جامعة نيويورك أبوظبي

NYU | ABU DHABI

# PSYCH-UH 2218: Language Science

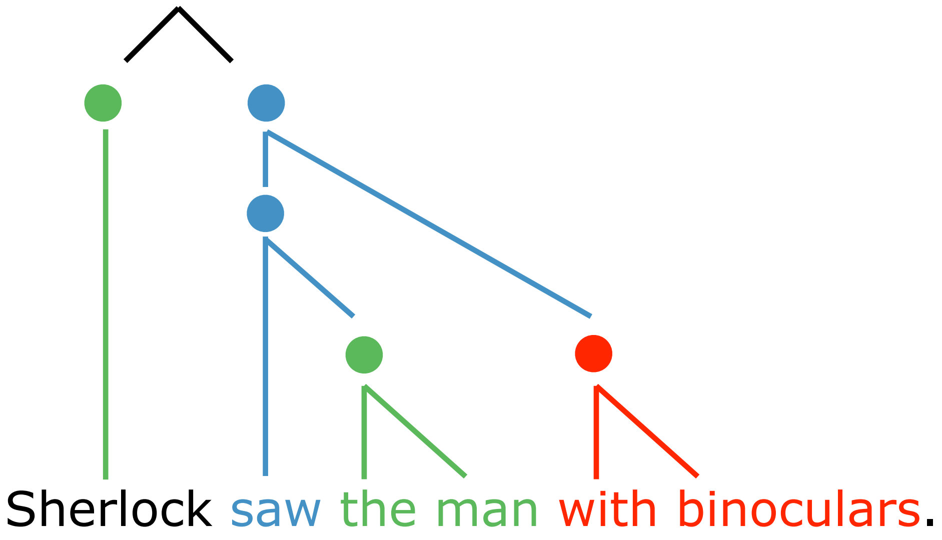
## Class 15: Syntax - phrase structure rules

Prof. Jon Sprouse
Psychology

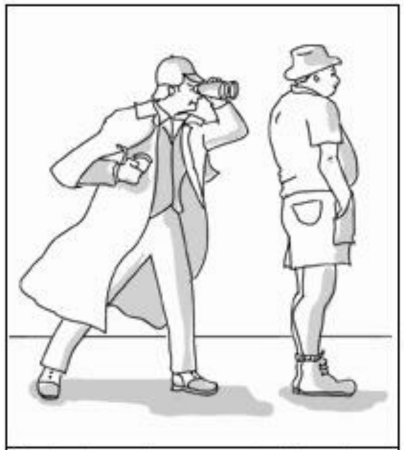
# What is our goal?

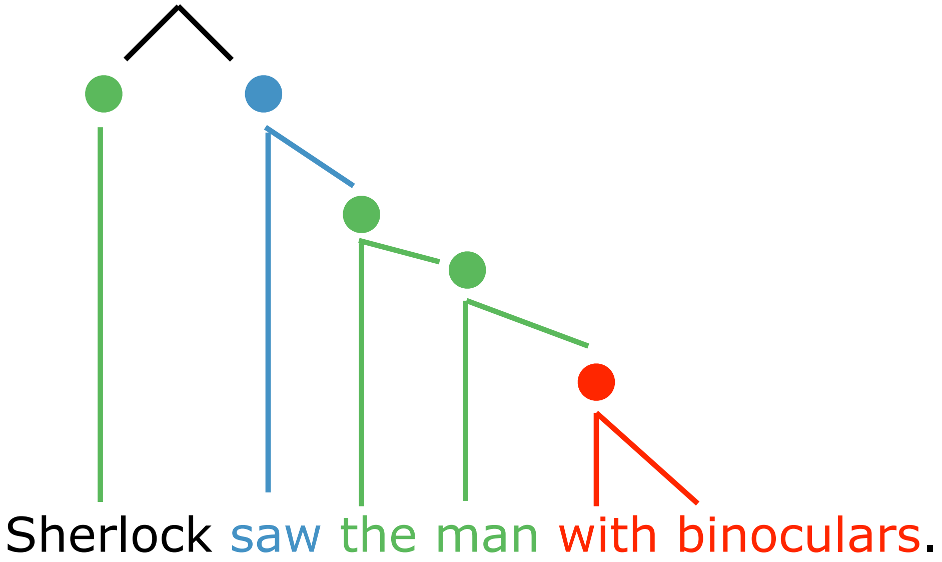
# Remember: ambiguity suggests hierarchical structure

