3. Syntax

1. Introduction

1.1 What's Syntax?

When studying morphology one discovers that there is a hierarchical structure to the way morphemes are put together. For instance, the morphemes in a word like “rehospitalize” are lined up in a linear order as in (1).

1. re + hospital + ize

a linear order for rehospitalize

Note that just talking about the order (using + signs) doesn't quite capture the fact that the morphemes re- and -ize have a different relationship to the root hospital. The morpheme re- can only be attached to verbal stems, so we have to make hospital into a verb first (hospital + ize) and then attach the prefix re-. We use a tree to represent this more complex relationship among the morphemes:

2.          Verb
    Aff             Verb
        re-   Noun   Aff
            hospital -ize

This structure is "hidden", because in speech we just "line up" the morphemes. But a structure is needed to account for our knowledge of what words we can and cannot build using these morphemes. When it comes to sentences, where words are put together with other words, we could talk about the result as an ordering of the parts of speech, as shown below.

3. I watched the movie about snakes on a plane in August

But we are going to see that this is not enough to capture your knowledge of sentences. Instead, when whole words are put together with other words into sentences there is a structure (hierarchical relationship) between them - a structure similar, but not identical, to morphological structure. Again, this structure is "hidden" from us - but we have evidence that it is there. Sentences are made up of words, but we can show that there's another level of organization in between - that is, that sentences are made up of phrases, and phrases are made of words. By the end of this section you will see that what we really need to do is represent the sentence above with a tree:
You need to know right now is that certain words are grouped together before being grouped with other words. Unlike a morphological tree, though, the top of this tree is a sentence and each of the branches ends in a word. This tree has all the information the linear order representation in (4) has (part of speech, order of words) but it also has something more. The words are grouped together into little mini-trees, and at the top of each mini tree there is a symbol °. This represents the kind of thing that words are grouped into - and we'll call them phrases. These phrases are put together with other words and phrases to form a sentence.

Syntax is the study of these kinds of structures. It is concerned with the ways that words are assembled into phrases (the little trees marked with dots), and phrases are assembled into sentences. The next section will describe what kinds of word groups count as phrases, and how to identify them. That is, we will see that there is "hidden structure" to sentences. This is one of the most amazing things about language and linguistics.

2. Categories and Structure
A fundamental fact about words all human languages has is that they can be grouped together into a relatively small number of classes called syntactic categories. This classification reflects a category of factors, including the type of meaning that words express, the types of affixes that they take, and the type of structures in which they can occur.
2.1 Categories of words

5. Lexical Categories

<table>
<thead>
<tr>
<th>Lexical Categories</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun (N)</td>
<td>Harry, girl, wheat, policy, moisture, bravery</td>
</tr>
<tr>
<td>Verb (V)</td>
<td>arrive, discuss, melt, hear, remain, dislike</td>
</tr>
<tr>
<td>Adjective (Adj)</td>
<td>good, tall, intelligent, beautiful, clean</td>
</tr>
<tr>
<td>Preposition (P)</td>
<td>to, in, on, near, at, by</td>
</tr>
<tr>
<td>Adverb (Adv)</td>
<td>slowly, quietly, now, always, perhaps</td>
</tr>
</tbody>
</table>

6. Functional Categories

<table>
<thead>
<tr>
<th>Functional Categories</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determiner (Det)</td>
<td>the, a</td>
</tr>
<tr>
<td>Auxiliary Verb (Aux)</td>
<td>will, can, may, must, should, could</td>
</tr>
<tr>
<td>Modal Non-modal</td>
<td>have, be</td>
</tr>
<tr>
<td>Conjunction (Con)</td>
<td>and, or, but</td>
</tr>
<tr>
<td>Degree word (Deg)1</td>
<td>too, so, very, more, quite</td>
</tr>
</tbody>
</table>

To help you distinguish the different categories, here are a few words:

Noun: could be a person, place, thing, quality or act.
Determiner: signal presence of a noun, we say that it determines the noun.
Adjectives: designate a property, attributes entities denoted by a noun; describes or modify a noun.
Verb: express action, state/existence, occurrence.
Preposition: express a relationship between a noun with a verb, with an adjective, or to another noun.
Adverb: typically denote properties and attributes of the actions, sensations, and states designated by verbs. Modifier of a noun, an adjective or an adverb.
Conjunction: connects words with phrases and sentences.

