

HW7
Mike Roylance
roylance@uw.edu

Describe and discuss your work in a write-up file

I completed this assignment using Python with nltk. My solution is organized into the following files and directories:

Location	Description	Purpose
docs/		folder that contains all the documents needed for this assignment: grammar.cfg and coref_sentences.txt.
docs/grammar.cfg		file that contains the context free grammar for the sentences to be parsed
docs/coref_sentences.txt		file that contains tuples of sentences, with the first containing noun phrases as antecedents and the second with pronouns to be resolved
source/		folder that contains all the source code
source/main.py		entry point script that reads in the files from the user and prints out the result to the console.
source/tests.py		tests around functionality
source/hobbs.py		class that takes in a couple of parsed sentences (as nltk trees) and runs the hobbs algorithm to determine the pronouns and antecedents.
source/utils.py		utility module to load in the nltk trees and return a tuple
source/traverseTree.py		this is a tree structure that has linking relationships set to traverse the syntactical trees for the different X and p paths.
source/rules.py		class that calculates all the rules associated with selected a pronoun or a potential antecedent. also calculates if two itemIndexes are in agreement with each other (and therefore acceptable).
hw7.cmd		command file used by condor. this calls hw7.sh with the parameters of docs/grammar.cfg docs/coref_sentences.txt and results
hw7.sh		file to handle calling the Python file main.py with specific parameters

Include problems you came across and how (or if) you were able to solve them, any insights, special features, and what you learned. Give examples if possible.

The first problem I encountered was how I approached this homework assignment - I thought that I could rely just on nltk to traverse the tree structures for the different X (NP and S). My first stab, which I got working albeit in a different order, just started at the beginning of each sentence and worked it's way from top to bottom. This was fine, until I realized the order in which X is compared is important, especially when starting from the second sentence's pronoun. I fixed this by creating a new class, `traversableTree.py`, that handles the relationship to allow for moving up and down the tree.

The class, `Hobbs.py`, takes in two nltk Trees that represent the first and second sentence. It also instantiates a singleton `Rules.py` class that contains the rules for identifying antecedents. It loops through the different lists and selectively filters them based on the criteria.

The `rules.py` class identifies what type of parts of speech are acceptable for pronouns (PRP, PossPro) and potential antecedents (S, SBAR, NP). It also contains specific dictionaries about what parts of speech are considered plural (NNS and them/their). It also encodes gender parts of speech, although in this case

For gender, I didn't see any male or female specific words in the sentences or grammar file. While I do instantiate "female" and "male" dictionaries, my rules only check the male one because it is a clone of the female one. For gender, I'm only check if it is "u" (unknown) or "m" (male).

Plurality, however, I feel works fine. I give the plurality dictionary the part of speech (NNS), but it also has specific pronouns that are plural as well (they, them).

I was going to implement this with just the grammar file, but I like the approach of having it in the `Rules.py` class because there are certain rules that cannot be coded in the grammar file that must be coded in `Rules.py` - such as the correct part of speech of antecedents or the function to determine if a given antecedent actually comes before the pronoun to be compared with.

My file is printing out similarly to how the `simple_example.txt` is printing out - however the instructions on the website indicate that if a programmatic approach is taken then to just print out all the antecedents. I did go a bit further and specified agreement as well. I have an extra space to indicate when the file is finished processing.

Starting on line #50 in `Hobbs.py`, I manually list out the steps taken in code as specified on the web site.

As specified on the discussion board, I added "X:" in front of every new X.

I have also **bolded** the results, my program is printing all the “Accept” with “Correct”, however I am updating “Correct” to include “Incorrect” if the result is inaccurate.

Here are the results:

	Corrects	Rejects	Incorrect Correct
Test 1	1	1	0
Test 2	1	1	0
Test 3	1	1	0
Test 4	2	4	2
Test 5	1	2	1

They (S (NP (Nom (NNS Scientists))) (VP (VBD rescued) (NP (Det a) (Nom (Nom (NN mouse)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .)) (S (NP (PRP They)) (VP (VP (VP (VBD published) (NP (Det the) (Nom (NN research)))) (ADVP (Adv today))) (ADVP (Adv online))) (PUNC .))

(S (NP (PRP They)) (VP (VP (VP (VBD published) (NP (Det the) (Nom (NN research)))) (ADVP (Adv today))) (ADVP (Adv online))) (PUNC .))

