

# Natural Language Processing: Part II

## Overview of Natural Language Processing (L90): Part III/ACS

Exercises for lectures 6 to 8

### 1 Lecture 6

#### 1.1 Pre-lecture

A very simple form of semantic representation corresponds to making verbs one-, two- or three- place logical predicates. Proper names are assumed to correspond to constants. The first argument should always correspond to the subject of the active sentence, the second to the object (if there is one) and the third to the indirect object (i.e., the beneficiary, if there is one). Give representations for the following examples:

1. Kim likes Sandy  
Answer: like(Kim, Sandy)
2. Kim sleeps
3. Sandy adores Kim
4. Kim is adored by Sandy (note, this is passive: the *by* should not be represented)
5. Kim gave Rover to Sandy (the *to* is not represented)
6. Kim gave Sandy Rover

#### 1.2 Post-lecture

1. Using the sample grammar provided in the notes, produce a derivation for the semantics of:
  - Kitty sleeps.
  - Kitty gives Lynx Rover.
2. Extend the grammar so that *Kitty gives Rover to Lynx* gets exactly the same semantics as *Kitty gives Lynx Rover*. You can assume that *to* is semantically empty in this use.
3. Go through the RTE examples given in the lecture notes, and decide what would be required to handle these inferences correctly.

### 2 Lecture 7

#### 2.1 Pre-lecture

- Without looking at a dictionary, write down brief definitions for as many senses as you can think of for the following words:
  1. plant
  2. shower
  3. bass

If possible, compare your answers with another student's and with a dictionary.

- Using a BNC search (or another suitable corpus search tool: i.e., one that doesn't weight the results returned in any way), go through at least 10 sentences that include the verb *find* and consider whether it could have been replaced by *discover*. You might like to distinguish between cases where the example 'sounds strange' but where meaning is preserved from cases where the meaning changes significantly.

## 2.2 Post-lecture

1. Give hypernyms and (if possible) hyponyms for the nominal senses of the following words:
  - (a) horse
  - (b) rice
  - (c) curtain
2. Choose three nouns from WordNet with between two and five senses. For each noun, find 10 or more sentences in the BNC which use that noun and assign a sense to each occurrence. Swap the unannotated data with one or two other people and ask them to repeat the exercise independently. Calculate the percentage agreement between you. Discuss each case where you disagreed (if you've got a lot of disagreements, just look at a subset of the cases) and see if you can work out where your assumptions differed. Can you come to an agreement you all feel happy with on any of these cases? If you've done this exercise with three people, how many of these agreed decisions correspond to the original majority decision?

## 3 Lecture 8

### 3.1 Pre-lecture

Without looking at a dictionary, write down brief definitions for as many senses as you can think of for the following words:

1. give
2. run

If possible, compare your answers with another student's and with a dictionary. How does this exercise compare with the pre-lecture exercise for lecture 7?

### 3.2 Post-lecture

1. Using a BNC search (or another suitable corpus search tool: i.e., one that doesn't weight the results returned in any way), find 10 or more sentential contexts for *shower*. For each of the different notions of context described in the lecture, find the features which a distributional model might associate with *shower*. You may want to use an online dependency parser: the Stanford dependency format is one of the most popular approaches (see [nlp.stanford.edu/software/stanford-dependencies.shtml](http://nlp.stanford.edu/software/stanford-dependencies.shtml), online demo at <http://nlp.stanford.edu:8080/corenlp/>). If you use an online parser, note that the output is unlikely to be perfectly accurate.
2. List some possible advantages and disadvantages of using Wikipedia as a corpus for experiments on distributional semantics compared with:
  - (a) The BNC
  - (b) A 2 billion word corpus of American newspaper text
  - (c) The UKWac corpus (see <http://www.sketchengine.co.uk/documentation/wiki/Corpora/UKWac>)
  - (d) The Google 5-gram corpus <http://googleresearch.blogspot.co.uk/2006/08/all-our-n-gram-are->  
html