

School of Computing and Information Systems
The University of Melbourne
COMP90042
NATURAL LANGUAGE PROCESSING (Semester 1, 2022)
Workshop exercises: Week 9

Discussion

1. What differentiates **probabilistic CYK parsing** from **CYK parsing**? Why is this important? How does this affect the algorithms used for parsing?
2. What is a **probabilistic context-free grammar** and what problem does it attempt to solve?
3. A hidden Markov model assigns each word in a sentence with a tag, e.g.,

Donald/NNP has/VBZ small/JJ hands/NNS

The probability of the sequence is based on the tag-word pairs, and the pairs of adjacent tags. Show how this process can be framed as a CFG, and how the various probabilities (e.g., observation, transition, and initial state) can be assigned to productions. What are the similarities and differences between CYK parsing with this grammar, and the HMM's Viterbi algorithm for finding the best scoring state sequence?

4. Using typical dependency types, construct (by hand) a dependency parse for the following sentence: *Yesterday, I shot an elephant in my pyjamas.* Check your work against the output of the online GUI for the Stanford Parser (<https://corenlp.run/>).
5. In what ways is **(transition-based, probabilistic) dependency parsing** similar to **(probabilistic) CYK parsing**? In what ways is it different?