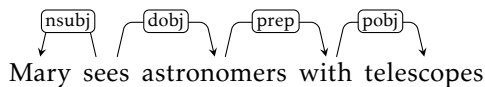


Dependency Parsing exercises:

Transition-based parsing

Deadline: 22.05.2017. Please send the homework to petit.jean@phil.hhu.de and cranenburgh@phil.hhu.de with subject "dependency homework" and an attachment named "ex4_lastname1_lastname2.pdf".

1. Enumerate the configurations an arc-standard transition-based parser goes through when parsing the sentence:



A transition is a left-arc, right-arc, or shift operation (LA, RA, or S). At each step, indicate the operation, the contents of the stack, the input buffer, and which dependency is added, if any:

- initial state: [Mary] [sees astronomers with telescopes] \emptyset
 - LA_{SBJ} [] [sees astronomers with telescopes] Mary $\xleftarrow{\text{SBJ}}$ sees
 - ...
2. It turns out that we have a corpus with some more information on the kinds of attachments telescopes tend to have:
 - Mary is an astronomer
 - Mary works with telescopes
 - An astronomer sees stars through telescopes

In this corpus, all prepositions are attached to verbs! Assume that an arc-standard model is trained with the following feature templates:

- the word below the top of the stack
- the word on top of the stack
- the first word in the input buffer
- the second word in the input buffer

An example of a configuration after a LA_{SBJ} transition, and its features would look as follows:

- [Mary] [is an astronomer] \emptyset
- (ε , Mary, is, an)

- (a) We expect to arrive at a different analysis if we train on this corpus. Assume we train a simple nearest neighbor classifier with $k = 1$, i.e., predict the transition of the single most similar configuration in the training data.

Which configuration, transition, and feature will be responsible for changing the attachment of the preposition as compared to the analyses in the previous exercises?

- (b) Suppose we have a larger corpus about telescopes, and it turns out there is a lot of ambiguity with respect to telescopes and their attachments. What kind of feature would work best: part of speech tags, lemmas, or word forms? What are the trade offs?