Intro to Computational Cognitive Science: HW5

This homework is due on Friday, October 19, at 6:00 PM. Please write your answers in a PDF named CompCogSci_HW5_grammar_FIRST_LAST.pdf (except replace FIRST and LAST with your first and last names) and submit the PDF to the following Dropbox folder:
https://www.dropbox.com/request/lJn2HPtCyAwICsF8wSKa

A. Context-free grammar

Here is a fragment of English grammar:
S → NP VP
NP → Det N
VP → V NP
Det → the
Det → my
N → walrus
N → manatee
N → whale
V → saw
V → visited

Question 1: Example sentence. Write a sentence that this grammar generates.

Question 2: Intransitive verbs. Modify the grammar to include the following verbs: *smiled*, *chuckled*. Note that these verbs are intransitive, meaning that they require a subject but no direct object, while *saw* and *visited* are transitive verbs, meaning that they require a subject and a direct object. Therefore, your modified grammar should generate the sentences in (1) (as well as many other similar sentences), but it should not generate the sentences in (2) (the asterisks in (2) indicate that these sentences are ungrammatical):

(1)   a. my walrus saw the manatee

          b. the manatee visited the whale

          c. my walrus smiled

          d. the whale chuckled

(2)   a. * my walrus smiled the manatee

          b. * the manatee chuckled the whale

          c. * my walrus saw

          d. * the whale visited

In your answer for this problem, please include the entire modified CFG, even including the rules that you did not have to modify.

Question 3: Subject-verb agreement. Modify the original grammar (the one without *smiled* and *chuckled*) so that it allows singular and plural nouns—that is, the nouns it includes should be *walrus*, *walruses*, *manatee*, *manatees*, *whale*, and *whales*. In addition, your modified grammar should use present-tense verbs instead of past-tense verbs—that is, the verbs it includes should be *see*, *sees*, *visit*, and *visits*.

    To make this work properly, you will have to make sure that the subjects and verbs agree in each sentence. That is, if the subject of the sentence is singular, then the verb of that sentence should also be singular; while if the subject is plural, the verb should also be plural. This means that your modified grammar should generate the sentences in (3) (as well as many other similar sentences), but it should not generate the ungrammatical sentences in (4):

(3)   a. the walruses see the whale

         b. the walrus sees the whale

         c. the walrus sees the whales

         d. the manatees visit the walrus

(4)   a. * the walruses sees the whale

         b. * the walrus see the whale

         c. * the walrus see the whales

         d. * the manatees visits the walrus

As before, in your submission, please include the entire modified CFG, even including the rules that you did not have to modify.

B. Finite-state machine

Consider the following English suffixes:

- *-al*: Attaches to a noun N and creates an adjective meaning "relating to N." Example: *industrial* means "relating to industry."

- *-ize*: Attaches to an adjective A and creates a verb meaning "to make (something) A." Example: *industrialize* means "to make (something) industrial."

- *-ation*: Attaches to a verb V and creates a noun meaning "the process of Ving." Example: *industrialization* means "the process of industrializing."

Notice something interesting about these suffixes: You can apply them in a loop to create an endless cycle of increasingly long words:

- *industry* "industry"

- *industrial* "relating to industry"

- *industrialize* "to make (something) industrial"

- *industrialization* "the process of industrializing"

- *industrializational* "relating to industrialization"

- *industrializationalize* "to make (something) industrializational"

- *industrializationalization* "the process of industrializationalizing"

- *industrializationalizational* "relating to industrializationalization"

- ...

- *industrializationalizationalizationalizationalize* "to make (something) industrializationalizationalizationalizational"

- ...

Of course, the longer these words get, the harder they are to process, and the less likely you are to encounter them in the wild. However, much as we argued that the English language allows an unbounded number of center embeddings even though human processing capabilities struggle beyond two levels of center embedding, we will also say that English allows this process of suffixation to continue unboundedly to create arbitrarily long words.

Question 4: Morphology. Create a finite-state machine that generates all possible words in the endless loop illustrated above (*industry, industrial, industrialize*, etc.). Make sure that your finite-state machine does not generate any words that would be invalid in English, such as *industriation, industrializationize*, or *industrialal*. Also be careful to get the spelling right: You should generate *industry* and *industrial*, but not **industri* or **industryal*. Similarly, you should generate *industrialize* and *industrialization*, but not **industrializ* or **industrializeation*.

Note: The following website is useful for drawing finite-state machines:
http://madebyevan.com/fsm/

Alternately, you can draw it by hand and include a picture or scan in your submission. If you do this, just make sure that the drawing is clean and readable.

For this problem, feel free to include more than one letter on a single arc of the finite state machine (if you only include one letter per arc, the finite state machine will be much larger and clunkier). For example, the following two finite state machines are equivalent: