Syntax

Evan Halloran

November 2025

1 Generative Grammar

There are six main branches of linguistics. In order of consequence, they are

$$\text{phonetics} \rightarrow \text{phonology} \rightarrow \text{morphology} \rightarrow \text{syntax} \rightarrow \text{semantics} \rightarrow \text{pragmatics}$$

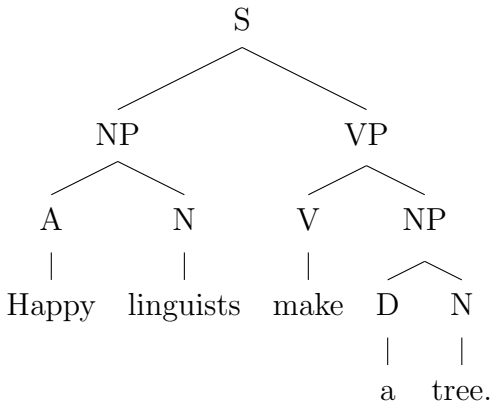
Phonetics is the study of the minimal units, or phones, that make up a language. In spoken language, the phones are utterances, while in signed language they are hand gestures. Phonology studies how sounds pattern in language. For instance, speakers of American English pronounce the /t/ phoneme differently in the word "type" than in the word "button." The reason as to why the /h/ phoneme can appear anywhere within a word except the end in English is also a matter of phonology. Morphology deals with the smallest units of meaning, and how these units combine to make words. The word "undesirable" is made up of three morphemes: a prefix, a root, and a suffix. "Desirable" is a word, but "undesire" is not. The word "ruthless" is commonplace, but "ruth" is hardly ever said. Syntax studies how words are systematically organized to form sentences. Semantics studies how these sentences convey meaning. For example, it's not too difficult to understand what's being asked in the question "who did gave John the book?" despite not abiding standard syntactic rules; conversely, the sentence "colorless green ideas sleep furiously"[1] is perfectly grammatical and yet it has no discernible meaning. Lastly, pragmatics deals with the use of language in the broader context of communication. If your partner mentions that they could really use a hand with dinner,

[1]This sentence was proposed by Noam Chomsky when he famously argued that syntax deserved its own field of linguistic study separate from semantics. Prior to this, probabilistic hypotheses suggested that the more likely a sentence were to be used in spoken conversation, the more likely it were to be grammatical. But the sentence above has probably never been uttered before Chomsky despite being a perfectly acceptable sentence of English. Even semantically absurd sentences can be syntactically sound.

they are soliciting your help despite not asking a direct question. If your coworker made the same comment, however, it's unlikely you would reach the same conclusion.

The study of syntax has undergone many developments since its beginnings in the early 20th century. Like other scientists, linguists formulate hypotheses and test them rigorously to establish theories. These theories try to address the main mystery of syntax—how it is that humans are able to systematically construct grammatical sentences of meaning to communicate with others. Perhaps the longest lasting and most influential of these frameworks is Generative Grammar (GG), which has its roots in the work of Noam Chomsky in the 1950s. Generative Grammar has undergone many developments and reformulations over the years. Some theories of syntax that fall under GG include Transformational Grammar, Government and Binding Theory, Principles and Parameters, and the recent Minimalist Program. These notes deal mostly with the Principles and Parameters framework, but will draw upon some other theories as needed. Some popular non-GG theories include Lexical-Functional Grammar (LFG) and Head-Driven Phrase Structure Grammar (HPSG).

One of the main purposes of the Generative Grammar program is to create syntax trees to highlight relationships between phrases within the sentence. Let's look at an example of a syntax tree. The highest node (root) in each tree is always S, which stands for sentence. The sentence node has exactly two children, a noun phrase (NP) and a verb phrase (VP). These phrases loosely correspond to the notions of subject and predicate one learns about in grade school. The noun phrase subject "happy linguists" consists of an adjective and a noun. The verb phrase predicate consists of a verb, "make," and a direct object NP "a tree." Finally, the direct object NP is made of a determiner (or article) and a noun. This is the general structure of a syntax tree.



The above tree is generated by a series of recursive rules that define what constitutes each phrase of the sentence. For instance, we know that a sentence is always made of exactly

one noun phrase subject and one verb phrase predicate. This rule can be written as S → NP VP, where the constituents to the right of the arrow produce the constituent on the left. The order of these constituents matter: we cannot have a sentence that begins with a verb phrase (yet). A full list of phrase structure rules might then look like this:

S → NP VP

NP → (D)(A)* N (PP)

VP → (Aux)* V (NP)(NP)(PP)* (CP)[Adv]

PP → P NP

CP → C S

X → X conj X

The first portion of these notes will be devoted to building and refining this list of phrase structure rules from grammatical English sentences. Before that, however, let's discuss what each of the rules mean. The first rule is the sentence requirement that was already discussed. The second rule is used to construct noun phrases and consists of a determiner, an adjective, a noun, and a prepositional phrase. In our phrase structure rules, the usage of parentheses around an argument indicates that it is optional—thus the only required part of a noun phrase is the noun itself. The asterisk following the adjective means that we are free to have as many adjectives in a noun phrase as we like. The following are noun phrases that can be generated by our rules:

dog

the dog

the big dog

the big, burly dog over there

The rule for verb phrases is the most convoluted of our phrase structure rules.[2] Similar to noun phrases, the only required part of a verb phrase is the verb. We are allowed any number of auxiliary verbs (modals such as *could, would, can*; tenses such as *have, been, will*; and moods such as *were*). A verb phrase is allowed two optional noun phrases: a

[2]This is due to the fact that English is a "head-initial" language, which means that modifiers follow the word or phrase (head) they modify. In English, the bulk of a sentence comes *after* the verb and is used to modify the predicate.

direct object and an indirect object. An arbitrary number of prepositional phrases may follow after the objects. An optional complementizer phrase (embedded sentence) is also permitted. A complementizer phrase is a subordinate clause nested within a larger sentence, and a complementizer is the word that introduces the clause. For example, in the sentence "*I know that Mary likes Max*," the complementizer is "*that*" and the complementizer phrase is "*that Mary likes Max*." Finally, a verb phrase can have an adverb, and its location in the phrase is largely unconstrained. This freedom to be placed anywhere is indicated by the brackets in our phrase structure rules. Note the different positions of the adverb in each of the following verb phrases:

Max [quickly biked to school after eating breakfast].

Max [biked quickly to school after eating breakfast].

Max [biked to school quickly after eating breakfast].

Max [biked to school after eating breakfast quickly].

Interestingly, the last of these sentences has a slightly different meaning from the other three (the eating was quick, not the biking). Going forward, syntax trees will be used to resolve any semantic ambiguity resulting from word order. The last of our phrase structure rules is different from the others encountered in that it can be used to construct any of the phrases mentioned up to this point. Simply put, it states that "any X can be made from two X's joined by a conjunction." A conjunction is a word such as *and, but, or*, etc. and is used to link two of the same heads or phrases. Observe how the rule is applied to the following constituents:

N → N conj N: Jake and I

NP → NP conj NP: the big dog and the lazy fox

VP → VP conj VP: sleep well but don't let the bed bugs bite

PP → PP conj PP: to the left or to the right

A → A conj A: hungry and angry

Together these phrase structure rules are used to generate a decent number of English sentences in a relatively simple manner, but they are not complete, that is, they do not generate the full corpus of grammatical English sentences. There are grammatical sentences that cannot be generated by these rules (e.g. *Are you hungry?*), as well as ungrammatical

sentences that **are** generated (e.g. **I can should do better*). We will have to amend these rules in the future as difficulties arise—in fact, we will discover that these phrase structure rules are flawed entirely and the need for a new theory will become apparent.

2 Parts of Speech

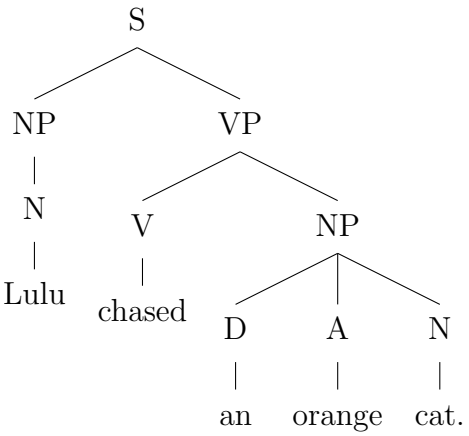
Before discussing how we might create our list of phrase structure rules, it's good to review the different parts of speech in the English language.

3 Constituency, Trees, and Rules

In this section, we will systematically build up our phrase structure rules, providing motivation from grammatical English sentences. The nature of these rules was discussed in Section 1. Here are the provisional rules to start with:

$$\text{S} \rightarrow \text{NP VP}$$
$$\text{NP} \rightarrow \text{(D)(A) N}$$
$$\text{VP} \rightarrow \text{V (NP)(PP)}$$
$$\text{PP} \rightarrow \text{P NP}$$

Let's begin applying these rules by making a syntax tree for the sentence "Lulu chased an orange cat."



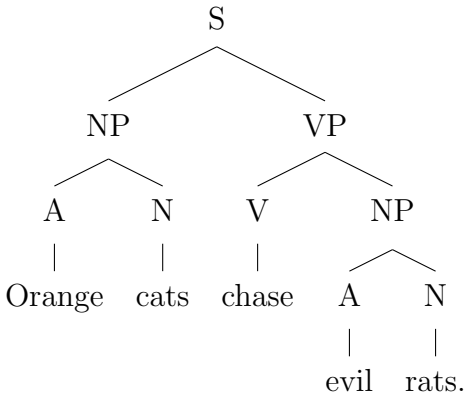
The structure of this tree is similar to the one discussed in Section 1. To see the other

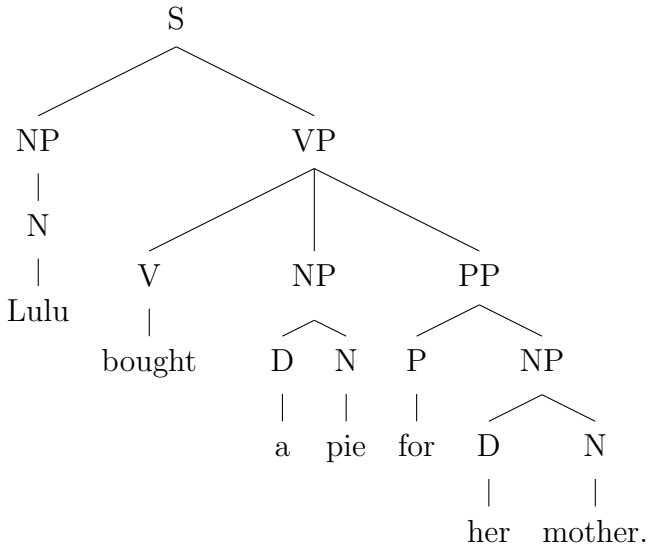
rules in action, let's make syntax trees for the following sentences:

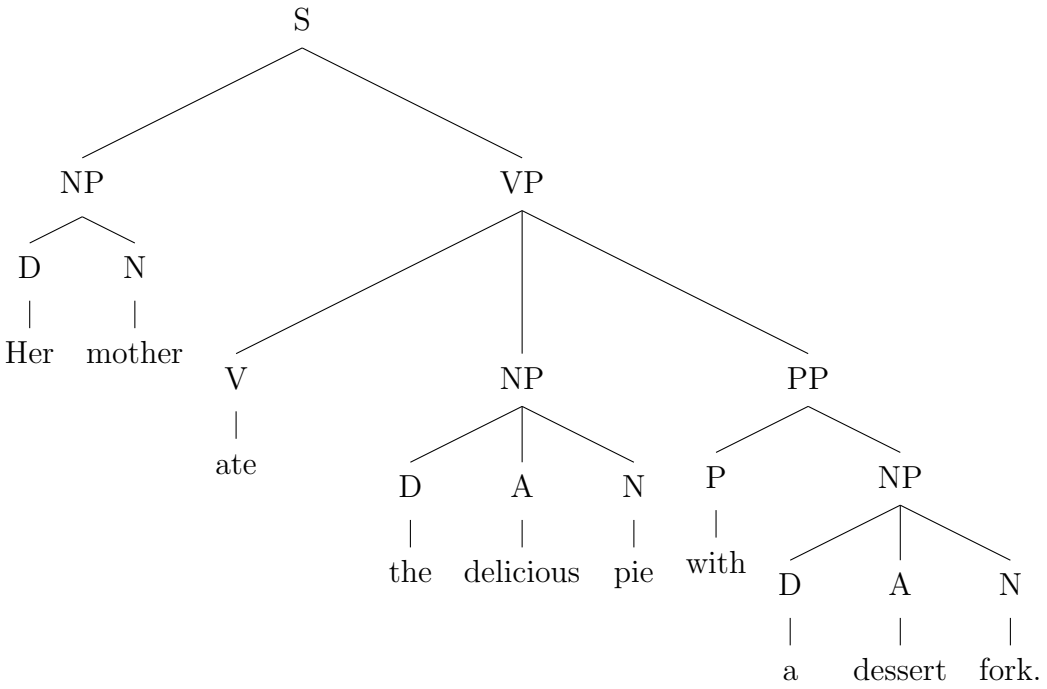
Orange cats chase evil rats.

Lulu bought a pie for her mother.

Her mother ate the delicious pie with a dessert fork.

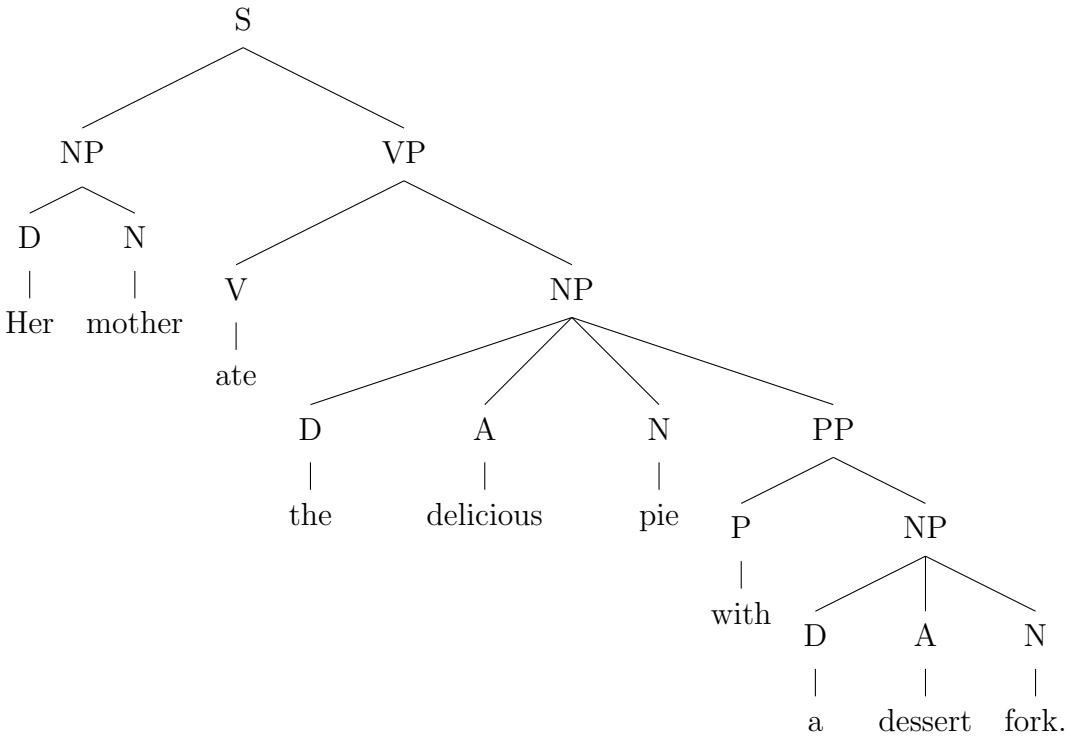






We should pause to examine the structure of the last tree. Notice that the sentence ends with the prepositional phrase [PP with a dessert fork]. Currently, this PP is attached to the

verb phrase. What would happen if the PP (legally) attached to the NP instead?



This tree is perfectly acceptable given our phrase structure rules, and yet it is semantically wrong. To see why, we should ask what it means for the PP to attach to the NP versus the VP. In the first case, we are treating the noun phrase

[NP the delicious pie [PP with a dessert fork]]

as one constituent. But this noun phrase is precisely what her mother ate, meaning if the PP were to form a constituency with the NP, then the mother is eating both a delicious pie and a dessert fork! Clearly this is wrong. What actually is happening is that the mother is eating a delicious pie, and she is using a dessert fork to eat. In terms of constituencies, the prepositional phrase [PP with a dessert fork] is a constituent of the verb phrase:

[VP ate [NP the delicious pie] [PP with a dessert fork]]

We will see more examples of a single sentence having multiple permissible syntax trees. These sentences are called semantically ambiguous, because the true meaning cannot be inferred by the one-dimensional string of words itself. In technical terms, it is not always possible to tell which adjacent phrases will form constituencies just based on word order. When a sentence is realized as a two dimensional tree-diagram, however, the meaning is

always clear because the constituencies are apparent from the nodal structure.

Lets make trees for the following sentences:

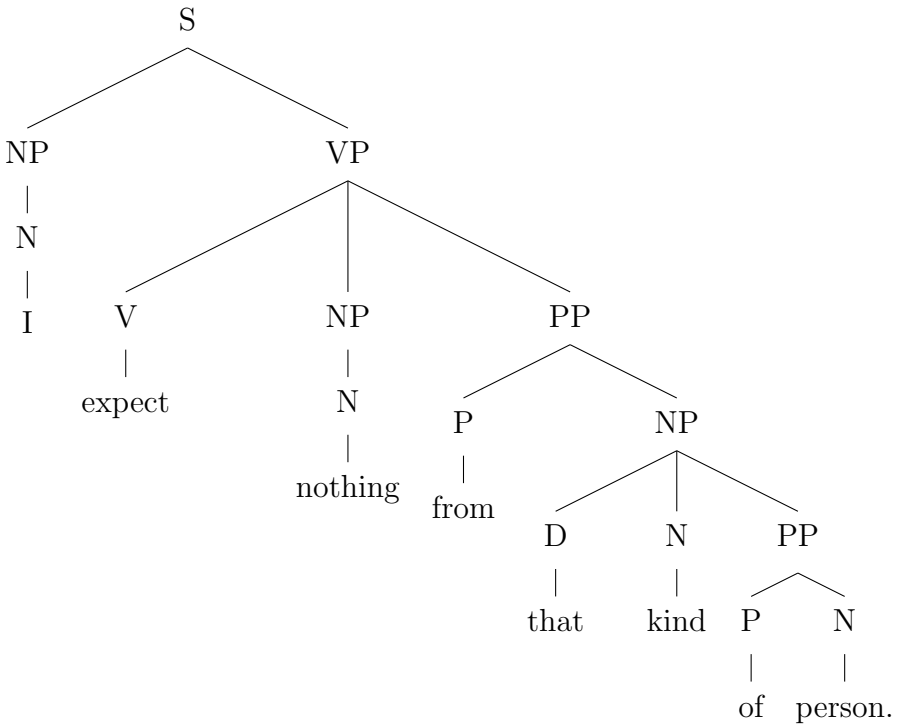
I expect nothing from that kind of person.

Lulu will sing a song during class.

Somebody from the main office was asking questions about her.

The dispute about her singing is affecting the school.

The first sentence can easily be created from our rules.



The second sentence is harder to construct since it has an auxiliary verb. We need to amend our verb phrase rule, but before doing so we must determine how many auxiliaries

can fit in a verb phrase. Here are some sentences with different amounts of auxiliaries:

I go to school.

I will go to school.