7. Barack Obama became the new President of the United States.
   noun noun verb det adj noun prep det noun noun

2.2 A reliable criterion: distribution
A reliable criterion for determining a word's category involves the type of elements with which it can co-occur (call "distribution").

8. Distribution

<table>
<thead>
<tr>
<th>Category</th>
<th>Distributional property</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun (N)</td>
<td>occurrence with a determiner</td>
<td>a car, a wheat</td>
</tr>
<tr>
<td>Verb (V)</td>
<td>occurrence with an auxiliary</td>
<td>has gone, will stay</td>
</tr>
</tbody>
</table>

---

1 Don’t worry too much about this: we will use them to recognize adverbs but we will mostly consider them as adverbs because they actually are.

2 Refer also to the Morphology handout where is detailed the different parts of speech.
Adjective (Adj) occurrence with a degree word | very rich, too big

Of course, a verb cannot occur with a determiner or degree word in these sorts of patterns and a noun cannot occur with an auxiliairy:

8. *the destroy
9. *very arrive
10. *will destruction

You can use distributional tests with confidence when it is necessary to categorize unfamiliar words.

12. Practice on your own. You have 5 min.

The boss of my firm eats lunch near the river bank.

The teacher often organised a discussion.

Marianne could become quite fond of Larry.

3. Phrase structure

As we said before, sentences are not formed by simply stringing words together like beads on a necklace. Rather, sentences have a hierarchical design in which words are grouped together into successively larger structural units.

We can see a relationship between words and phrases by noting that whenever you find a single noun in a very simple sentence, you can replace it with a group of words, all related to each other and crucially containing a noun, and the sentence will still be grammatical. So for example, starting with the simple sentence,

13. John runs

We could replace the single noun John with a group of related words:

14. The skinny little guy in the neon tank top runs

The sentence is still fine - and, if you know the John I know, it means the same thing. This group of words is called a noun phrase it's a phrase that's built around a central noun (guy), and the whole phrase can be traded for a single noun and the whole sentence will still make sense. Notice that while we can replace this noun phrase (abbreviated "NP") with a noun and still have a good sentence, we can't replace it with any other part of speech (remember that an asterisk means ungrammatical).

15. *Swim (V) runs
16. *Fast (Adj) runs
17. *Down (P) runs
We will also talk about two other kinds of phrases right now: verb phrases (VP) and prepositional phrases (PP). A verb phrase is a group of words that can substitute for a single verb, or that can have a single verb substituted for them; a prepositional phrase is "unsurprisingly" a group of words that can substitute for a single preposition, or that can have a single preposition substituted for them.

NP: shortest: it
longer: the utterly horrid estate of Lady Catherine de Bourgh

PP: shortest: out (in: Lizzy's gone out.)
longer: in the spacious entrance hall

VP: shortest: sleeps (in: Jane sleeps.)
longer: reads a very good book on politics.

4. Constituency tests

When we talked about figuring out what part of speech a single word was, we said that while intuitive definitions like "A noun is a person, place, or thing" might be useful guides, to be really precise we needed to use structural definitions of categories like nouns to figure out whether a given word was a noun. Similarly, while an intuitive definition of a noun phrase like "A noun phrase is a group of words that describe a noun, and can go in sentences where nouns can go" may be a useful guide, we'll use formal tests called constituency tests to prove whether groups of words are particular kinds of phrases.

Constituency test  A way of testing whether a group of words forms a phrase. There are different constituency tests to check whether groups of words are different types of phrases.

The constituency test for noun phrases is the pronoun test, where you replace a group of words that you think might be a noun phrase with a pronoun; if the resulting sentence is grammatical, and means the same thing as the original sentence (that is, the resulting sentence is a shorter way of saying the original one), the group of words is a NP.

