5 Lexical Semantics [Lecture 5]

Exercise 5.1
(a) Give brief definitions of the following terms. Illustrate the definitions with examples.
   (i) hyponymy
   (ii) meronymy
   (iii) antonymy

(b) Describe Yarowsky’s (1995) technique for word sense disambiguation and illustrate how it would disambiguate the following two senses of sake:
   Sense 1: sake, interest (a reason for wanting something done: “for your sake”, “died for the sake of his country”)
   Sense 2: sake, saki, rice beer (Japanese alcoholic beverage made from fermented rice, usually served hot)

6 Distributional Semantics [Lecture 6, Lecture 7]

Exercise 6.1
(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

(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).

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

Exercise 6.2

The distributional hypothesis states that the meaning of a word can be defined by its use and, therefore, it can be represented as a distribution of contexts in which the word occurs in a large text corpus.

(a) Describe four different types of context that can be used for this purpose.

(b) The contexts can be weighted using Pointwise Mutual Information (PMI).

   Explain, giving formulae, how PMI is calculated and how individual probabilities are estimated from a text corpus.

(c) Some words occur very rarely in the corpus. How does this affect their PMI scores as contexts?

(d) The goal of distributional word clustering is to obtain clusters of words with similar or related meanings. The following clusters have been produced in two different noun clustering experiments:

   Experiment 1:

   | carriage bike vehicle train truck lorry coach taxi |
   | official officer inspector journalist detective |
   | constable policeman reporter |
   | sister daughter parent relative lover cousin friend |
   | wife mother husband brother father |

   Experiment 2:

   | car engine petrol road driver wheel trip steering |
   | seat highway sign speed |
   | concert singer stage light music show audience |
   | performance ticket |
   | experiment research scientist paper result |
   | publication laboratory finding |

   (i) How are the clusters produced in the two experiments different with respect to the similarity they capture? What lexico-semantic relations do the clusters exhibit?

   (ii) The same clustering algorithm, K-means, was used in both experiments. What was different in the setup of the two experiments that resulted in the different kinds of similarity captured by the clusters?
8 Compositional Semantics [Lecture 8]

Exercise 8.1

(a) What deficiencies of linguistic distributional models does multimodal semantics address?

(b) Describe Bag-of-Visual-Words (BoVW) model and how it would apply to learning visual representations of linguistic concepts.

(c) Explain the difference between feature level fusion and scoring level fusion in building multimodal representations.

(d) Discuss how the bias of visual data would affect certain tasks with visual representation.

Exercise 8.2

(a) How is syntax encoded in lexical function models? What deficiencies of vector mixture models do lexical function models address?

(b) Given the toy examples of adjective matrices and noun vectors below,

\[
\begin{array}{c|cc}
\text{old} & \text{runs} & \text{barks} \\
\hline
\text{runs} & 0.5 & 0.2 \\
\text{barks} & 0 & 0.8 \\
\end{array}
\quad
\begin{array}{c|cc}
\text{dead} & \text{runs} & \text{barks} \\
\hline
\text{runs} & 0 & 0.2 \\
\text{barks} & 0.3 & 0.1 \\
\end{array}
\quad
\begin{array}{c|cc}
\text{dog} & \text{runs} & \text{barks} \\
\hline
\text{dog} & 1 & 5 \\
\text{cat} & 4 & 0 \\
\end{array}
\]

(i) What do the matrices show about the semantics of the adjective?

(ii) Compute the vector representations of the phrases old dog and dead dog using lexical functional model. How do the adjectives affect the semantics of the noun when deriving an Adj Noun phrase?