Homework 1: Syntax

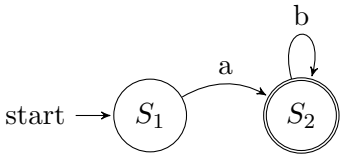
LING 5001 Introduction to Linguistics

FSAs and CFGs

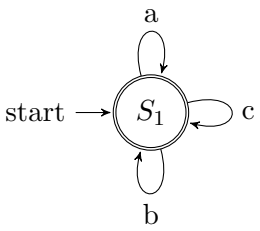
First, we will explore the properties of Finite State Automata and Context-Free Grammars as mathematical objects. That is, we will start by examining "languages" that consist of small strings of lower-case letters before moving onto natural language.

1.) (10 points) The following five Finite State Automata generate an infinite number of strings. For each Finite State Automata, provide three strings that it generates.

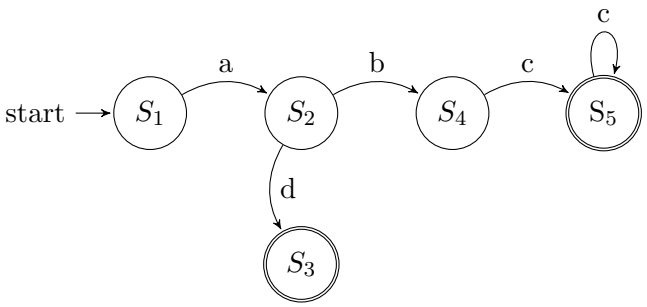
1.



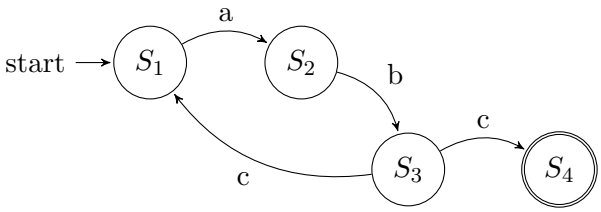
2.



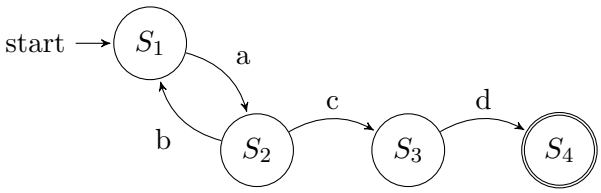
3.



4.



5.



Next, we will examine the relationship between Contet-Free Grammars (CFGs) and Finite State Automata (FSAs). For the purposes of this exercise, our start symbol will be `S`, our non-terminal nodes (i.e., the equivalent of NP, VP, ...) will be capital letters (`X, Y, Z, ...`), and our terminal nodes will be lower-case letters (*a, b, c...*).

Some CFGs can be rewritten as FSAs. These CFGs must have a specific form in order for this to be possible. Each rewrite rule must contain one terminal symbol, and at most one non-terminal symbol. In all rewrite rules, the non-terminal node must be on the same side as the terminal node, i.e., all non-terminal nodes must be on the left of the terminal node, or all non-terminal nodes must be on the right.

For example, the following CFG can be rewritten as the following FSA, because every rewrite symbol has zero or one non-terminal symbols, and because all the non-terminal symbols are on the right. In order to see that these are equivalent, look at the relationship between the recursive step in the CFG (i.e., the rules that contain the symbol `Y` on the left side of the rewrite rule or on the right side of the rewrite rule) and state S_3 in the FSA. You may want to draw a few trees to convince yourself that these describe the same language (namely, an *a*, followed by a *b*, followed by zero to infinity *c*s, followed by a *d*).

- **CFG**:

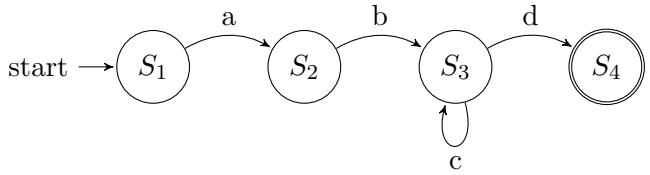
  S → *a* X

  X → *b* Y

  Y → *c* Y

  Y → *d*

- **FSA**:



2.) (5 points) Translate the following CFG into an FSA. (I would suggest contrasting the CFG with the example above. What is the big difference between these two? How can you modify the above FSA to capture this difference? Again, drawing trees will help immensely here.)