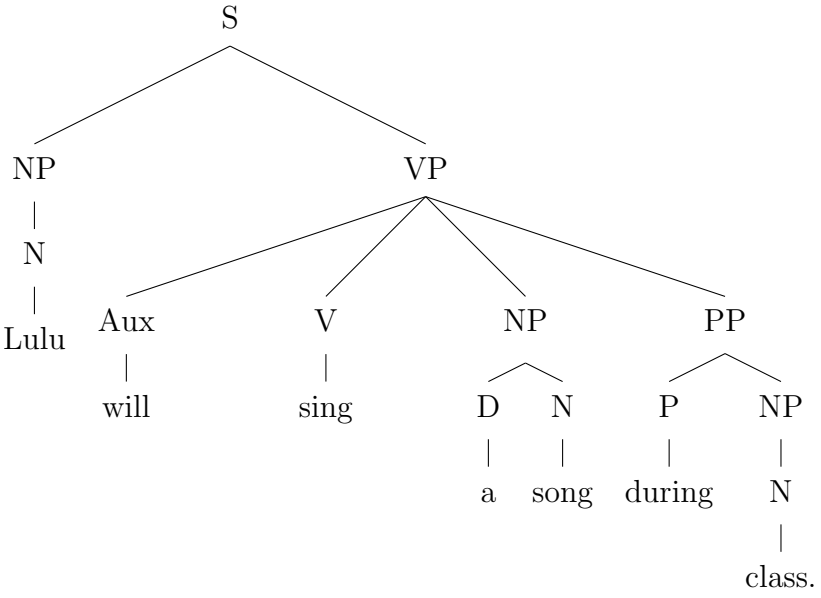
I would have gone to school.

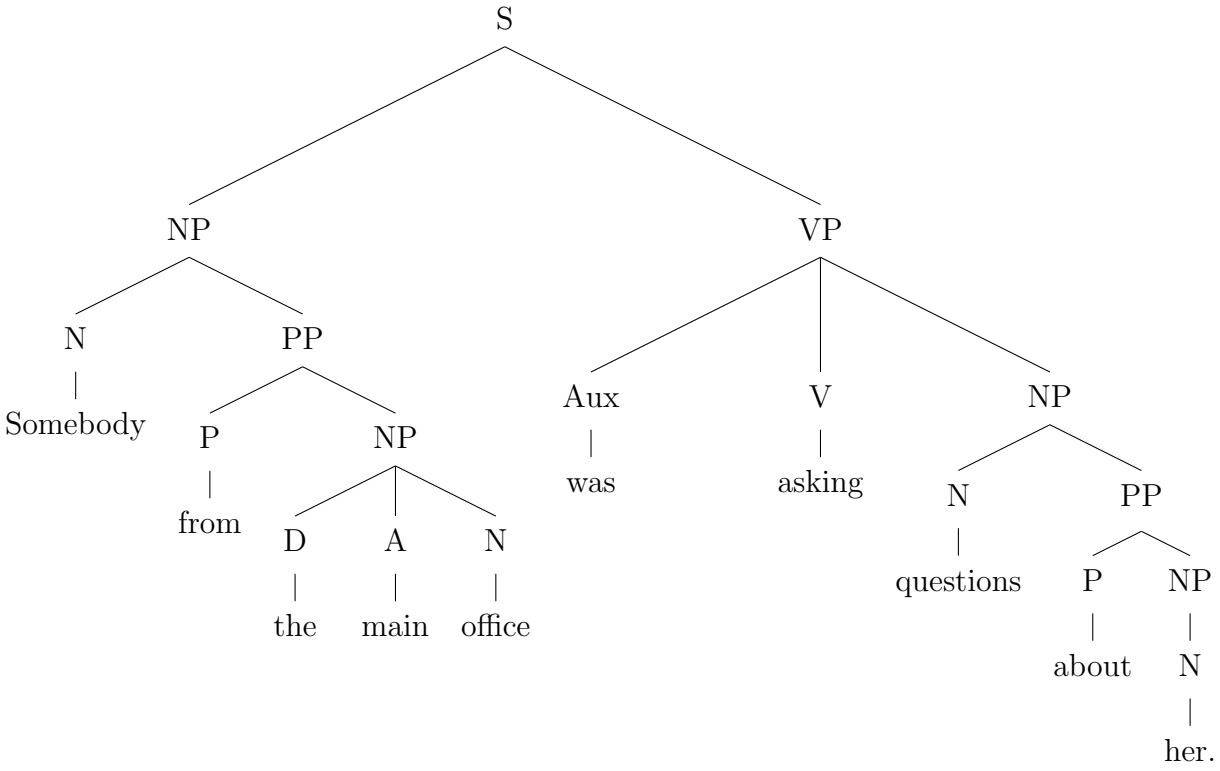
I would have been going to school.

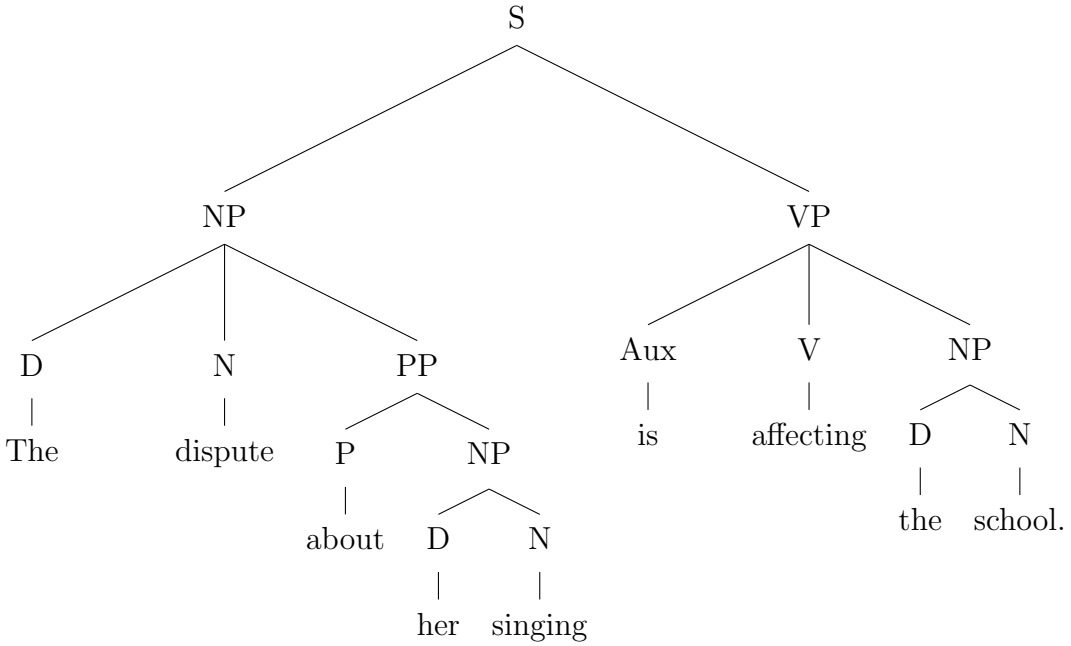
Let's fix the verb phrase rule and stipulate a maximum of three auxiliaries per phrase, as well as an arbitrary amount of prepositional phrases:

$$\text{VP} \rightarrow (\text{Aux})^3 \text{ V } (\text{NP})(\text{PP})^*$$

Back to the trees:







