

# A pretty simple English CFG This is (much of) the English grammar in Eisenstein, ch. 9.

**Sentences**

The most common production rule for sentences is,

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

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:

$$S \rightarrow ADVP NP VP \quad \text{Unfortunately Abigail ate the kimchi.} \quad (8.28)$$

$$S \rightarrow S CC S \quad \text{Abigail ate the kimchi and Max had a burger.} \quad (8.29)$$

$$S \rightarrow VP \quad \text{Eat the kimchi.} \quad (8.30)$$

where ADVP is an adverbial phrase (e.g., *unfortunately, very unfortunately*) and CC is a coordinating conjunction (e.g., *and, but*).<sup>7</sup>

**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:

$$NP \rightarrow NN | NNS | NNP | PRP \quad (8.31)$$

$$NP \rightarrow DET NN | DET NNS | DET NNP \quad (8.32)$$

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.,  $PRP \rightarrow I | you | we | \dots$

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:

$$NP \rightarrow ADJP NN | ADJP NNS | DET ADJP NN | DET ADJP NNS \quad (8.33)$$

$$NP \rightarrow CD NNS | DET CD NNS | \dots \quad (8.34)$$

<sup>7</sup>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:

$$NP \rightarrow NN NN | NN NNS | DET NN NN | \dots \quad (8.35)$$

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:

$$NP \rightarrow NP CC NP \quad \text{e.g., the red and the black} \quad (8.36)$$

$$NP \rightarrow NP PP \quad \text{e.g., the President of the Georgia Institute of Technology} \quad (8.37)$$

$$NP \rightarrow NP SBAR \quad \text{e.g., the bicycle that I found outside} \quad (8.38)$$

$$NP \rightarrow NP VP \quad \text{e.g., a bicycle made of titanium} \quad (8.39)$$

These recursive productions are a major source of ambiguity, because the VP and PP non-terminals 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 constituents**

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

$$PP \rightarrow IN NP \quad \text{United States of America} \quad (8.59)$$

$$PP \rightarrow TO NP \quad \text{He gave his kimchi to Abigail} \quad (8.60)$$

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

$$SBAR \rightarrow IN S \quad \text{She said that it was spicy} \quad (8.61)$$

$$SBAR \rightarrow S \quad \text{She said it was spicy} \quad (8.62)$$

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

$$ADVP \rightarrow RB RBR \quad \text{They went considerably further} \quad (8.63)$$

$$ADVP \rightarrow ADVP PP \quad \text{They went considerably further than before} \quad (8.64)$$

The tag RBR is a comparative adverb.

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

$$ADJP \rightarrow RB JJ \quad \text{very hungry} \quad (8.65)$$

$$ADJP \rightarrow RBR JJ \quad \text{more hungry} \quad (8.66)$$

$$ADJP \rightarrow JJS JJ \quad \text{best possible} \quad (8.67)$$

$$ADJP \rightarrow RB JJR \quad \text{even bigger} \quad (8.68)$$

$$ADJP \rightarrow JJ CC JJ \quad \text{high and mighty} \quad (8.69)$$

$$ADJP \rightarrow JJ JJ \quad \text{West German} \quad (8.70)$$

$$ADJP \rightarrow RB VBN \quad \text{previously reported} \quad (8.71)$$

The tags JJR and JJS refer to comparative and superlative adjectives respectively. All of these phrase types can be coordinated:

$$PP \rightarrow PP CC PP \quad \text{on time and under budget} \quad (8.72)$$

$$ADVP \rightarrow ADVP CC ADVP \quad \text{now and two years ago} \quad (8.73)$$

$$ADJP \rightarrow ADJP CC ADJP \quad \text{quaint and rather deceptive} \quad (8.74)$$

$$SBAR \rightarrow SBAR CC SBAR \quad \text{whether they want control or whether they want exports} \quad (8.75)$$

**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*).<sup>9</sup> Each of these forms can constitute a verb phrase on its own:

$$VP \rightarrow VB | VBZ | VBD | VBN | VBG | VBP \quad [9.41]$$

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):

$$VP \rightarrow MD VP \quad \text{She will snack} \quad [9.42]$$

$$VP \rightarrow VBD VP \quad \text{She had snacked} \quad [9.43]$$

$$VP \rightarrow VBZ VP \quad \text{She has been snacking} \quad [9.44]$$

$$VP \rightarrow VBN VP \quad \text{She has been snacking} \quad [9.45]$$

$$VP \rightarrow TO VP \quad \text{She wants to snack} \quad [9.46]$$

$$VP \rightarrow VP CC VP \quad \text{She buys and eats many snacks} \quad [9.47]$$

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:

$$VP \rightarrow VBZ NP \quad \text{She teaches algebra} \quad [9.48]$$

$$VP \rightarrow VBG NP \quad \text{She has been teaching algebra} \quad [9.49]$$

$$VP \rightarrow VBD NP NP \quad \text{She taught her brother algebra} \quad [9.50]$$

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:

$$VP \rightarrow VBZ S \quad \text{Hunter wants to eat the kimchi} \quad [9.51]$$

$$VP \rightarrow VBZ SBAR \quad \text{Hunter knows that Tristan ate the kimchi} \quad [9.52]$$

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.

<sup>9</sup>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:

$$VP \rightarrow VBZ PP \quad \text{She studies at night} \quad (8.52)$$

$$VP \rightarrow VBZ ADVP \quad \text{She studies intensively} \quad (8.53)$$

$$VP \rightarrow ADVP VBG \quad \text{She is not studying} \quad (8.54)$$

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:

$$VP \rightarrow VBZ ADJP \quad \text{She is hungry} \quad (8.55)$$

$$VP \rightarrow VBP ADJP \quad \text{Success seems increasingly unlikely} \quad (8.56)$$

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:

$$VP \rightarrow VB PRT \quad \text{She told them to fuck off} \quad (8.57)$$

$$VP \rightarrow VBD PRT NP \quad \text{They gave up their ill-gotten gains} \quad (8.58)$$

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 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.

CS 485 CFG exercise, 3/7/2024    Name: \_\_\_\_\_

*Write answers below.*