

8 Natural Language Processing (AAC)

- (a) Use the following text (from Eliot’s ‘The Hollow Men’) to derive distributions for *eyes*, *here* and *valley*. Use a 5-word window including open- and closed-class words, ignore case and line-breaks and weight contexts by frequency.

The eyes are not here
There are no eyes here
In this valley of dying stars
In this hollow valley
This broken jaw of our lost kingdoms

[5 marks]

- (b) What is cosine similarity? Show how to calculate the cosine similarity between each pair of the distributions you produced above (i.e., eyes/here, eyes/valley, valley/here). [5 marks]

- (c) The following table shows some similarities for pairs of nouns calculated using distributions extracted from the British National Corpus.

ring	bracelet	0.64	ring	cat	0.13
ring	necklace	0.55	lord	knight	0.37
ring	finger	0.54	lord	ring	0.41

Discuss what these results suggest about distributional similarity with respect to human similarity judgements and with respect to the concept of synonymy used in lexical semantics. [4 marks]

- (d) Describe the notion of textual entailment, as investigated in the Recognising Textual Entailment (RTE) challenge. Outline how one could investigate the use of distributional methods to improve the performance of a bag-of-words method for this task. [6 marks]