Let's recap out phrase structure rules, and allow for an arbitrary amount of adjectives

and prepositional phrases in noun phrases:

$$\begin{aligned}
&\text{S} \rightarrow \text{NP VP} \\
&\text{NP} \rightarrow \text{(D)(A)* N (PP)*} \\
&\text{VP} \rightarrow \text{(Aux)}^3 \text{ V (NP)(PP)*} \\
&\text{PP} \rightarrow \text{P NP}
\end{aligned}$$

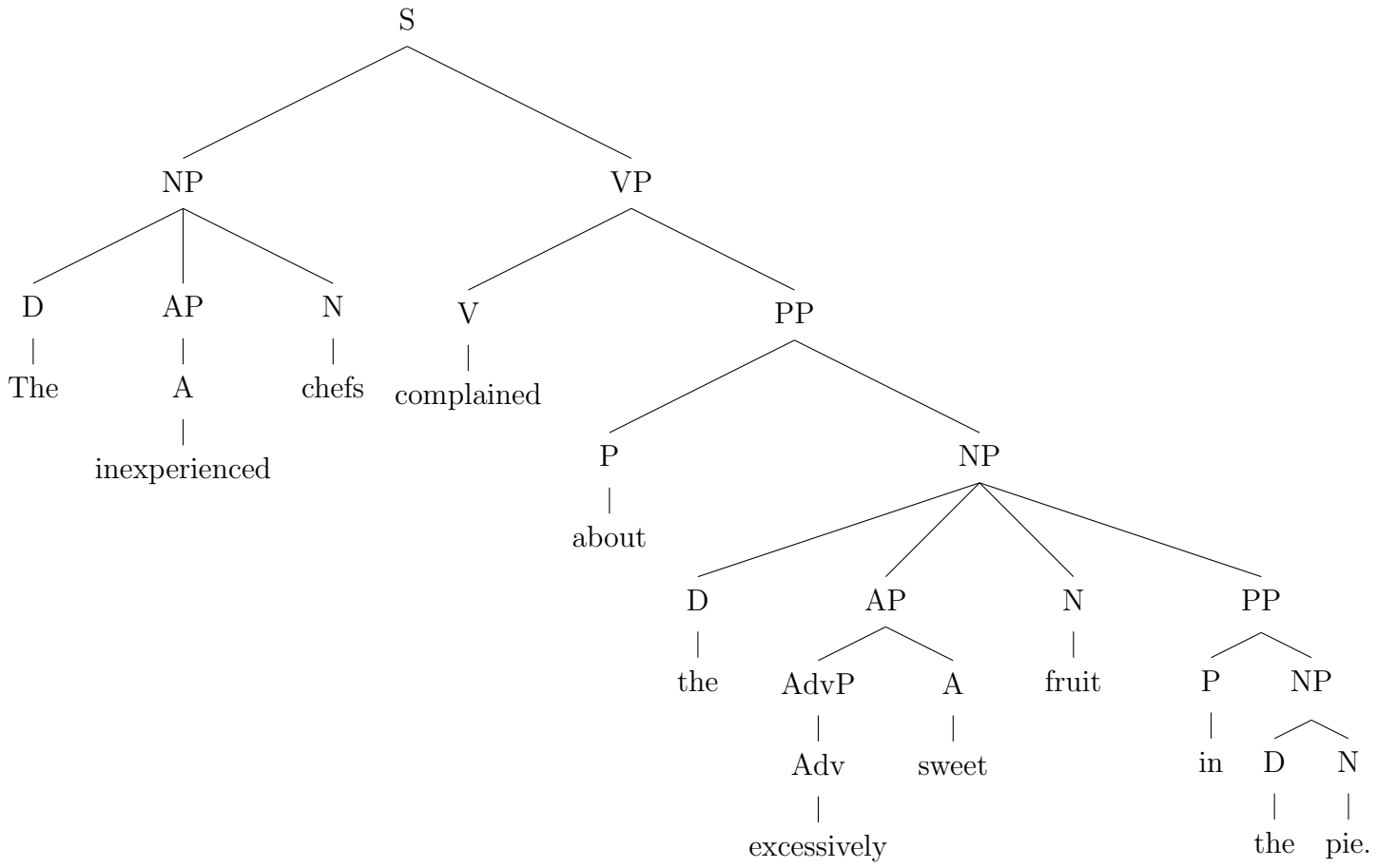
Some of the following sentences will require us to strengthen our rules one-by-one:

> The inexperienced chefs complained about the excessively sweet fruit in the pie.
>
> I will be meeting Lulu after the concert.
>
> I know that pig.
>
> I know that pigs fly.

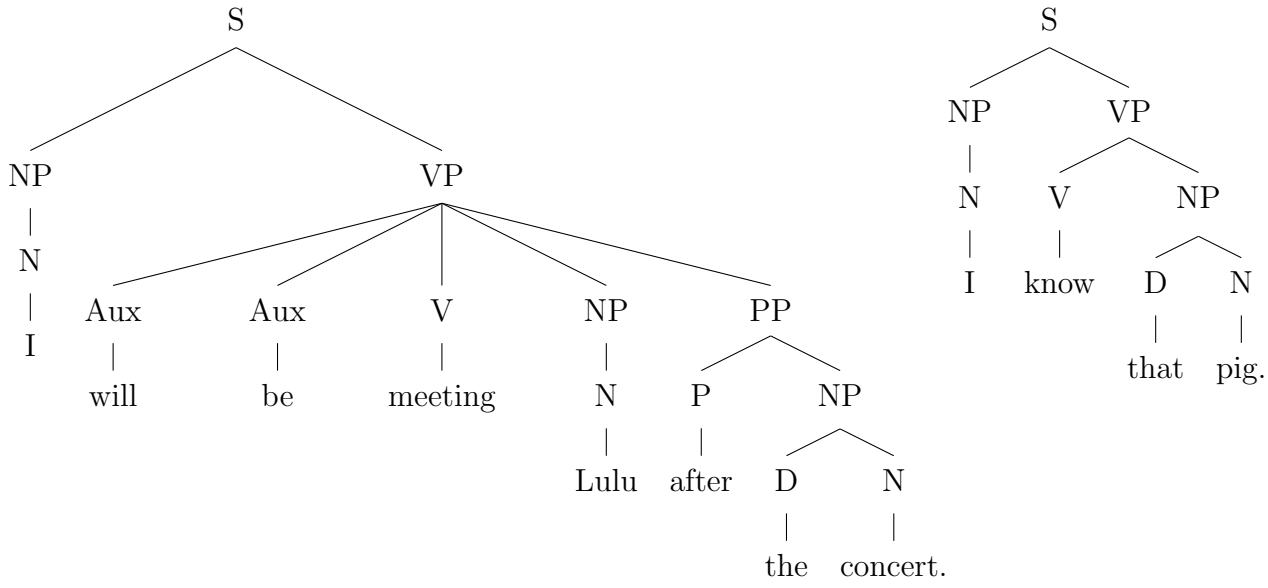
We currently have no way of dealing with adverbs such as "excessively." It seems that the phrase [excessively sweet] forms a constituency that is used to modify [fruit]. This constituent is an adjective phrase (AP) and is made of an optional adverb phrase (AdvP) and a non-optional adjective. Let's update our phrase structure rules to allow for adjective and adverb phrases:

