CS 485 CFG exercise, 10/8/2024 Name: _____

Use the grammar on the next page (and if you need sensible modifications or to introduce additional rules, note them). Assume there are reasonable POS tags (not included in the grammar). With them:

1. Draw a parse tree for:
   I fly between Philadelphia and Atlanta

2. Invent a new sentence that has a syntactic ambiguity. Draw it and two legitimate parse trees for it.

Sentences

The most common production rule for sentences is,

$$\text{S} \rightarrow \text{NP VP} \tag{8.27}$$

which accounts for simple sentences like *Abigail ate the kimchi* — as we will see, the direct object *the kimchi* is part of the verb phrase. But there are more complex forms of sentences as well:

S →ADVP NP VP	*Unfortunately Abigail ate the kimchi.*	(8.28)
S →S CC S	*Abigail ate the kimchi and Max had a burger.*	(8.29)
S →VP	*Eat the kimchi.*	(8.30)

where ADVP is an adverbial phrase (e.g., *unfortunately*, *very unfortunately*) and CC is a coordinating conjunction (e.g., *and*, *but*).[7]

Noun phrases

Noun phrases refer to entities, real or imaginary, physical or abstract: *Asha*, *the steamed dumpling*, *parts and labor*, *nobody*, and *the rise of revolutionary syndicalism in the early twentieth century*. Noun phrase productions include "bare" nouns, which may optionally follow determiners, as well as pronouns:

$$\text{NP} \rightarrow \text{NN} \mid \text{NNS} \mid \text{NNP} \mid \text{PRP} \tag{8.31}$$
$$\text{NP} \rightarrow \text{DET NN} \mid \text{DET NNS} \mid \text{DET NNP} \tag{8.32}$$

The part-of-speech tags NN, NNS, and NNP refer to singular, plural, and proper nouns; PRP refers to personal pronouns, and DET refers to determiners. The grammar also contains terminal productions from each of these tags, e.g., PRP → *I* | *you* | *we* | ....

Noun phrases may be modified by adjectival phrases (ADJP; e.g., *the small Russian dog*) and numbers (CD; e.g., *the five pastries*), each of which may optionally follow a determiner:

$$\text{NP} \rightarrow \text{ADJP NN} \mid \text{ADJP NNS} \mid \text{DET ADJP NN} \mid \text{DET ADJP NNS} \tag{8.33}$$
$$\text{NP} \rightarrow \text{CD NNS} \mid \text{DET CD NNS} \mid \ldots \tag{8.34}$$

[7] Notice that the grammar does not include the recursive production S → ADVP S. It may be helpful to think about why this production would cause the grammar to overgenerate.

Some noun phrases include multiple nouns, such as *the liberation movement* and *an antelope horn*, necessitating additional productions:

$$\text{NP} \rightarrow \text{NN NN} \mid \text{NN NNS} \mid \text{DET NN NN} \mid \ldots \tag{8.35}$$

These multiple noun constructions can be combined with adjectival phrases and cardinal numbers, leading to a large number of additional productions.

Recursive noun phrase productions include coordination, prepositional phrase attachment, subordinate clauses, and verb phrase adjuncts:

NP →NP CC NP	*e.g., the red and the black*	(8.36)
NP →NP PP	*e.g., the President of the Georgia Institute of Technology*	(8.37)
NP →NP SBAR	*e.g., the bicycle that I found outside*	(8.38)
NP →NP VP	*e.g., a bicycle made of titanium*	(8.39)

These recursive productions are a major source of ambiguity, because the VP and PP nonterminals can also generate NP children. Thus, the *the President of the Georgia Institute of Technology* can be derived in two ways, as can *a bicycle made of titanium found outside*.

Other contituents

The remaining constituents require far fewer productions. **Prepositional phrases** almost always consist of a preposition and a noun phrase,

PP → IN NP	*United States* ***of America***	(8.59)
PP → TO NP	*He gave his kimchi* ***to Abigail***	(8.60)

Similarly, complement clauses consist of a complementizer (usually a preposition, possibly null) and a sentence,

SBAR → IN S	*She said* ***that it was spicy***	(8.61)
SBAR → S	*She said* ***it was spicy***	(8.62)

Adverbial phrases are usually bare adverbs (ADVP → RB), with a few exceptions:

ADVP → RB RBR	*They went* ***considerably further***	(8.63)
ADVP → ADVP PP	*They went* ***considerably further than before***	(8.64)

The tag RBR is a comparative adverb.