S → *a* X

X → *b* Y

Y → *c* Y

Y → *c* Z

Z → *d*

3.) (5 points) Translate the following CFG into an FSA. Is there any relation between this FSA and the one you wrote before? What does this tell you about the non-terminal nodes `X, Y, Z`?

S → X *d*

X → X *c*

X → Y *c*

Y → Z *b*

Z → *a*

Limitations of Context-Free Grammars

The following Context-Free Grammar is a subset of English. It can generate two grammatical sentences and two ungrammatical sentences.

`S` → `NP VP`

`NP` → *the key*

`NP` → *the keys*

`VP` → *is old*

`VP` → *are old*

4.) (5 points) Write out the four sentences that this grammar generates, and mark with an asterisk (*) which sentences are ungrammatical in Standard American English.

5.) (10 points) Find a way to change the set of rewrite rules in this grammar so that it only generates the grammatical sentences. (Hint: The most obvious solution to this puzzle has two different S nodes. Please do not posit agreement rules, transformations, or selectional restrictions in this exercise!)

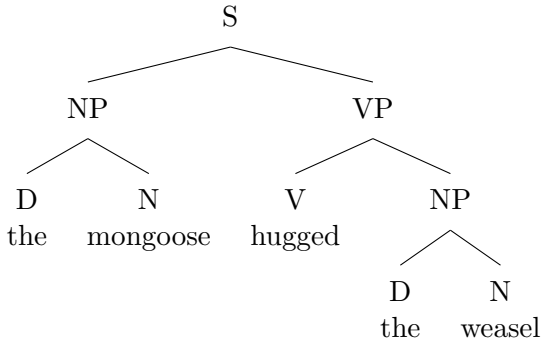
In *Syntactic Structures*, Chomsky argued that FSA were insufficient for modeling human language. This is because, in order to capture long-distance dependencies (*either* ... *or* ...; *if* ... *then* ...), the FSA would need to duplicate large swaths of the Finite State Automata. There was an important generalization that was missed in how an FSA would accomodate such sentences. Namely, the structure that could be sandwiched in between *either* and *or* was exactly the same as the structure that could be sandwiched in between *if* and *then*.

6.) (5 points) The solution that you provided to this puzzle likely missed a generalization of some kind. What generalization did you miss? Hint: Remember that non-terminal node labels are arbitrary in CFGs.

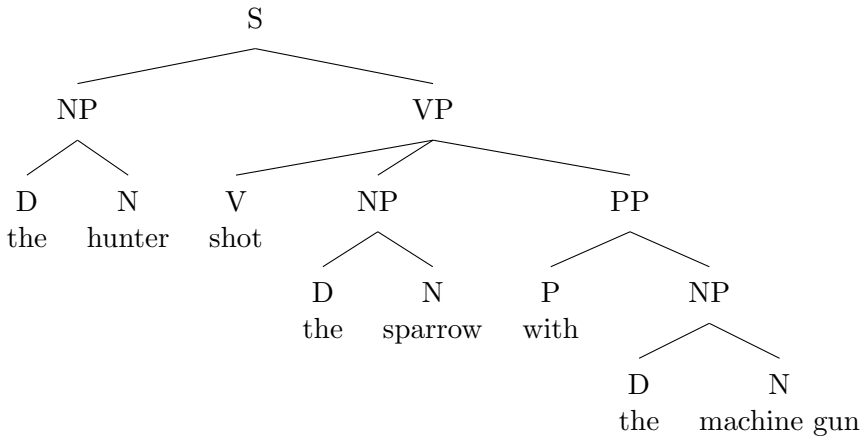
Ambiguity

As we discussed in class, linguists usually aren't interested in CFGs as mathematical objects for describing strings. Instead, we usually attribute some theoretical significance to the non-terminal nodes (i.e., constituents). We saw that constituency tests were a good diagnostic for determining which words constituted a phrase, and that often these phrases were semantically related. For instance, sentences typically cleave into a noun phrase and a verb phrase. The noun phrase (NP) consists of a noun and its modifiers, and the verb phrase (VP) consists of verb and its modifiers. Thus, all words in an NP are semantically related, and all words in a VP are semantically related.

