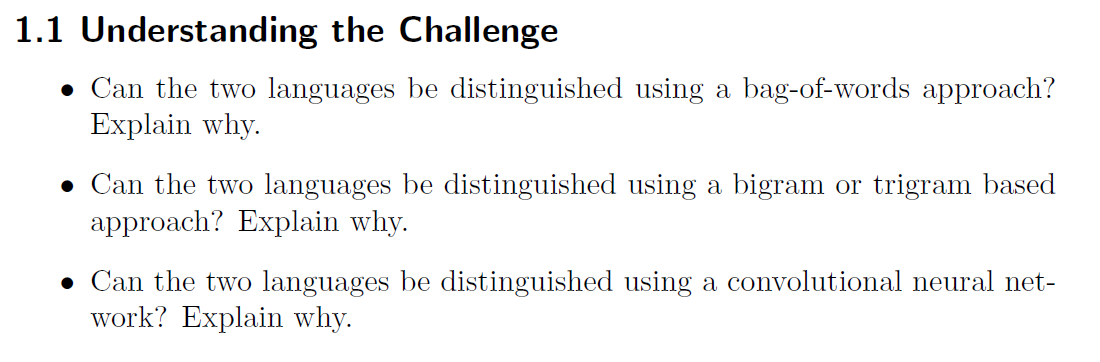
**ID 011862141**



* The two languages certainly cannot be distinguished using bag-of-words approach. The reason is bag of words does not retain information on the order of the components (characters in this case)
* The bigrams and trigrams that would be encountered are the same between the two languages. Bigram/Trigram approach cannot encode the order of the segments of words within the long string.
* Convolutional neural networks could potentially solve this problem:
  + As long as the overall sequence is below some max length, hard-coding the (potentially very long) length of maximum “interaction” length.
  + Using a hierarchical CNN, longer and longer sequences can be considered together and so it can potentially learn the syntax.