CS585 CFG exercise, 11/14/17

This is (most of) the English grammar in the Eisenstein reading.

Sentences

The most common production rule for sentences is,

$$S \rightarrow NP\ VP \qquad (8.27)$$

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:

$S \rightarrow ADVP\ NP\ VP$	*Unfortunately Abigail ate the kimchi.*	(8.28)
$S \rightarrow S\ Cc\ S$	*Abigail ate the kimchi and Max had a burger.*	(8.29)
$S \rightarrow VP$	*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:

$NP \rightarrow Nn \mid Nns \mid Nnp \mid Prp$	(8.31)
$NP \rightarrow Det\ Nn \mid Det\ Nns \mid Det\ Nnp$	(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., $Prp \rightarrow I \mid you \mid we \mid \ldots$.

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:

$NP \rightarrow ADJP\ Nn \mid ADJP\ Nns \mid Det\ ADJP\ Nn \mid Det\ ADJP\ Nns$	(8.33)
$NP \rightarrow Cd\ Nns \mid Det\ Cd\ Nns \mid \ldots$	(8.34)

[7]Notice that the grammar does not include the recursive production $S \rightarrow ADVP\ S$. 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:

$$NP \rightarrow Nn\ Nn \mid Nn\ Nns \mid Det\ Nn\ Nn \mid \ldots \qquad (8.35)$$

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:

$NP \rightarrow NP\ Cc\ NP$	*e.g., the red and the black*	(8.36)
$NP \rightarrow NP\ PP$	*e.g., the President of the Georgia Institute of Technology*	(8.37)
$NP \rightarrow NP\ SBAR$	*e.g., the bicycle that I found outside*	(8.38)
$NP \rightarrow NP\ VP$	*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,

$PP \rightarrow In\ NP$	*United States of America*	(8.59)
$PP \rightarrow To\ NP$	*He gave his kimchi to Abigail*	(8.60)

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

$SBAR \rightarrow In\ S$	*She said that it was spicy*	(8.61)
$SBAR \rightarrow S$	*She said it was spicy*	(8.62)

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

$ADVP \rightarrow Rb\ Rbr$	*They went considerably further*	(8.63)
$ADVP \rightarrow ADVP\ PP$	*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:

$ADJP \rightarrow Rb\ Jj$	*very hungry*	(8.65)
$ADJP \rightarrow Rbr\ Jj$	*more hungry*	(8.66)
$ADJP \rightarrow Jjs\ Jj$	*best possible*	(8.67)
$ADJP \rightarrow Rb\ Jjr$	*even bigger*	(8.68)
$ADJP \rightarrow Jj\ Cc\ Jj$	*high and mighty*	(8.69)
$ADJP \rightarrow Jj\ Jj$	*West German*	(8.70)
$ADJP \rightarrow Rb\ Vbn$	*previously reported*	(8.71)

The tags Jjr and Jjs refer to comparative and superlative adjectives respectively.

All of these phrase types can be coordinated:

$PP \rightarrow PP\ Cc\ PP$	*on time and under budget*	(8.72)
$ADVP \rightarrow ADVP\ Cc\ ADVP$	*now and two years ago*	(8.73)
$ADJP \rightarrow ADJP\ Cc\ ADJP$	*quaint and rather deceptive*	(8.74)
$SBAR \rightarrow SBAR\ Cc\ SBAR$	*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**).[8] Each of these forms can constitute a verb phrase on its own:

$$VP \rightarrow Vb \mid Vbz \mid Vbd \mid Vbn \mid Vbg \mid Vbp \qquad (8.40)$$

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 infitival *to* (To):

$VP \rightarrow Md\ VP$	*She will snack*	(8.41)
$VP \rightarrow Vbd\ VP$	*She had snacked*	(8.42)
$VP \rightarrow Vbz\ VP$	*She has been snacking*	(8.43)
$VP \rightarrow Vbn\ VP$	*She has been snacking*	(8.44)
$VP \rightarrow To\ VP$	*She wants to snack*	(8.45)
$VP \rightarrow VP\ VP$	*She buys and eats many snacks*	(8.46)

Each of these productions uses recursion, with VP appearing on the right-hand side. This enables the creation of very 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:

$VP \rightarrow Vbz\ NP$	*She teaches algebra*	(8.47)
$VP \rightarrow Vbg\ NP$	*She has been teaching algebra*	(8.48)
$VP \rightarrow Vbd\ NP\ NP$	*She taught her brother algebra*	(8.49)

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:

$VP \rightarrow Vbz\ S$	*Asha wants to eat the kimchi*	(8.50)
$VP \rightarrow Vbz\ SBAR$	*Asha knows that Boyang eats the kimchi*	(8.51)

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 nonterminals, indicating whether the main verb can be conjugated.

[8]It bears emphasis that 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:

$VP \rightarrow Vbz\ PP$	*She studies at night*	(8.52)
$VP \rightarrow Vbz\ ADVP$	*She studies intensively*	(8.53)
$VP \rightarrow ADVP\ Vbg$	*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:

$VP \rightarrow Vbz\ ADJP$	*She is hungry*	(8.55)
$VP \rightarrow Vbp\ ADJP$	*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:

$VP \rightarrow Vb\ PRT$	*She told them to fuck off*	(8.57)
$VP \rightarrow Vbd\ PRT\ NP$	*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.

Using this grammar, and reasonable POS tags, please:

1. Draw parse trees for these phrases:
   - from Denver
   - redeye flights
   - I need to fly between Philadelphia and Atlanta.

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