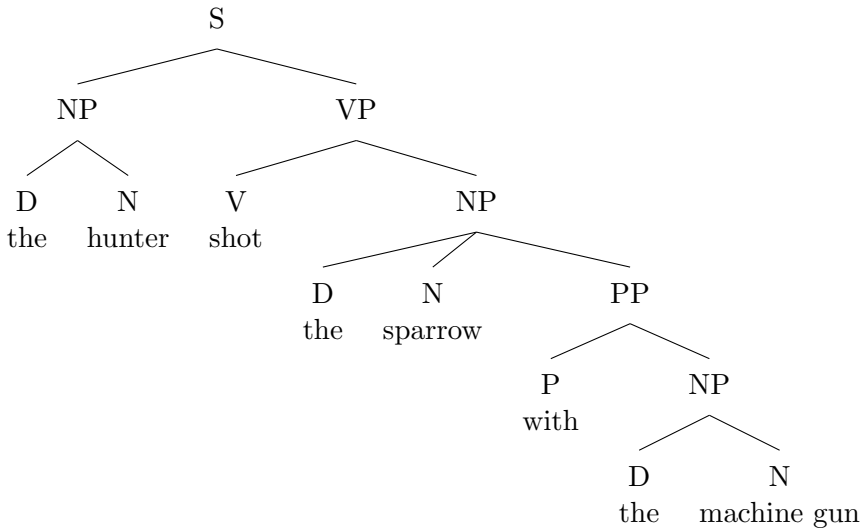
(1) The mongoose hugged the weasel.



We often find that there are sentences with multiple analyses available. These different analyses may lead to different meanings. For instance, the sentence *The hunter shot the sparrow with the rifle* can either have this structure, with an unsurprising meaning (i.e., the shooting was with the rifle):



. . . or this structure, with a more amusing meaning (i.e., the sparrow was also armed):

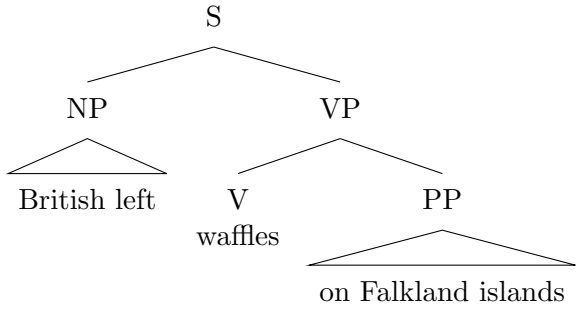


We can paraphrase these two meanings a number of different ways. For instance, the unsurprising

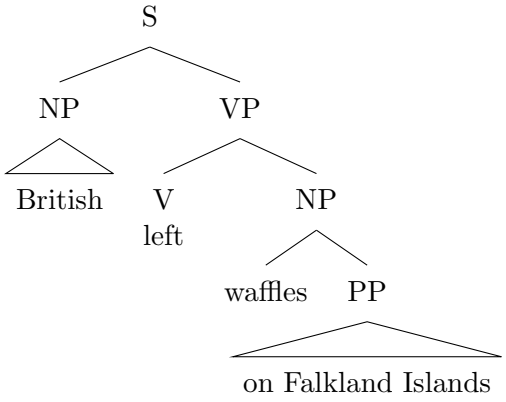
meaning could be paraphrased as *Using a rifle, the hunter shot the sparrow.* The unusual meaning could be paraphrased as *The hunter shot the sparrow that was holding a rifle.*

7.) (20 points) Below are four pairs of trees for newspaper headlines. Newspaper headlines are often are quite telegraphic, which can lead to some bizarre ambiguities that professors of Intro to Linguistics find amusing. For each pair, one tree represents the likely intended meaning, and the other tree represents the likely unintended meaning. Tell me which tree generates the intended meaning, and give a paraphrase for both meanings.[1]

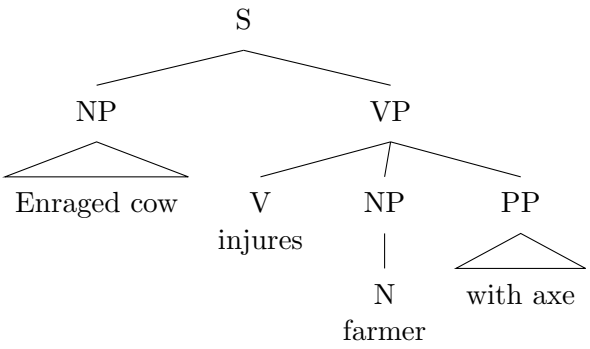
1. A.)