Pronoun test  Replace a group of words with a pronoun, like he, she, him, her, it, they, or them. If any of these replacements results in a grammatical sentence with the same meaning as the original sentence, the group of words is a noun phrase. If none of the replacements has this result, the words are not a noun phrase.

18  a. Bingley and Darcy meet in the parlour.
    b. They meet in the parlour.

19  a. I watched the movie about snakes on a plane last week.
    b. I watched it last week

The new sentence is grammatical and means the same thing as the original sentence, so the bolded words in the first sentence form a noun phrase.
20  a. Most people like the guy in that movie about snakes.
   b. Most people like him.

The new sentence is grammatical and means the same thing as the original sentence, so the bolded words in the first sentence form a noun phrase.

21  a. I watched the movie about snakes last week.
   b. * I watched the it last week.

The new sentence is not grammatical, so the bolded words in the first sentence do not form a noun phrase.

22  a. Bob threw away his homework.
   b. Bob threw it.

The new sentence is grammatical, so it is not marked with a (*). However, it does not mean the same thing as the original sentence, so the bolded words in the first sentence do not form a noun phrase.

Note: It is important to understand that Phrases can be just single words. We can prove this using the pronoun test on a sentence like “Bob likes cheese”. The sentence “Bob likes it” is grammatical and has the same meaning as the original sentence. Since cheese can be replaced with the pronoun it, cheese is a one-word NP.

There's also a constituency test for verb phrases: the do-form test.

**Do-form test** Replace a group of words with a form of do so. If the replacement results in a grammatical sentence with the same meaning as the original sentence, the group of words is a verb phrase. If the replacement does not have this result, the words do not constitute a verb phrase.

23  a. I'm going to watch the movie.
   b. (I said I'd watch the movie, and) I'm going to do so.

The new sentence is grammatical, so the bolded words in the first sentence form a verb phrase.

The part in parentheses is added to provide more context, so that the sentence with "do so" sounds more natural. Let's look at another example.

24 a. I watched the movie about snakes in August.
   b. (I wanted to watch the movie about snakes in July but,) I did so in August.  

The new sentence is grammatical, so the bolded words in the first sentence form a verb phrase.

__________________________

3 Note that we keep using the past tense when we replaced the verb by do so, so we got ‘did so’.
25 a. I watched the movie about snakes in August. b. *I did so about snakes in August.

The new sentence is not grammatical, so the bolded words in the first sentence do not form a verb phrase in this.

The constituency test for prepositional phrases is the there/then test.

**There/then test** Replace a group of words with there or then. If the replacement results in a grammatical sentence with the same meaning as the original sentence, the group of words is a prepositional phrase. If the replacement does not have this result, the words are not a prepositional phrase.

I watched the movie about snakes in August → I watched the movie about snakes then.

The new sentence is grammatical, so the bolded words in the first sentence form a prepositional phrase.

26 a. I watched the movie in Brooklyn. b. I watched the movie there.

The new sentence is grammatical, so the bolded words in the first sentence form a prepositional phrase.

27a. I wanted to watch the movie about snakes in Brooklyn. b. *I wanted there.

The new sentence is not grammatical, so the bolded words in the first sentence do not form a prepositional phrase. Also, we should take note of the fact that here to is not a preposition. It indicates the infinitival form of watch.

**The Movement Test** There's another constituency test that usually identifies NPs, VPs, and PPs. When one of the other tests doesn't work, the movement test often does. (We'll see in class that the above constituency tests are imperfect, and we sometimes need to perform more than one.) The movement test allows us to move a group of words to the front of the sentence. Let's look at the sentences from above.

**NP**
(i) a. I watched the movie about snakes on a plane last week.
   b. The movie about snakes on a plane is what I watched last week.

**VP**
(ii) a. I watched the movie about snakes.
   b. Watch the movie about snakes is what I did.

**PP**
(iii) a. I watched the movie in Brooklyn.
   b. In Brooklyn is where I watched the movie.
Identifying phrases in a sentence can often be slightly easier than this. There is an important rule about phrases that helps you identify them:

Every N is part of an NP.