Adjectival phrases extend beyond bare adjectives (ADJP → JJ) in a number of ways:

ADJP → RB JJ	*very hungry*	(8.65)
ADJP → RBR JJ	*more hungry*	(8.66)
ADJP → JJS JJ	*best possible*	(8.67)
ADJP → RB JJR	*even bigger*	(8.68)
ADJP → JJ CC JJ	*high and mighty*	(8.69)
ADJP → JJ JJ	*West German*	(8.70)
ADJP → RB VBN	*previously reported*	(8.71)

The tags JJR and JJS refer to comparative and superlative adjectives respectively.

All of these phrase types can be coordinated:

PP →PP CC PP	*on time and under budget*	(8.72)
ADVP →ADVP CC ADVP	*now and two years ago*	(8.73)
ADJP →ADJP CC ADJP	*quaint and rather deceptive*	(8.74)
SBAR →SBAR CC SBAR	*whether they want control or whether they want exports*	(8.75)

Verb phrases

Verb phrases describe actions, events, and states of being. The PTB tagset distinguishes several classes of verb inflections: base form (VB; *she likes to snack*), present-tense third-person singular (VBZ; *she snacks*), present tense but not third-person singular (VBP; *they snack*), past tense (VBD; *they snacked*), present participle (VBG; *they are snacking*), and past participle (VBN; *they had snacked*).[9] Each of these forms can constitute a verb phrase on its own:

$$\text{VP} \rightarrow \text{VB} \mid \text{VBZ} \mid \text{VBD} \mid \text{VBN} \mid \text{VBG} \mid \text{VBP} \tag{9.41}$$

More complex verb phrases can be formed by a number of recursive productions, including the use of coordination, modal verbs (MD; *she should snack*), and the infinitival *to* (TO):

VP → MD VP	*She* ***will snack***	[9.42]
VP → VBD VP	*She* ***had snacked***	[9.43]
VP → VBZ VP	*She* ***has been snacking***	[9.44]
VP → VBN VP	*She has* ***been snacking***	[9.45]
VP → TO VP	*She wants* ***to snack***	[9.46]
VP → VP CC VP	*She* ***buys and eats*** *many snacks*	[9.47]

Each of these productions uses recursion, with the VP non-terminal appearing in both the LHS and RHS. This enables the creation of complex verb phrases, such as *She will have wanted to have been snacking*.

Transitive verbs take noun phrases as direct objects, and ditransitive verbs take two direct objects:

VP → VBZ NP	*She* ***teaches algebra***	[9.48]
VP → VBG NP	*She has been* ***teaching algebra***	[9.49]
VP → VBD NP NP	*She* ***taught*** *her brother algebra*	[9.50]

These productions are *not* recursive, so a unique production is required for each verb part-of-speech. They also do not distinguish transitive from intransitive verbs, so the resulting grammar overgenerates examples like **She sleeps sushi* and **She learns Boyang algebra*. Sentences can also be direct objects:

VP → VBZ S	*Hunter* ***wants to eat the kimchi***	[9.51]
VP → VBZ SBAR	*Hunter* ***knows that Tristan ate the kimchi***	[9.52]

The first production overgenerates, licensing sentences like **Hunter sees Tristan eats the kimchi*. This problem could be addressed by designing a more specific set of sentence non-terminals, indicating whether the main verb can be conjugated.

The first production overgenerates, licensing sentences like **Asha sees Boyang eats the kimchi*. This problem could be addressed by designing a more specific set of sentence non-terminals, indicating whether the main verb can be conjugated.

[8] It bears emphasis the principles governing this tagset design are entirely English-specific: VBP is a meaningful category only because English morphology distinguishes third-person singular from all person-number combinations.

Verbs can also be modified by prepositional phrases and adverbial phrases:

VP → VBZ PP	*She* ***studies at night***	(8.52)
VP → VBZ ADVP	*She* ***studies intensively***	(8.53)
VP → ADVP VBG	*She is* ***not studying***	(8.54)

Again, because these productions are not recursive, the grammar must include productions for every verb part-of-speech.

A special set of verbs, known as **copula**, can take **predicative adjectives** as direct objects:

VP → VBZ ADJP	*She* ***is hungry***	(8.55)
VP → VBP ADJP	*Success* ***seems increasingly unlikely***	(8.56)

The PTB does not have a special non-terminal for copular verbs, so this production generates non-grammatical examples such as **She eats tall*.

Particles (PRT as a phrase; RP as a part-of-speech) work to create phrasal verbs:

VP → VB PRT	*She told them to* ***fuck off***	(8.57)
VP → VBD PRT NP	*They* ***gave up their ill-gotten gains***	(8.58)

As the second production shows, particle productions are required for all configurations of verb parts-of-speech and direct objects.