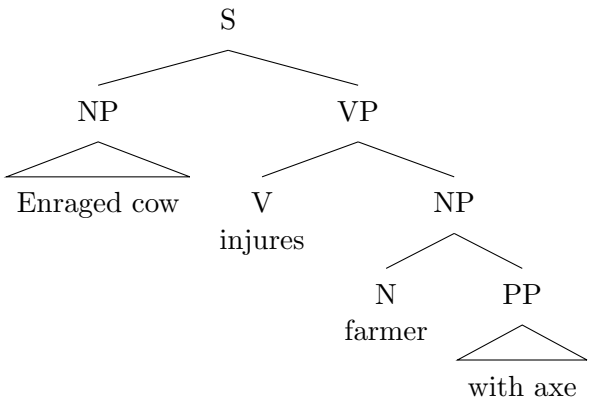
B.)



2. A.)

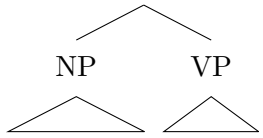


B.)

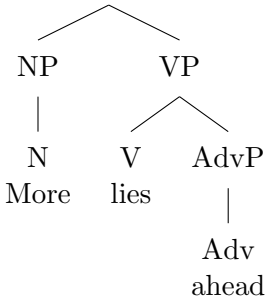


3. Reagan wins on budget, but ...

A.)

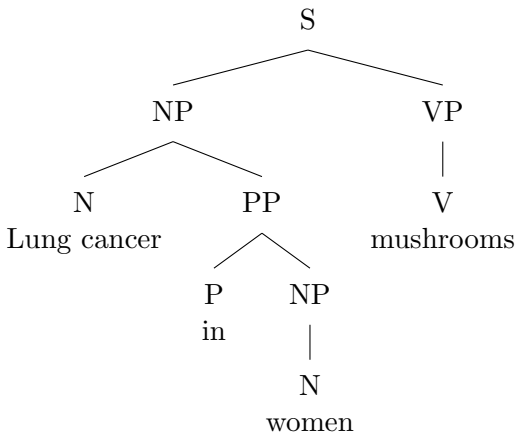


B.)

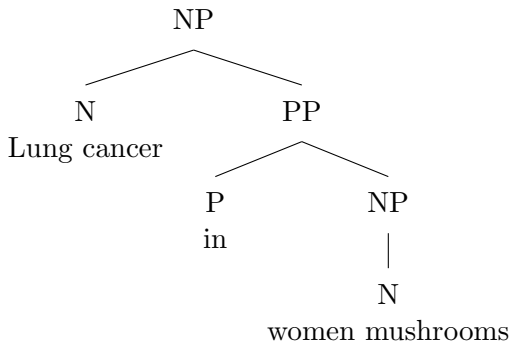


[1]These are taken from `http://www.departments.bucknell.edu/linguistics/synhead.html`

4. A.)



B.)



Selectional Restrictions

As we discussed in class, verbs in English (and potentially all languages) are quite picky about who they let into their verb phrase. We demonstrated this with the following examples. The notation $[_{\text{VP}} \ldots]$ is shorthand for marking off what occurs in a verb phrase, and $[_{\text{PP}} \ldots]$ for what occurs in a prepositional phrase, and so on.

(2) a. The armadillo $[_{\text{VP}}$ ate.]
b. The armadillo $[_{\text{VP}}$ ate a pomegranate.]

(3) a. *The armadillo $[_{\text{VP}}$ devoured.]
b. The aramdillo $[_{\text{VP}}$ devoured a pomegranate.]

(4) a. The armadillo $[_{\text{VP}}$ dined.]
b. *The armadillo $[_{\text{VP}}$ dined a pomegranate.]