Every V is part of a VP.

Every P is part of a PP.

The technical term for this relationship between N and NP, V and VP, and so on, is head. Every NP has as its head an N, and so on for each phrase. So we can generalize this rule to say that any phrase XP (where X stand in for any part of speech) has an X in it - that is, has a head X.

This means that when you are asked to find all of the phrases in a sentence, if the sentence has 3 nouns, 1 preposition, and 1 verb (like the one below), you should find 3 NPs, 1 PP, and at least one VP. Try to find the phrases in the following sentence:

28a. My young neighbor won a kangaroo during the weekend.

4.2 Exercises on constituency tests

Every N is part of an NP.

Every V is part of a VP.

Every P is part of a PP.

The technical term for this relationship between N and NP, V and VP, and so on, is head. Every NP has as its head an N, and so on for each phrase. So we can generalize this rule to say that any phrase XP (where X stand in for any part of speech) has an X in it – that is, has a head X.

This means that when you are asked to find all of the phrases in a sentence, if the sentence has 3 nouns, 1 preposition, and 1 verb (like the one below), you should find 3 NPs, 1 PP, and at least one VP. Try to find the phrases in the sentence:

28b. The little guy watched the movie in Brooklyn

Det Adj N V det N Prep N

Noun phrases (NP) in dotted-line boxes:

The little guy watched the movie in Brooklyn

Pronoun test: he watched the movie in Brooklyn. “The little guy” is a NP.
Movement test: The movie is what the little guy watched in Brooklyn. “The movie” is a NP.

NP “Brooklyn”:
The pronoun or the movement tests won’t work for the last NP, “Brooklyn”, but I still expect you to know that it’s a NP anyway.
The little guy watched the movie in Brooklyn

Prepositional phrase (PP) in dotted-line circle:

There / then test: The little guy watched the movie there. “In Brooklyn” is a PP.

Verb-phrases (VP) in solid-line circles:

Do form test: (the little guy said he would watch the movie), and he did so in Brooklyn. “Watched the movie” is a VP.

Movement test: Watch the movie in Brooklyn is what the little guy did. Watch the movie in Brooklyn is a VP.

Sentence (S) in a solid-line box:

There are a few things that we should take note of here. The first is that there are two VPs, even though there is only one verb. We have the larger VP, watched the movie in Brooklyn, and the smaller VP, watched the movie. We’ll cover this in more detail when we outline our phrase structure rules for verb phrases. Also, we see that VPs start with verbs and PPs start with prepositions. However, we see that NPs start with either a noun or a determiner. Finally, these diagrams represent the fact that phrases often contain other phrases – for example, the PP has an NP inside it, and so does the smaller VP. The big VP has both a VP and a PP in it.

Another exercise:
- What about this sentence: Sarah and Peter ate croissants in Paris on Monday. How many NPs, VPs and PPs are there? Run different tests to find all the phrases in the sentence.
Try now to find the 6 phrases in the following sentence:
Clyde got a passionate love letter from Stacy.

4.3 Phrase structure rules: an overview
In studying the structure of sentences, a major goal is to provide a general description of what sentences, NPs, VPs, PPs, look like in different languages. We do that by writing general rules for the structure of each of these kinds of phrases. These rules are called phrase structure rules (PSRs).

Let’s say we’ve figured out (as we likely have by now) that every single English sentence has a noun phrase and then a verb phrase. We can write a schematic rule that says this:

\[(29) \ S \rightarrow NP \ VP\]

If you were to read that rule in words, it would sound like “A sentence consists of a noun phrase followed by a verb phrase.” To be really explicit, the pieces of this rule mean the following:

\[(30) \ S = \text{“a sentence”} \quad \rightarrow = \text{“consists of”} \quad NP = \text{“a noun phrase”} \quad VP = \text{“a verb phrase”}\]

The left-to-right order in the rule corresponds to left-to-right order in a sentence. That’s where the “noun phrase followed by a verb phrase” part comes from – the order of the constituents after the arrow.