Remember that earlier we saw that sentences can be ambiguous. We know from our work in morphology that this suggests **hierarchical structure**:

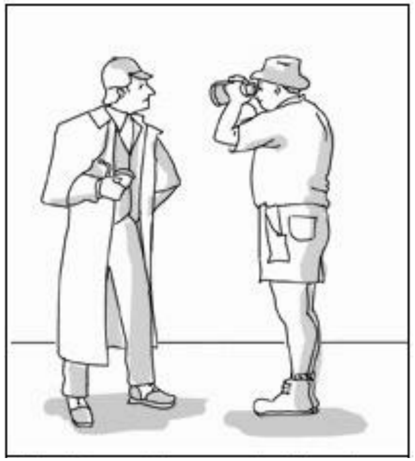


Sherlock saw the man with binoculars.





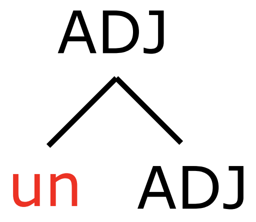
Sherlock saw the man with binoculars.



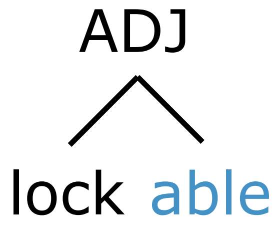
# Remember: hierarchical structure comes from structure-building rules

Remember that in morphology we saw that every "triangle" in our hierarchical structure comes from a structure-building rule:

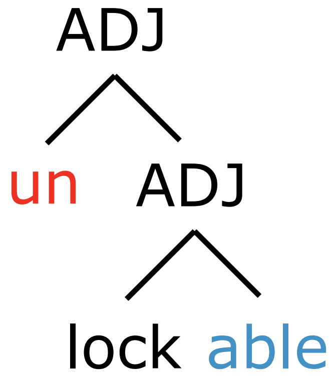
**ADJ → un + ADJ**



**ADJ → VERB + able**



And by applying these rules, we can derive the hierarchical structure for a multi-morphemic word:



Well, we want to do the same thing for syntax: find the rules that will give us the hierarchical structure!

# We want to find the structure-building rules for syntax. We call them phrase-structure rules.

We've already seen structure-building rules for creating complex words. Now let's try to come up with some structure building rules for constructing sentences.

The first step is to label the syntactic categories of words. Syntactic categories are the units that will go into our rules. **This is the abstractness that we need to handle novel sentences!**

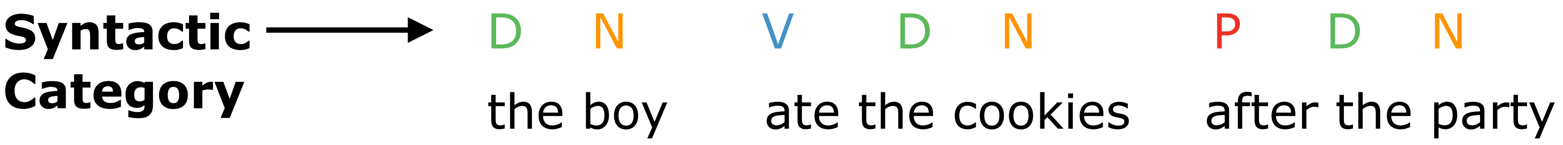
D = determiner

N = noun

V = verb

P = preposition

To save space, we can use the first letter of each syntactic category instead of the full name. No big deal.