(5) a. The armadillo $[_{\text{VP}}$ depended $[_{\text{PP}}$ on $[_{\text{NP}}$ his friend the marten]]].
b. *The armadillo $[_{\text{VP}}$ depended $[_{\text{PP}}$ in $[_{\text{NP}}$ his friend the marten]]].
c. *The armadillo $[_{\text{VP}}$ depended $[_{\text{PP}}$ at $[_{\text{NP}}$ his friend the marten]]].
d. *The armadillo $[_{\text{VP}}$ depended $[_{\text{PP}}$ with $[_{\text{NP}}$ his friend the marten]]].
e. *The armadillo $[_{\text{VP}}$ depended $[_{\text{PP}}$ on]]

There are many technical ways of describing the selectional restrictions of verbs. For our purposes, we will just say:

- *eat* may have an object NP.
- *devour* must have an object NP.
- *dined* must not have an object NP.
- *depend* must have a PP with the head *on*;
  on must have an NP.

8.) (5 points) Describe the selectional restrictions of the verbs *sleep*, *laugh*, and *die*. Do you notice any relation between the verb and the noun that it can select?

(6) a. The groundhog slept a long, restful sleep.
b. *The groundhog slept a long, restful nap.
c. *The groundhog slept a fluffy pillow.
d. *The groundhog slept a comfortable bed.
e. The groundhog slept.

(7) a. The jackal laughed a hearty, joyous laugh.
b. *The jackal laughed a hearty, joyous chuckle.
c. *The jackal laughed a funny joke.
d. *The jackal laughed a comedy routine.
e. The jackal laughed.

(8) a. The inchworm died a sudden, tragic death.
b. *The inchworm died a sudden, tragic demise.
c. *The inchworm died a horrible accident.
d. *The inchworm died a stomping shoe.
e. The inchworm died.

Next, let's look at the selectional restrictions of the verbs *want*, *consider*, and *desire*. All three of these verbs take clauses as objects. However, they place different demands on whether the clause has the complementizer *that*, and whether the clause is finite (i.e., has a present tense marker on the verb, *-s*) or non-finite (i.e., uses *to* before the verb).

(9) a. The cobbler wants [$_{\text{clause}}$ the customer to pay].
b. The cobbler wants [$_{\text{clause}}$ that the customer pays].[2]
c. *The cobbler wants [$_{\text{clause}}$ the customer pays].
d. *The cobbler wants [$_{\text{clause}}$ that the customer to pay].

(10) a. The knight considers [$_{\text{clause}}$ the dragon to be terrifying].
b. *The knight considers [$_{\text{clause}}$ that the dragon is terrifying].
c. *The knight considers [$_{\text{clause}}$ the dragon is terrifying].
d. *The knight considers [$_{\text{clause}}$ that the dragon to be terrifying].

(11) a. *The director desires [$_{\text{clause}}$ the actress to pose].
b. The director desires [$_{\text{clause}}$ that the actress poses].
c. *The director desires [$_{\text{clause}}$ the actress poses].
d. *The director desires [$_{\text{clause}}$ that the actress to pose].

9.) (5 points) State the selectional requirements for each of the three verbs *want*, *consider* and *desire*, specifying the selectional restrictions each verb has on the complementizer of the embedded clause and the finiteness of the embedded clause.

There is a bigger generalization here – when there is a complementizer *that* in the embedded clause, the verb must be finite. When there is no complementizer *that*, the verb must be non-finite. In order to more accurately capture this generalization, syntacticians have posited that clauses are built out of two phrases – a CP (complementizer phrase) and a TP (tense phrase). The head of the CP is either *that*, or a null complementizer, which I mark here as $\emptyset$. The head of the TP is either

[2]Notice that Standard American English also makes available a subjunctive form of the embedded verb for all the (b) examples, ie.., *the cobbler wants that the customer pay.* We are setting this construction aside.

the finiteness on the verb (which I will mark here as [FIN]), or the non-finite marker *to*. Thus, the structure of these sentences is as below:

(12) a. The cobbler wants [$_{CP}$ ∅ [$_{TP}$ the customer to pay]].
b. The cobbler wants [$_{CP}$ that [$_{TP}$ the customer [FIN] pays]].
c. *The cobbler wants [$_{CP}$ ∅ [$_{TP}$ the customer [FIN] pays]].
d. *The cobbler wants [$_{CP}$ that [$_{TP}$ the customer to pay]].