Syntax rules can, of course, build things other than sentences:

\[(31) \ NP \rightarrow Adj \ N\]

This rule says that a noun phrase consists of an adjective followed by a noun, like in pretty girls. We know, though, that not all English noun phrases have adjectives in them. You can start a sentence with just a noun, as in Girls laugh or Mary laughs. Other times, you can put several adjectives in a noun phrase: small smart pretty girls laugh, etc. Because adjectives are optional and because of our ability to continually add adjectives, we are going to end up revising the NP rule, so don’t commit this one to memory.

Our constituency tests showed us we also have verb phrases, VPs. Here’s an attempt at a phrase structure rule for a verb phrase such as watch a movie.

\[(32) \ VP \rightarrow V \ NP\]

We have the verb watch and the NP a movie. This phrase structure rule itself has another phrase in it (so did the Sentence phrase structure rule). This means that one phrase structure rule leads to another, and that’s why we can have phrases inside of other phrases.

NOTE: The VP watch a movie provides another reason for us to revise our NP rule. We have the determiner a, and presently, our NP rule doesn’t indicate a place for determiners.
4.4 Syntactic Trees

Throughout the rest of this chapter, we are going to be using our phrase structure rules to generate syntactic trees, so before we go into any more detail about the nature of phrase structure rules, let’s learn the basics of tree-drawing.

Syntactic trees are much like morphological trees, but instead of building words, the trees build words into phrases and sentences. The three phrase structure rules proposed above would be represented by the following trees.

(33) Phrase Structure Rule Corresponding Tree Diagram

S → NP VP

<table>
<thead>
<tr>
<th>Phrase Structure Rule</th>
<th>Corresponding Tree Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP VP</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>VP</td>
</tr>
<tr>
<td>NP → Adj N</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Adj</td>
</tr>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>VP → V NP</td>
<td>VP</td>
</tr>
<tr>
<td></td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>NP</td>
</tr>
</tbody>
</table>

(NOTE: These phrase structure rules are just examples to get us started. We are going to keep S → NP VP, but the other two will be changed slightly, so don’t commit them to memory.)

You can put these altogether when you have a whole sentence:

A syntax tree is a representation of the structure of a single sentence – of the way that the individual words are grouped into phrases and eventually form the full sentence. A tree shows every phrase that’s in the sentence, so if your project is to draw a tree, it’s a really excellent idea to start by labeling every word with its part of speech, and then using constituency tests to find all of the phrases in the sentence.
Question: Do the constituency tests for NPs and VPs confirm that we have grouped the words in this sentence correctly? If not, we’ve made a mistake.

Very importantly, the groupings of words into phrases and into the sentence in a syntax tree absolutely must be an accurate representation of the phrase structure of the language. You’ll usually be asked to draw a syntax tree after you’ve written phrase structure rules for a language. Other times, you might be given a tree and asked to write the rules. If you just make sure that you’re identifying all phrases when both writing phrase structure rules and drawing trees, and that the phrases in your tree can be generated using your phrase structure rules, this shouldn’t be a problem.

Important properties of trees
- Every word must be labeled with its part of speech.
- Every phrase present in the sentence must be correctly drawn in the tree.
- Every word must be connected into the tree.
- The groupings of words into phrases and phrases into sentences as shown in the tree must be consistent with the language’s phrase structure rules.

Some Terminology:
The category that appears to the left of the arrow in the phrase structure rule is the mother and the categories that appear to the right of the arrow are the daughters. Daughters that have the same mother are sisters. Let’s consider the phrase structure rule $S \rightarrow NP \ VP$ and its corresponding tree:

```
    S
     /\       
    NP  VP   
      /\     
     daughter daughter
```

The S node is the mother and its daughters are the NP and VP nodes. NP and VP are sisters.

4.5 Phrase structure rules

4.5.1 The structure of noun phrases

Now let’s return to the issue of why (31) doesn’t make for a good NP rule. We’ve already said that it doesn’t include determiners, it doesn’t allow for more than one adjective, and it doesn’t tell us which elements are optional. We’ll also see that NPs can have PPs and sentences inside of them. Here are some more examples of NPs.

