Natural Language Processing - Supervision 2

A.H. Zaidi

5 Dependencies

(a) Give a parse tree (or a bracketed structure) for the following sentences, and show how comparable dependency structures could be constructed using the algorithm described in the lecture notes (assuming a suitable oracle).

(i) the dog wants fish

(ii) a small dog wants a small carrot

(iii) they fish in rivers in December

6 Compositional semantics

(a) Using the sample grammar provided, produce a derivation for the semantics of:

(i) Kitty sleeps.

(ii) Kitty gives Lynx Rover

7 Lexical semantics

(a) Give brief definitions of the following terms. Illustrate the definitions with examples.

(i) hyponymy

(ii) meronymy

(iii) antonymy

(b) Choose three nouns from WordNet with between two and five senses. For each noun, find 5 or more sentences in the BNC which use that noun and assign a sense to each occurrence.

8 Distributional semantics

(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.

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

(d) Discuss what these results of (c) suggest about distributional similarity with respect to human similarity judgments and with respect to the concept of synonymy used in lexical semantics.

(e) 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.

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

(ii) 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.

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