Problem Set 9 (and last!)

Assigned: May 1
Due: May 8
Note: Because of the end of the semester, this will be only be accepted late until the time of the final exam, 12 PM May 11.

Problem 1 (6 points out of 10)

Consider again the grammar used in problem set 8.

```
S → NP VP | S Conj S
NP → { Adj } Noun | NP Conj NP
VP → Verb { NP } | VP Conj VP
```

It is convenient to think of sentences as ending with a period with the part of speech `Period`.

The corresponding 3-gram model would require the following states:

S0: `*S* *S*`
S1: `*S* Adj`
S2: `*S* Noun`
S3: `Adj Noun`
S4: `Conj Adj`
S5: `Conj Noun`
S6: `Conj Verb`
S7: `Noun Conj`
S8: `Noun Verb`
S9: `Noun Period`
S10: `Verb Adj`
S11: `Verb Conj`
S12: `Verb Noun`
S13: `Verb Period`

Let us write P(SX→SY) to mean, "the probability that the kth state is SY given that the $k-1$st state is SX." We posit the following non-zero transition probabilities.

P(S0 → S1) = 0.6	P(S0 → S2) = 0.4		
P(S1 → S3) = 1			
P(S2 → S7) = 0.3	P(S2 → S8) = 0.7		
P(S3 → S7) = 0.2	P(S3 → S8) = 0.5	P(S3 → S9) = 0.3	
P(S4 → S3) = 1			
P(S5 → S7) = 0.1	P(S5 → S8) = 0.6	P(S5 → S9) = 0.3	
P(S6 → S10) = 0.2	P(S6 → S11) = 0.1	P(S6 → S12) = 0.3	P(S6 → S13) = 0.4
P(S7 → S4) = 0.2	P(S7 → S5) = 0.5	P(S7 → S6) = 0.3	
P(S8 → S10) = 0.1	P(S8 → S11) = 0.2	P(S8 → S12) = 0.3	P(S8 → S13) = 0.4
P(S10 → S3) = 1			
P(S11 → S4) = 0.1	P(S11 → S5) = 0.3	P(S11 → S6) = 0.6	
P(S12 → S7) = 0.3	P(S12 → S9) = 0.7		

The base probabilities of the tags are:
P(`Adj`) = 0.1. P(`Conj`) = 0.1. P(`Noun`) = 0.4. P(`Verb`) = 0.3. P(`Period`) = 0.1

The probabilities of the tags given the words, P(T|W) are given in this table

	and	boots	brown	own	people	socks	.
`Adj`	0	0	0.7	0.4	0	0	0
`Conj`	1	0	0	0	0	0	0
`Noun`	0	0.9	0.2	0.	0.9	0.8	0
`Verb`	0	0.1	0.1	0.6	0.1	0.2	0
`Period`	0	0	0	0	0	0	1

A. Consider the sentence "People own boots and brown socks." What are the three possible taggings with non-zero probability? (Hint: work from left to right in the sentence; for the first few transitions, there is actually only one choice.)

B. Which is the most probable of those taggings? (Hint: since the first few tags are the same in all three, all you have to compute is the product over the states where they differ.)

C. Find a sentence consisting of two words followed by a period and an associated tagging such that the tagging has non-zero probability in the above k-gram model but is not a possible tagging for a complete sentence i.e. a tree rooted by `S` in the CFG above.

Problem 2 (4 points out of 10)

The purpose of this assignment is to illustrate some of the difficulties in interpreting natural text. Do your best with it. I think this should not be too hard, but if you find that you're having trouble with one part or another because the English is too difficult, make a note of that, and I'll make allowances in grading. In any case, grading here will be lenient; I want you to engage with these issues, not to come up with the correct solution.

The two sentences quoted below are quotations from
"I Was General Counsel of the N.S.A. America Has a Problem With Secrets."
Glenn S. Gerstell, *New York Times,* April 24, 2023

> 1. When Congress allocates funds to spy agencies, they are more likely to spend them on new spying techniques that might produce richer intelligence, rather than on protective measures that lower the risk of compromise.
>
> 2. From pharmaceutical companies to defense contractors working on the cutting edge of the digital revolution, private companies deploy technology in an effort to prevent theft of industrial secrets so that samples, models and blueprints don't walk out the door.

I am not expressing any opinion about the content of this opinion piece; merely using it as an example for issues in language interpretation.

A. What are the grammatically possible referents for "they"and "them" in sentence 1? What are the correct referents? How might a knowledge-based system decide on the correct referent?

B. In a dependency grammar, what is the path from "compromise" in sentence 1 to the root, "are"? (I'm not expecting a precise answer here — I haven't given you anything close to enough information — but think about it, and write something plausible.)

C. Find a compound noun in sentence 2; that is, a phrases consisting of two nouns in a row. Discuss how the meaning of the compound is related to the meanings of the two words. There are some examples of this in the notes on Compound Nouns on the course web page.

D. Find two figurative expressions in sentence 2 (admittedly, both are clichés).

E. WordNet `http://wordnetweb.princeton.edu/perl/webwn` is a valuable, manually constructed, online resource for information about words in English. (Versions of WordNet also exist for many

other languages.) How many meanings of "work" as an verb are listed? (I have to admit there are a couple I had never heard of.)

F. Optional, not to hand in: If this kind of thing interests you, look up the entry for "work" as a verb in the Oxford English Dictionary. To get to the OED, go to the NYU online catalog `http://bobcat.library.nyu.edu`; log in using your NetId; and search for "Oxford English Dictionary". The first result will be labelled as a Database. Click on that, and you will get to the online OED.