(34) a. cats
     b. black cats
     c. the black cats
     d. the small black cats
     e. the small black cats on the couch
f. the small black cats that John owns

Let’s take the NPs in (34a-c) first. Here’s an attempt at a phrase structure rule:

(35) NP → (Det) (Adj) N

This rule says that an NP consists of an optional determiner, an optional adjective, and a noun (which is required).

This looks good, right? We would draw the tree like (36) for (34c):

(36)           NP
              Det    Adj    N
              the    black    cats

This is what is called a flat phrase structure. That’s because below the NP level, all the other categories are lined up (technically, they are all daughters to the NP). But there is evidence that this isn’t so. There is another kind of constituency test that distinguishes different constituents within an NP. You can often replace parts of an NP with the word one.

(37) a. the red shoes → the red ones

    b. the silly linguistics teachers → those silly ones

    c. the small green bottles of beer → the small ones

    d. some intelligent students of language → some intelligent ones

These replacements suggest that the adjective and the noun, and even the noun itself, form a constituent to the exclusion of the determiner. But we still want the whole NP to come together as an NP. To do so, we’re going to add in another “level” to our phrases. We have heads (the categories at the bottom that are just above the actual words) and we have the phrases, the XPs on top of the phrases. Let’s say there can be a level in-between, and we will notate this with an apostrophe/prime symbol next to the category of phrase. So, in the tree in (31), we have N'.

(38)     NP
          Det      N'
          the
          Adj
          Small
          N' replaceable by one(s)
          N' bottles
          PP
          P
          NP
          N'
          N
          of
          N' beer
With this extra level of structure we can account for why the word *ones* can replace *bottles of beer* but leave the adjective *small*. That’s because its forms a constituent to the exclusion of the adjective (and the determiner). In general, these extra levels of structure – which we will call Bar-levels – are where we attach modifiers, such as adjectives and some prepositional phrases. The number of bar-levels is, in principle, infinite. In this way, it is recursive (it can ‘recur’ over and over again). This is why we could keep adding adjectives in an NP, as in (34d). So we are going to replace the rule in (35) with the following:

(39) a. \( \text{NP} \to (\text{Det}) \text{N'} \)
    b. \( \text{N'} \to \text{Adj N'} \)
    c. \( \text{N'} \to \text{N} \)

The rule in (39a) says that an NP consists of an optional determiner and necessarily an N’. The rules in (39b-c) tell us what an N’ can consist of. It can consist of an adjective and another N’ or it can consist of just an N. It’s the rule in (39b) that allows us to have an infinite number of adjectives. Because we have N’ on both sides of the arrow, this phrase structure rule can keep recalling itself, as in (40).

(40) Several small black scary angry cats

\[
\text{NP} \\
\text{Det} \quad \text{N'} \\
\text{Several} \\
\quad \text{Adj} \quad \text{N'} \\
\quad \text{small} \\
\quad \text{Adj} \quad \text{N'} \\
\quad \text{black} \\
\quad \text{Adj} \quad \text{N'} \\
\quad \text{scary} \\
\quad \text{Adj} \quad \text{N'} \\
\quad \text{angry} \\
\quad \text{N} \\
\text{cats}
\]

We can also have prepositional phrases inside of NPs, as we saw in (34e) *the small black cats on the couch* (and which we see in the tree in (42)). The rules for this are in (41) and the tree for (34e) is in (42).

(41) a. \( \text{N'} \to \text{N'} \text{ PP} \)
    b. \( \text{PP} \to \text{P NP} \)
Finally, we need to account for (34f) the small black cats that John owns. Here, we have a sentence – that John owns – inside of our noun phrase. This sentence is different from the other sentences we’ve looked at because it begins with what we call a complementizer, that. (Of course, we can also say this NP without the complementizer, as in the small black cat John owns. In this case, we would say that the complementizer is null, or silent.) Complementizers include words such as that, if, and whether. The rules that account for (34g) are in (43) and the tree is in (44).

\begin{align*}
(43) \text{a. } & \mathbf{N'} \rightarrow \mathbf{N'} \text{ CP} \\
\text{b. } & \text{ CP } \rightarrow \text{ C } \text{ S}
\end{align*}

(44) NP

With the exception of the VP rule, all the other rules used to generate this tree should be familiar at this point. The rule S $\rightarrow$ NP VP was introduced in (30) and schematized in (33). So, here are the rules we need to account for NPs:
4.5.2 Phrase Structure Rules: The Structure of Verb Phrases

We’ve spent a lot of time looking at how noun phrases are formed. Since we have some background, dissecting verb phrases will be much easier. Let’s look at the verb phrases in (42).