$$\begin{aligned}
&\text{NP} \rightarrow \text{(D)(AP)* N (PP)*} \\
&\text{VP} \rightarrow \text{(Aux)}^3 \text{ V (NP)(PP)*[AdvP]} \\
&\text{AP} \rightarrow \text{(AdvP) A} \\
&\text{AdvP} \rightarrow \text{Adv}
\end{aligned}$$

It may seem strange that we are introducing an adverb phrase whose only constituent is a single adverb. Though the existence of an adverb phrase is unnecessary right now, there will be a need for adverb phrases in the future. In fact, it is beneficial to define a phrase for every given part of speech, and so it is good practice to define them early. That being said, we can now we can complete the tree.



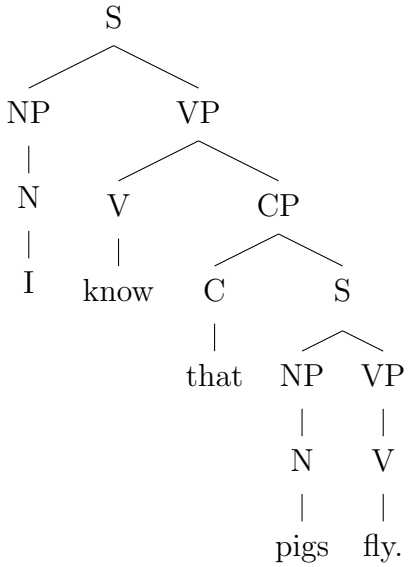
The next two trees are easy.



The next sentence contains a subordinate clause, so we need to introduce the complementizer

rule from Section 1.

$$\text{VP} \rightarrow (\text{Aux})^3 \text{ V } (\text{NP})(\text{CP})(\text{PP})^*[\text{AdvP}]$$
$$\text{CP} \rightarrow \text{C S}$$



These next sentences contain conjunctions, and will require another rule:

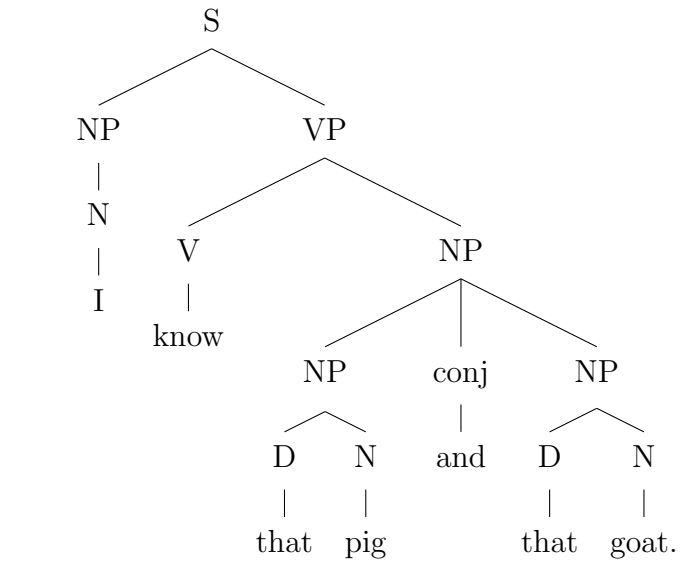
I know that pig and that goat.

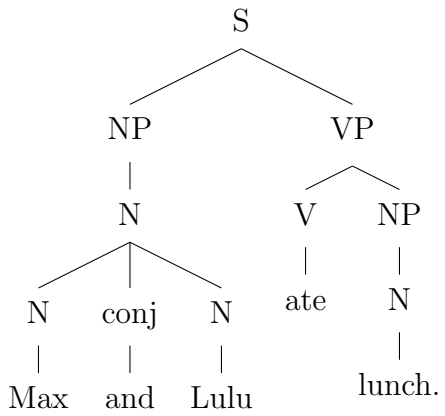
Max and Lulu ate lunch.

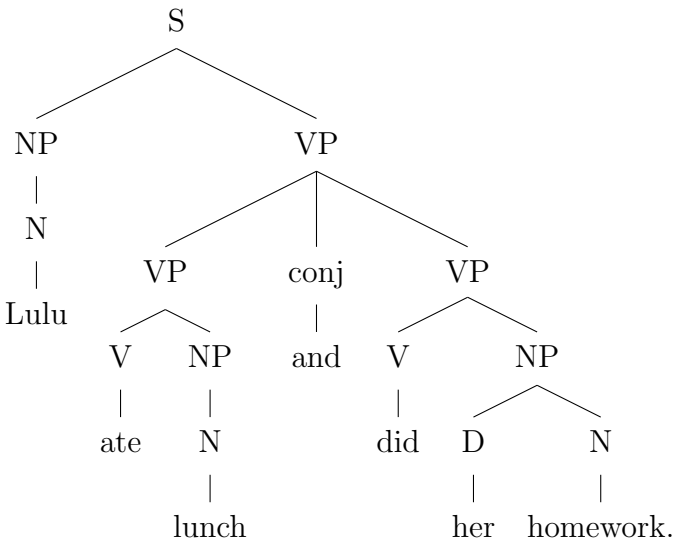
Lulu ate lunch and did her homework.

