

Finite-state machines exercise

9.19(0) Fall 2023, Instructor: Roger Levy

2023 October 2

Regular grammar of a fragment of English syntax

Every language has a limited inventory of PARTS OF SPEECH, for example here are a few in English:

Part of Speech	Examples
N(oun)	<i>dog, mud, justice, hearing</i>
V(erb)	<i>read, exists, be, walking, went, give</i>
Adj(ective)	<i>red, bright, former, tall</i>
P(reposition)	<i>on, toward, of, between</i>
Det(eterminer)	<i>the, a, most, my, whose, each, two, this</i>
Adv(erb)	<i>very, formerly, well, often</i>
Pron(oun)	<i>I, me, you, she, they, them</i>

Another part of syntax is PHRASAL CATEGORIES. One example is the NOUN PHRASE (NP). Roughly speaking, an NP is any string of words in a sentence that, in the proper context, can be substituted by a pronoun without changing the propositional content of that sentence:

Wilbur likes rolling in mud. →

Mud is soft and warm. Wilbur likes rolling in *it*. →

A very tall tree shades the bench. →

A very tall tree stands in the middle of the park. *It* shades the bench. →

Most rooms in her house have windows. →

Most rooms in her house are well illuminated in the day. *They* have windows.

Kim lent the youngest student in the kindergarten class a hand. →

The youngest student in the kindergarten class was having trouble. Kim lent *her* a hand.

Exercise: Taking into account the examples above, try to write a regular grammar of the set of English noun phrases over the part-of-speech symbol set $\Sigma = \{N, Adj, Det, P, Adv\}$,

by constructing a finite-state automaton that accepts all and only sequences of these part-of-speech sequences that are instantiated by some actual English noun phrase. At every stage of constructing your automaton, test by trying to come up with legitimate English NPs that it does not accept.¹

¹Note that in this exercise you will not be able to avoid OVERGENERATING at the level of specific English word sequences—that is, accepting certain word sequences that really should be rejected. For example, you will need to accept the sequence Det N in order to account for the NP *a dog*, but once you do, you won't be able to rule out **a mud*. This is OK for this exercise, but it points to the ultimate need for richer representations of grammar. Also, please note that the complete grammar of English NPs is not regular, as we will see later in the semester—but the complete grammar of English NPs constructed from the part-of-speech inventory I've given you *is* regular!