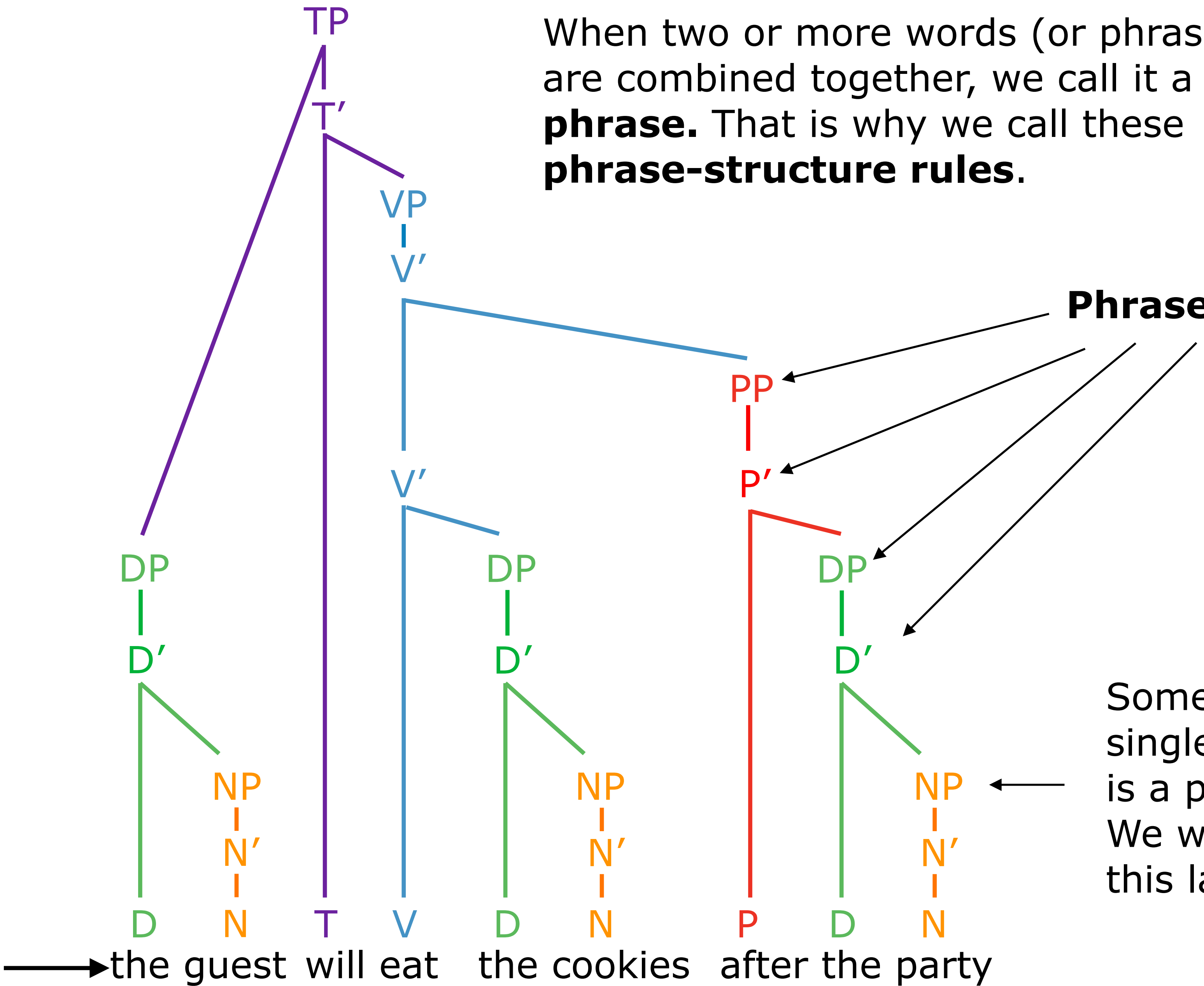
**Syntactic Category**

D N V D N P D N

the boy ate the cookies after the party

# We want to find the structure-building rules for syntax. We call them phrase-structure rules.

When two or more words (or phrases) are combined together, we call it a **phrase.** That is why we call these **phrase-structure rules**.