(S (NP (Nom (NNS Scientists))) (VP (VBD rescued) (NP (Det a) (Nom (Nom (NN mouse)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

(NP (Nom (NNS Scientists)))

Accept

Correct

(NP (Det a) (Nom (Nom (NN mouse)) (Nom (ADJP (Adj immune)) (Nom (NN system))))

Reject - number

They (S (NP (Det The) (Nom (ADJP (Adj immune)) (Nom (NN response)))) (VP (AUX (AuxZ is)) (VP (VBN alerted) (PP (Prep by) (NP (Nom (ADJP (Adj dendritic)) (Nom (NNS cells)))))) (PUNC .)) (S (NP (PRP They)) (VP (VP (VBP capture) (NP (Nom (ADJP (Adj infected)) (Nom (NNS cells)))) (Conj and) (VP (VBP display) (NP (Nom (Nom (NNS fragments)) (PP (Prep of) (NP (Det the) (Nom (NN pathogen)))))) (PUNC .))

(S (NP (PRP They)) (VP (VP (VBP capture) (NP (Nom (ADJP (Adj infected)) (Nom (NNS cells)))) (Conj and) (VP (VBP display) (NP (Nom (Nom (NNS fragments)) (PP (Prep of) (NP (Det the) (Nom (NN pathogen)))))) (PUNC .))

(S (NP (Det The) (Nom (ADJP (Adj immune)) (Nom (NN response)))) (VP (AUX (AuxZ is)) (VP (VBN alerted) (PP (Prep by) (NP (Nom (ADJP (Adj dendritic)) (Nom (NNS cells)))))) (PUNC .)) (NP (Det The) (Nom (ADJP (Adj immune)) (Nom (NN response))))

Reject - number

(NP (Nom (ADJP (Adj dendritic)) (Nom (NNS cells))))

Accept**Correct**

They (S (NP (Nom (NNS Scientists))) (VP (VBD rescued) (NP (Det a) (Nom (Nom (NN mouse)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .)) (S (NP (PRP They)) (VP (VBD restored) (NP (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

(S (NP (PRP They)) (VP (VBD restored) (NP (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

(S (NP (Nom (NNS Scientists))) (VP (VBD rescued) (NP (Det a) (Nom (Nom (NN mouse)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

(NP (Nom (NNS Scientists)))

Accept**Correct**

(NP (Det a) (Nom (Nom (NN mouse)) (Nom (ADJP (Adj immune)) (Nom (NN system))))

Reject - number

They (S (NP (Nom (NNS Scientists))) (VP (VBD restored) (NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .)) (S (NP (PRP They)) (VP (VBD injected) (NP (PRP them)) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj live)) (Nom (NN vaccine)))))) (PUNC .))

(S (NP (PRP They)) (VP (VBD injected) (NP (PRP them)) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj live)) (Nom (NN vaccine)))))) (PUNC .))

(S (NP (Nom (NNS Scientists))) (VP (VBD restored) (NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

(NP (Nom (NNS Scientists)))

Accept**Correct**

(NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system))))))

Reject - number

(NP (Nom (NNS mice)))

Accept**Incorrect**

(NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system))))

Reject - number

them (S (NP (Nom (NNS Scientists))) (VP (VBD restored) (NP (Nom (Nom (Nom (NN immunity))

(PP (Prep in) (NP (Nom (NNS mice)))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .)) (S (NP (PRP They)) (VP (VBD injected) (NP (PRP them)) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj live)) (Nom (NN vaccine)))))) (PUNC .))

(S (NP (PRP They)) (VP (VBD injected) (NP (PRP them)) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj live)) (Nom (NN vaccine)))))) (PUNC .))

(S (NP (Nom (NNS Scientists)) (VP (VBD restored) (NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

(NP (Nom (NNS Scientists)))

Accept

Incorrect

(NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

Reject - number

(NP (Nom (NNS mice)))

Accept

Correct

(NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))) (PUNC .))

Reject - number

They (S (NP (Nom (NNS Scientists)) (VP (VBD restored) (NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .)) (S (NP (PRP They)) (VP (AUX (AuxZ were)) (VP (VBN compromised) (PP (Prep by) (NP (Nom (NN sepsis)))))) (PUNC .)) (S (NP (PRP They)) (VP (AUX (AuxZ were)) (VP (VBN compromised) (PP (Prep by) (NP (Nom (NN sepsis)))))) (PUNC .))

(S (NP (Nom (NNS Scientists)) (VP (VBD restored) (NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

(NP (Nom (NNS Scientists)))

Accept

Incorrect

(NP (Nom (Nom (Nom (NN immunity)) (PP (Prep in) (NP (Nom (NNS mice)))) (PP (Prep with) (NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))))) (PUNC .))

Reject - number

(NP (Nom (NNS mice)))

Accept

Correct

(NP (Det a) (Nom (ADJP (Adj weak)) (Nom (ADJP (Adj immune)) (Nom (NN system)))) (PUNC .))

Reject - number