Looking at (46a-c) we see that VPs can contain a single verb, a verb followed by a preposition, or a verb followed by a preposition that is followed by a noun phrase. The following rules capture this:

(47)  a. VP → VP PP
      b. VP → V

Let’s draw a tree for (46c):

(48)     VP
      /   \
     VP   PP
       |    \
      V    NP
     jump into Det N'
     the   N
     pool

We have a VP, instead of just a V, as the first daughter because we’ll want to be able to account for phrases like jumped off the diving board into the pool. Here, we have two PPs, and our rule in (47a) would allow us to incorporate both of them by recalling the VP.

Of course, VPs can have noun phrases in them, as in (46e). Here’s our rule and the corresponding tree.

(49) VP → VP NP
Sometimes VPs contain auxiliary verbs (what you probably learned as helping verbs), as in (46f) *should play the piano.*

\[
(51) \text{VP} \rightarrow \text{AUX VP}
\]

\[
(52)
\begin{array}{c}
\text{Aux} \\
\text{should} \\
\text{VP} \\
\text{V} \\
\text{play} \\
\text{NP} \\
\text{Det} \\
\text{the} \\
\text{N'} \\
\text{piano}
\end{array}
\]

Finally, some VPs contain sentences, just as some NPs do, as in (46g) *heard that John plays the piano.*

\[
(53) \text{VP} \rightarrow \text{VP CP}
\]

\[
(54)
\begin{array}{c}
\text{VP} \\
\text{V} \\
\text{heard} \\
\text{CP} \\
\text{C} \\
\text{that} \\
\text{S} \\
\text{NP} \\
\text{VP} \\
\text{N'} \\
\text{V} \\
\text{plays} \\
\text{NP} \\
\text{N} \\
\text{John} \\
\text{Det} \\
\text{the} \\
\text{N'} \\
\text{piano}
\end{array}
\]

4.5 How to do syntactic analysis (and find phrase structure rules)

There will be (probably many) times when you’ll be given some sentences in a language that isn’t English, and you’ll be asked to write phrase structure rules that describe the language. Sometimes you’ll also be given some English data that we haven’t studied, and asked to write some new phrase structure rules that describe how the new data works. These phrase structure rule writing tasks are pretty similar. There’s a lot of discussion in this section, but the steps are underlined in the following paragraphs, and summarized at the end of the section.

Start by labelling every word with its part of speech. You’ll need to know all of the words’ parts of speech eventually, because you write rules that say which parts of speech (or which kinds of phrases) go where. Also, if you can look at sentences in terms of the types of words/phrases they contain, rather than literally what words they contain, this can really help with seeing similar structures in different-looking sentences.
Summary: To find phrase structure rules in English

1. Label all words with their part of speech.

2. Find all phrases in each sentence.

Every N, V, or P heads a NP, VP, or PP. So to find all phrases:
• Find each head (N, V, P).
• Use constituency tests to figure out what NP, VP, or PP that head is a part of.

3. Write individual phrase structure rules to describe the structure of each phrase (each NP, VP, PP, and S).

These rules should describe all of the pieces of the phrase, at the most general level possible. This means describing the contents of a phrase in terms of constituent phrases, instead of individual constituent words, whenever possible.

4. After you’ve done this for each sentence in the data, look at all the rules for a single kind of phrase (e.g. VP) and see if you can combine them at all (see below). Do this for each type of phrase: NP, VP, PP, and S.

