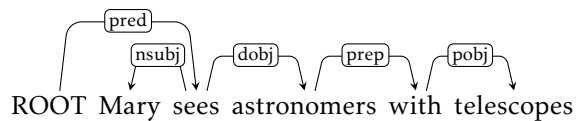


Dependency Parsing exercises: Transition-based arc-standard parsing

Deadline: 18.06.2018. Please send completed solutions to jakub.waszczyk@phil.uni-duesseldorf.de with subject "dependency homework".

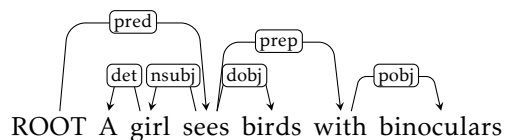
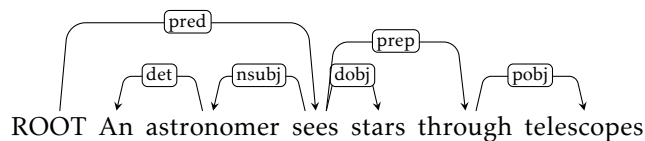
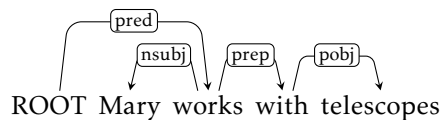
1. Enumerate the configurations an arc-standard transition-based parser goes through when parsing the sentence:



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: [ROOT] [Mary sees astronomers with telescopes] \emptyset
- SH: [ROOT Mary] [sees astronomers with telescopes]
- LA_{SBJ}: [ROOT] [sees astronomers with telescopes] + (Mary \xleftarrow{SBJ} sees)
- ...

2. It turns out that we have a small corpus with some more information on the kinds of attachments "sees" and "telescopes" tend to have:



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 and its features would look as follows:

- [ROOT] [Mary works with telescopes] \emptyset
- (ϵ , ROOT, Mary, works)

(a) We expect to arrive at a different analysis of the sentence in Ex. 1 if we train on this corpus. Assume we train a simple nearest neighbor classifier, i.e.:

- For a given parsing configuration, the parser chooses the transition of the single most similar configuration in the training dataset.

Which configuration, transition, and features in the training dataset will be responsible for changing the attachment of the preposition as compared to the analysis in the previous exercise?

(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 to deal with ambiguity: part of speech tags, lemmas, or word forms? What are the trade offs of different feature kinds?