HW4: Word Meaning and Word Similarity

Due: Dec 12th 2017, 23:59 EST

These are theoretical questions. You will submit your answers in PDF format.

Q1. Examine the entries for these lexemes {***imitation, synthetic, artificial, fake, simulated***} in WordNet. Do you agree with WordNet in terms of the number of senses assigned to each of these lexemes? Compare them to the Collins Cobuild entries for the same words.

Q2. The WordNet entry for the noun ***bat*** lists 6 distinct senses. Cluster these senses using the definitions of homonymy and polysemy given in the slides. For any senses that are polysemous, give an argument as to how the senses are related.

Q3. For a pair of languages that you are familiar with, find three cases of an ambiguous word in the first language for which the senses translate into different words and three cases of an ambiguous words for which at least two senses translate to the same word.

Q4. Collect a small corpus of 20 example sentences of varying lengths from any newspaper or magazine. Using WordNet, or any standard dictionary, determine how many senses there are for each of the open-class words in each sentence. How many distinct combinations of senses are there for each sentence? How does this number seem to vary with sentence length?

Q5. Using WordNet, or a standard reference dictionary, tag each open-class word in your corpus with its correct tag. Was choosing the correct sense always a straightforward task? Report on any difficulties you encountered.

Q6. Using the same corpus, isolate the words taking part in all the verb-subject and verb-object relations. How often does it appear to be the case that the words taking part in these relations could be disambiguated using only information about the words in the relation?

Q7. Evaluate two or three of the similarity methods from the publicly available Wordnet::Similarity package (Pedersen et al., 2004). You might do this by hand-labeling some word pairs with similarity scores and seeing how well the algorithms approximate your hand labels.