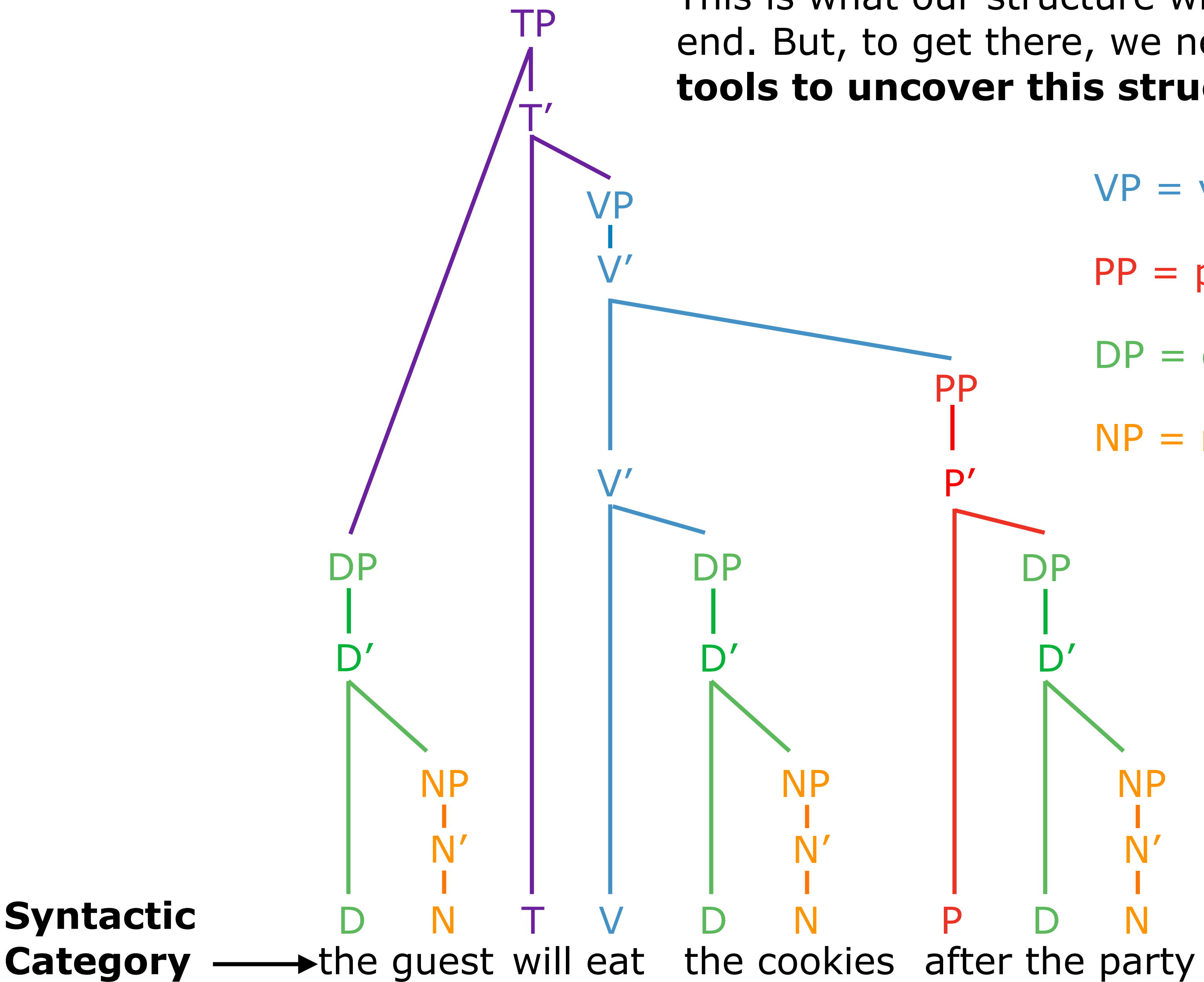
**Phrases**

Sometimes a single word is a phrase. We will see this later.

**Syntactic Category**

# We want to find the structure-building rules for syntax. We call them phrase-structure rules.

This is what our structure will look like in the end. But, to get there, we need **to develop tools to uncover this structure**.