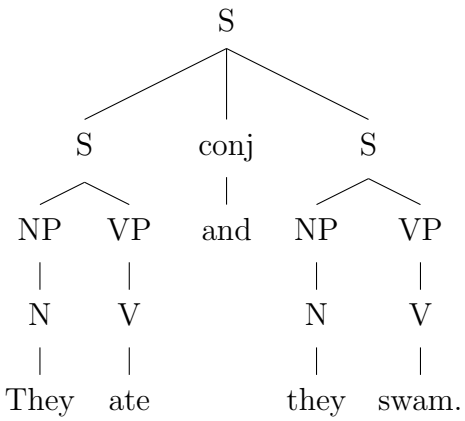
They ate and they swam.

$$\text{X} \rightarrow \text{X conj X}$$









How would we make a tree for the sentence "Lulu's pig can fly"? The possessive construction [Lulu's] appears to be one constituent, and it behaves like a determiner phrase because it tells us *which* pig can fly. Can this phrase be broken up? The study of syntax is focused on how discrete units of meaning combine with respect to formal rules to produce sentences. These units of meaning are usually the words themselves of the sentence, however in this case, [Lulu's] contains two units. For all intents and purposes, [Lulu] is a standalone unit that is being modified by the determiner ['s]. Affixes of this kind are often called particles, and some languages place a much larger emphasis on particles as their own part of speech (e.g. Japanese). We said earlier that every part of speech should have a corresponding

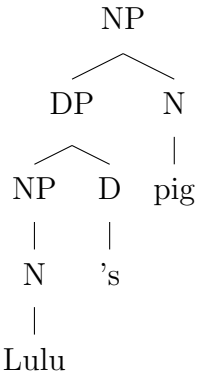
phrase, so now is a good time to introduce a determiner phrase.

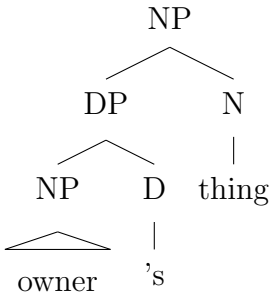
$$\text{NP} \rightarrow \text{(NP)(AP)* N (PP)*}$$
$$\text{DP} \rightarrow \text{(NP) D}$$

Using this framework, we can turn [Lulu's] into a determiner phrase as follows:

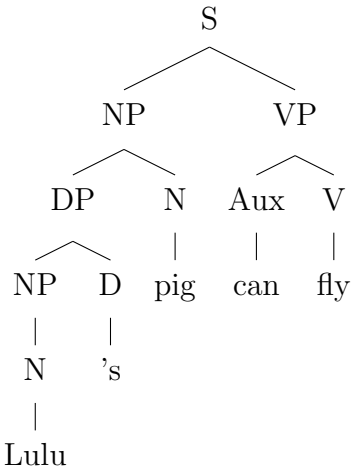
[DP [NP [N Lulu]] [D 's]]

This new determiner phrase then modifies [pig].

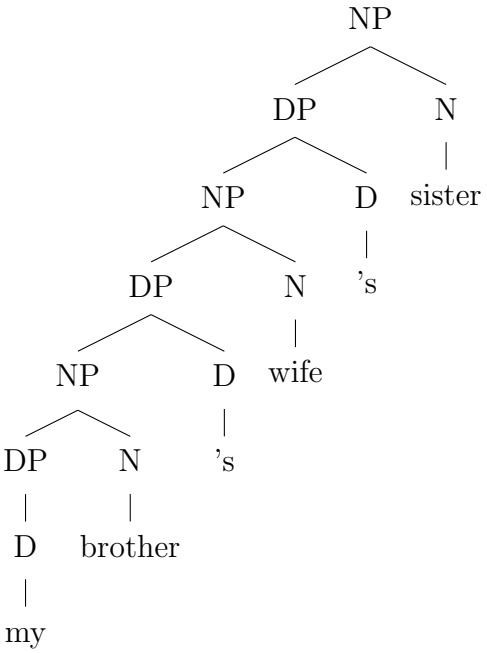




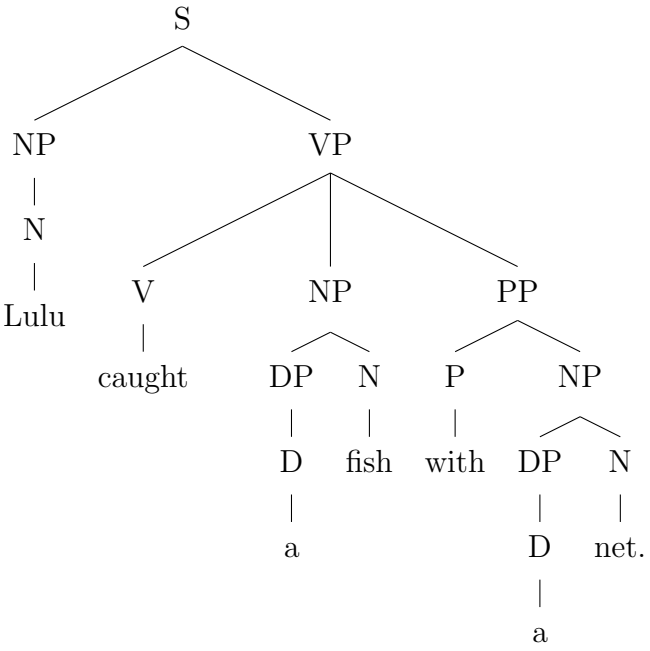
Here is the completed tree:

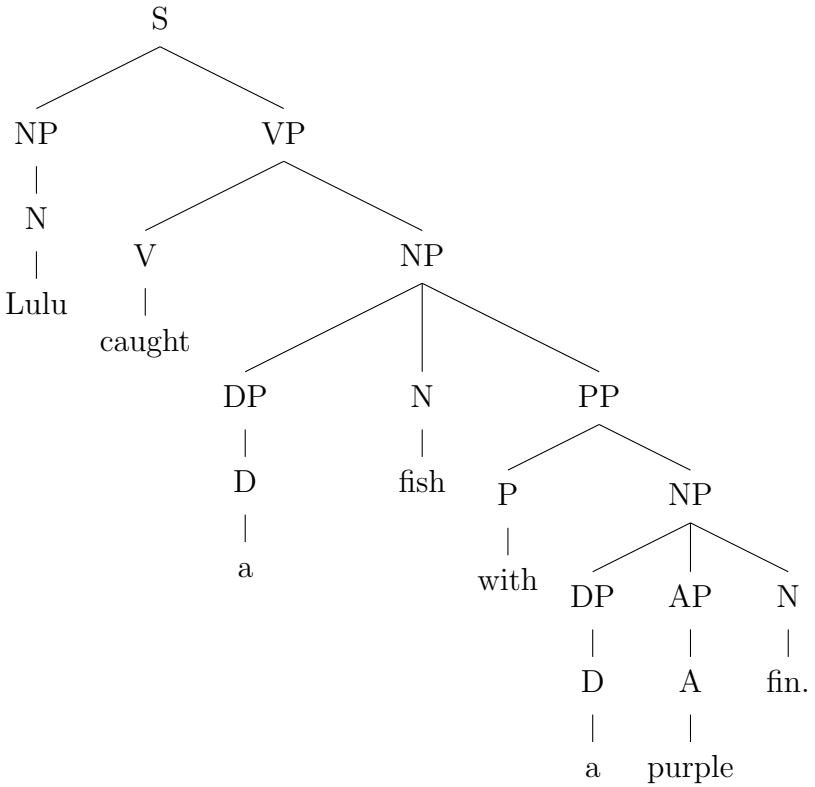


Compound possessives can get tricky because they involve nested determiner phrases:

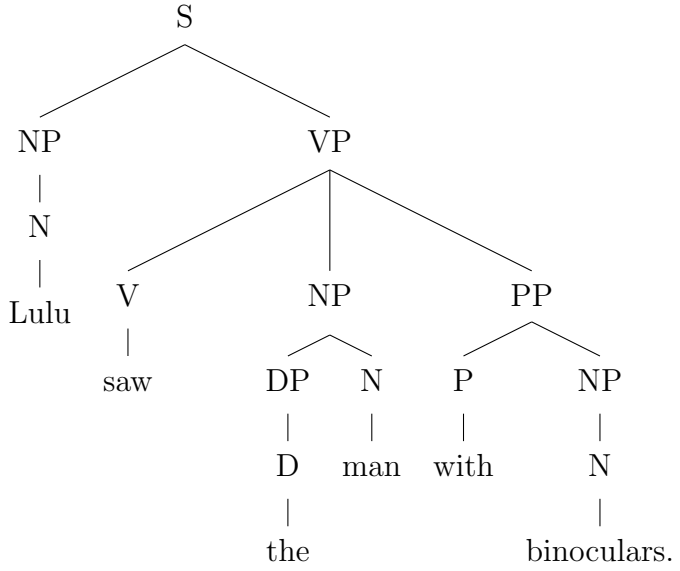


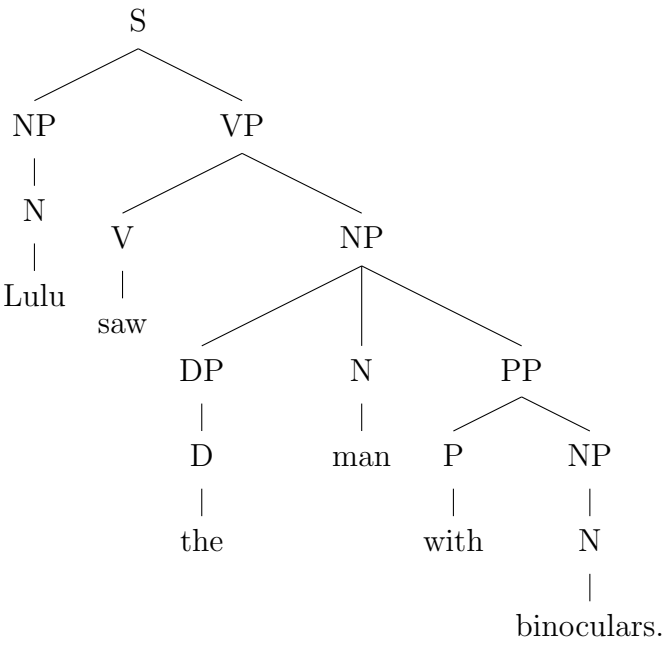
We will still use determiner phrase even when not dealing with possessives.





We said earlier that syntax trees can make clear any semantic ambiguity a sentence has. Consider the sentence "Lulu saw the man with binoculars." Two different trees for this sentence can be made.





The first tree refers to Lulu observing a man through her binoculars, while the second refers to Lulu observing a man who has binoculars.
Let's create a rule to deal with the following sentences:

Max gave Lulu the book.

Pizza is very delicious.

Currently, our verb phrase rule is used to deal with transitive verbs. These are verbs that act on a single object such as *kick, steal, want*, etc. The first sentence above contains a ditransitive verb, because it requires two arguments: a direct object and an indirect object. Notice how the removal of one of the arguments makes the sentence ungrammatical:

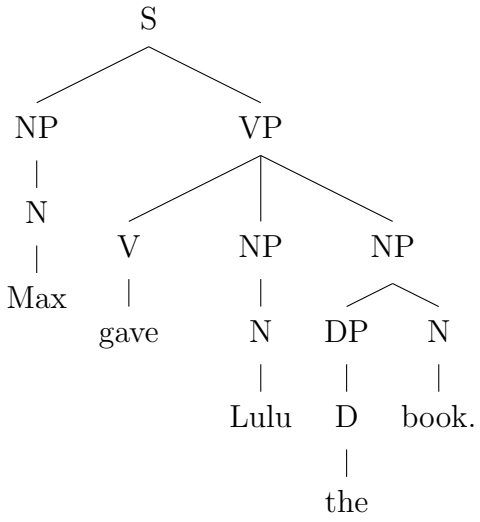
*Max gave Lulu.

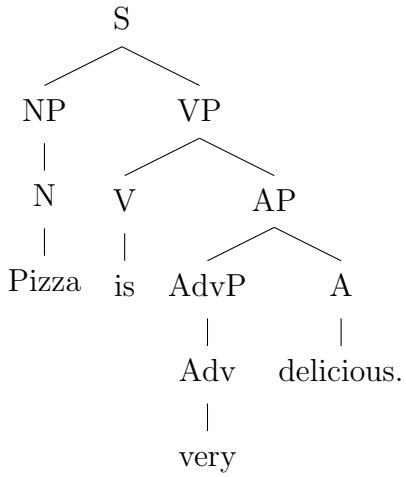
*Max gave the book.

The verb in the second sentence above is intransitive—it does not act on an object. We have to introduce a *second* verb phrase rule to deal with intransitive verbs. Both rules are stated below:

$$\text{VP} \rightarrow (\text{Aux})^3 \text{ V } (\text{NP})(\text{NP})(\text{PP})^*(\text{CP})[\text{AdvP}]$$

$$\text{VP} \rightarrow \text{V } (\text{AP})$$





We have accomplished a lot so far, so let's restate all of our rules in full before proceeding.

$S \rightarrow NP\ VP$

$NP \rightarrow (DP)(AP)^*\ N\ (PP)^*$

$DP \rightarrow (NP)\ D$

$AP \rightarrow (AdvP)\ A$

$AdvP \rightarrow Adv$

$VP \rightarrow (Aux)^3\ V\ (NP)(NP)(PP)^*(CP)[AdvP]$

$VP \rightarrow V\ (AP)$

$PP \rightarrow P\ NP$

$CP \rightarrow C\ S$

$X \rightarrow X\ conj\ X$

4 Structural Relations

5 Binding Theory

6 X-Bar Theory

7 Extending X-Bar Theory to Functional Categories

8 Constraining X-Bar Theory: Theta Theory

9 Theta Grids and Functional Categories

10 Head-to-Head Movement

11 DP Movement

12 *Wh*-movement and Locality Constraints

13 A Unified Theory of Movement