(13) a. The knight considers [$_{CP}$ ∅ [$_{TP}$ the dragon to be terrifying]].
b. *The knight considers [$_{CP}$ that [$_{TP}$ the dragon [FIN] is terrifying]].
c. *The knight considers [$_{CP}$ ∅ [$_{TP}$ the dragon [FIN] is terrifying]].
d. *The knight considers [$_{CP}$ that [$_{TP}$ the dragon to be terrifying]].

(14) a. *The director desires [$_{CP}$ ∅ [$_{TP}$ the actress to pose]].
b. The director desires [$_{CP}$ that [$_{TP}$ the actress [FIN] poses]].
c. *The director desires [$_{CP}$ ∅ [$_{TP}$ the actress FIN poses]].
d. *The director desires [$_{CP}$ that [$_{TP}$ the actress to pose]].

This permits a more straightforward description of the selectional restrictions of these verbs. With this analysis, we can specify whether the verb wants *that* in the CP next to it, or whether it wants ∅ in the CP next to it. Then, we can write a selectional restriction for *that* and a selectianal restriction for ∅ in order to capture the more general pattern between complementizers and finiteness.

10.) (10 points) With this analysis in mind, rewrite your selectional restrictions for *want*, *consider*, *desire*, and write the selectional restrictions for the complementizer *that* and the null complementizer ∅.

Movement

In class, we noticed a paradox: a violation of a verb's selectional restrictions could be forgiven in some circumstances. For instance, we see that *devour* requires an NP object next to it. However, *devour* is happy if there is a question phrase that is interpreted as its object earlier in the sentence.

(15) a. The monster will devour the victim.
b. *The monster will devour.
c. Which victim will the monster devour?

In generative syntax, the solution to this puzzle is to posit a **movement** relation. We say that the grammar of English permits the noun phrase *which victim* to play two roles. As its first role, *which victim* appears at the beginning of the sentence to mark it as a question. As its second role, it satisfies the selectional requirement of the verb *devour*. Because *which victim* is playing these two roles, we understand this sentence as a question about who is being devoured by the monster. Intuitively, we can think of movement theory as saying that *which victim* starts off next to *devour*, and then moves up to the beginning of the sentence. Alternatively, we can think that there is a silent *which victim* next to *devour* that is syntactically related to the non-silent *which building* earlier in the sentence. Either way, let's represent the movement dependency like this:

(16) Which victim will the monster devour <which victim>?

Where < and > can be understood as demarcating the "first" position of *which victim*, or the silent version of *which victim*.

Now, let's examine the verb *please*.

(17) a. The tasty victim pleased the monster.
b. *The tasty victim pleased.

11.) (5 points) Describe the selectional restrictions of *please*, given the data above.

Now, examine the following sentences. We again see a paradox – the verb *please* appears to not have its selectional requirements met. However, the sentences are grammatical.

(18) a. The monster is easy to please.
b. The monster is tough to please.
c. The monster is fun to please.

12.) (10 points) Explain why these sentences are grammatical using the theory of movement. In your answer, explain which noun phrase moved, and where it moved from. (Hint: In order to understand where a the noun phrase moved from, think of who is doing the pleasing, and who is being pleased, and compare it with (17).)

Now, turn your attention to the following sentences:

(19) a. The selfless victim was eager to please the monster.
b. The selfless victim was happy to please the monster.
c. *The selfless victim was easy to please the monster.
d. *The selfless victim was tough to please the monster.
e. *The selfless victim was fun to please the monster.
f. The monster is ready to eat.
g. The monster is ready to eat the victim.

11.) (5 points) Describe what this data shows us about the movement phenomenon under consideration. In your answer, include which word "decides" whether movement takes place. Also, describe how the movement changes the meaning of the sentence. In your answer, describe the different roles that *the monster* plays in (19-f) and (19-g). How does this compare to verb selectional restrictions (e.g., *devour* vs. *dine* vs. *eat*)?