LING571 – Hw9
Hobbs Algorithm

1. Overview

Hobb's algorithm was implemented in Python with the help of NLTK.

2. Methodology

I used NLTK to determine the POS tags for sentences in conjunction with a simple feature grammar to keep track of the NUMBER aspect of noun phrases. This provided almost everything we needed to complete the implementation of Hoggs Algorithm.

3. Results

A significant proportion (66.6%) of the pronouns were successfully resolved using Hobbs algorithm. This in itself doesn't seem very high but I remember hearing in class that in general, the accuracy of this algorithm is closer to 85+%. I also think that certain heuristics like binding theory would further enhance the effectiveness of the algorithm.

4. Challenges and Learning Outcomes

The NLTK library really let down the team this round. The parented tree structure is really not very nice to work with. I think that, finding ways around NLTK occupied more time than anything else in this assignment. In particular, using feature grammars and extracting the POS and feature information required casting a property of the parented tree to different objects; which is not intuitive and requires time to find work-arounds.

Otherwise, the wording of the algorithm is not very clear and could be improved. Personally, I see little value in having to put in effort to decode the algorithm description. It could be argued that it enhances our ability to think critically, but the same thing could be said of arguing with my mother-in-law; but, in my opinion, neither of these ventures pay dividends.

5. Insights

This assignment was interesting in two particular ways:

1. It was rewarding to see the Hobbs algorithm, which is a relatively simple algorithm, is able to largely solve the fairly complicated problem of pronoun resolution.
2. It was also nice to draw together a few skills that we learned throughout the course; particularly using NLTK and feature grammars to make this implementation relatively simple.

Although the algorithm performs relatively well, it is insufficient to solve more complex problems of pronoun ambiguity, for example. I am more convinced than ever that a semantically aware model is necessary to completely solve this class of problems. In the absence of this, I believe that a probabilistic model could improve on these baseline results.

6. Completeness

I completed the assignment.

Finally, …

Thank you very much for marking our assignments. 😊