On combining sets of phrase structure rules into single rules

Sometimes you can combine rules into a single, more efficient rule; other times you can’t. It’s not strictly wrong to leave rules uncombined, so when in doubt, don’t combine them. But it’s generally better to have one or two general rules than sixteen specific, redundant rules, so when it’s clearly safe, it’ll make me happier (and you might get a couple more points, and also your hand will get less tired) if you combine rules.

A case where you cannot combine rules:

Say you have the following rules:

\[
\begin{align*}
\text{VP} & \rightarrow \ V \ NP \\
\text{VP} & \rightarrow \ V \ PP
\end{align*}
\]

It’s not correct to combine these to form “\(\text{VP} \rightarrow \ V (\text{NP}) (\text{PP})\)”, because you don’t have evidence that a sentence can have both NP and PP after a V, or that a sentence can have just a V and neither a following NP nor a following PP, or for what order the two would be in if they could cooccur.

A case where you can combine rules:

If you have the following rules:

\[
\begin{align*}
\text{VP} & \rightarrow \ V \\
\text{VP} & \rightarrow \ V \ NP \\
\text{VP} & \rightarrow \ V \ PP \\
\text{VP} & \rightarrow \ V \ PP \ NP
\end{align*}
\]
…then it’s okay to combine them to “VP → V (PP) (NP)”, because you know that a single VP can have both NP and PP (or neither), and you know which order they go in when they combine.

### 4.6 Inventory of English Phrase Structure Rules

Basic Rules: These are the PSRs that we’ll be using for English. You should be able to provide example phrases and sentences that use the PSRs.

#### Important: Memorize this chart

<table>
<thead>
<tr>
<th>S → NP VP</th>
<th>NP → (D) N’</th>
<th>VP → VP PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → CP VP</td>
<td>NP → N’ PP</td>
<td>VP → VP NP</td>
</tr>
<tr>
<td>CP → C S</td>
<td>N’ → Adj N’</td>
<td>VP → AUX VP</td>
</tr>
<tr>
<td>PP → P (NP)</td>
<td>N’ → N</td>
<td>VP → V</td>
</tr>
<tr>
<td>αP → αP CON αP</td>
<td></td>
<td>VP → VP CP</td>
</tr>
</tbody>
</table>

There are a couple of rules here that haven’t yet been introduced:

S → CP VP and αP → αP CON αP

An example of S → CP VP is illustrated by the sentence *That Mary shops without coupons bothers John.* Here, instead of starting our sentence with a noun phrase, we are starting it with a complementizer phrase *that Mary shops without coupons*. The tree is shown in (55).

(55)

Finally, we want to be able to represent sentences in which phrases are conjoined with *and*, *or*, or *but*, as in *Theresa likes fine cuisine and good wine*. The rule αP → αP CON αP allows us to do this. This rule looks different from all the other phrase structure rules, but it’s quite simple to understand. This rule says that any category can consist of two members of itself. So, we can conjoin two VPs to form a larger VP or two NPs to form a larger NP, etc. Let’s see how this works.
Here we have two N’s conjoined, *fine cuisine* and *good wine*. This is the only time when a node will have three branches. Here we have two N’ daughters, as well as *and*. It’s very important to remember that **whenever we have conjunction, the two daughters are necessarily identical to the mother.** So, we couldn’t ever have a VP mother that has as its conjoined daughters a VP and a PP, for instance. We don’t ever put the exact rule $\alpha P \rightarrow \alpha P \text{ CON } \alpha P$ into our syntactic trees. Rather, we use the particular category that $\alpha$ represents; so in (53), we use N’.

**Question:** We could also have drawn this tree with two conjoined NPs. How would it look?

As was mentioned above, because a tree is a representation of a particular sentence, and phrase structure rules ideally describe all possible sentences in a language, a tree has to be consistent with phrase structure rules. This doesn’t mean that a tree has to contain every single constituent listed in a phrase structure rule. To explain this better, let’s consider a sentence like *Happy students watch movies*. This sentence demonstrates that while it’s possible to have an adjective in a NP (like *happy students*), it’s not required since the NP *movies* has no adjective.