* The formula for score was slightly modified:
  + Z = (count of triplets where central words are w1 and w2 and the context for both are the same) / (count of all triplets where the central words were w1 and w2)
  + Z = (count of triplets where central words are w1 and w2 and the context for both are the same) / (count of all triplets which had count of occurrences of the context of the words w1 and w2)
* This is a Naïve approach which is heavily dependent on the training data, i.e. the appearance of the words in the context in the training data.
* Some observations in tweets are:
  + There are a lot of spelling errors. eg.: “Thanks to Mr. Modi all my ungles and aunthies in Abu Dubhai will finally go to their own temble. Thiz iz zimbly zuperb”
  + Other languages are transliterated to English. eg.: “kullam khulla becho desh ko”
  + Extracting a tweet is hard as it spans several lines arbitrarily in certain cases.
  + Short forms are used in certain scenarios which is not present abundantly enough in the corpus to train for such occurrences.