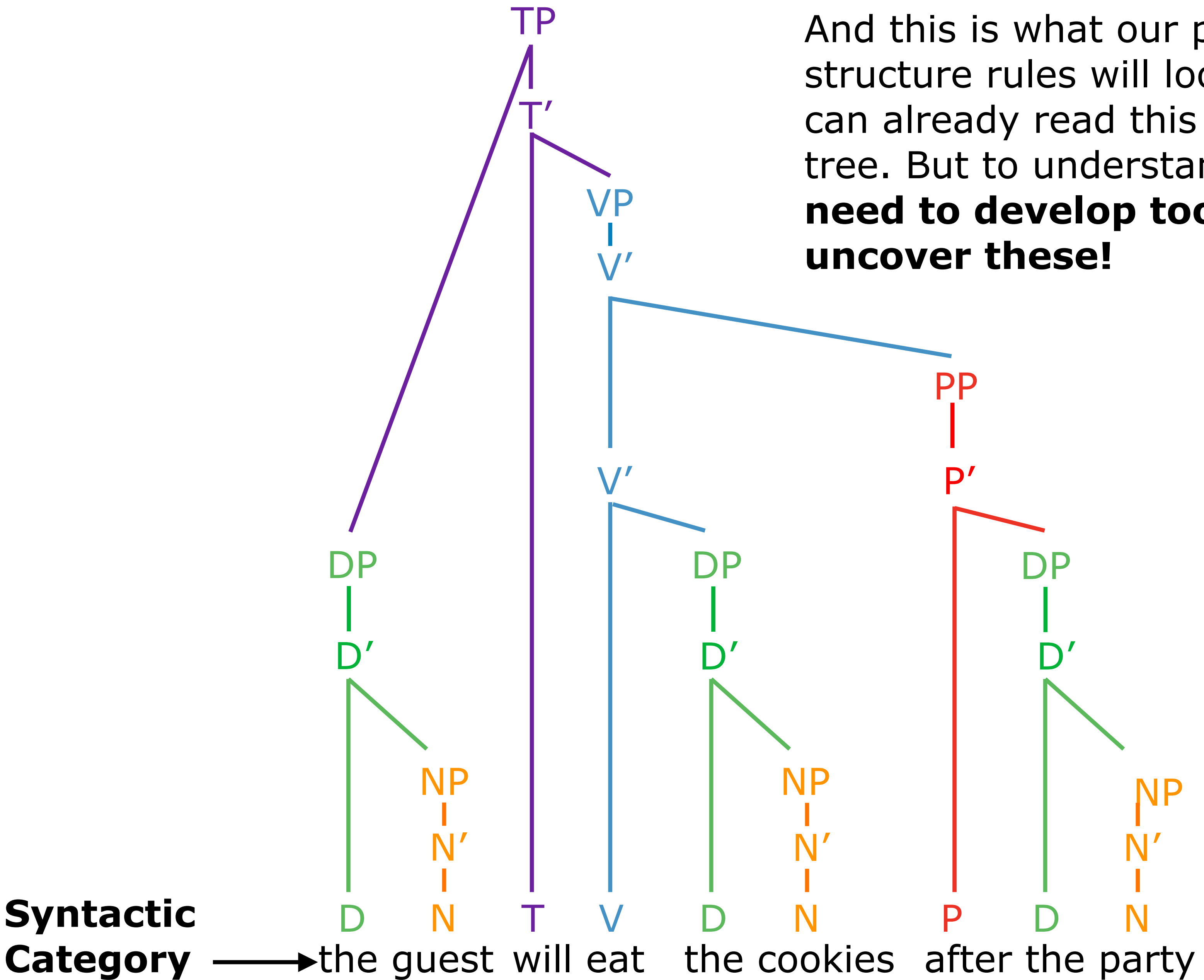
VP = verb phrase

PP = prepositional phrase

DP = determiner phrase

NP = noun phrase

# We want to find the structure-building rules for syntax. We call them phrase-structure rules.



And this is what our phrase structure rules will look like. You can already read this off of the tree. But to understand it, **we need to develop tools to uncover these!**

TP → DP T′

T′ → T VP

VP → V′

V′ → V′ PP

V′ → V DP

DP → D′

D′ → D NP

NP → N′

N′ → N

# Uncovering phrase structure rules with constituency tests

# What is a constituent?

A constituent is an item or items that combine to form a larger unit.

Psychologically, we would say that a constituent is a group of items that show some sort of behavior that suggests they act together as a unit.

Those behaviors can be turned into a diagnostic called a constituency test. The idea is that if a group of words act like a unit, we should be able to see consequences of this!

Here is a very straightforward example:

The guest will eat the cookies after the party.

They will eat the cookies after the party.

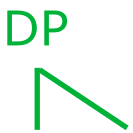
The words “the guest” behave as a unit in that they can be replaced together by another word — in this case, the pronoun “they”. That is it. That is an example of showing a behavior that suggests that they form a unit.

# Representing constituents with trees

The trees that we draw (for words or sentences) show us the constituents.

The big idea is that words that form a constituent must all be "under" a single node in the tree.

In the simplest case, maybe the words all have branches that go to the same node:



DP

The guest will eat the cookies after the party.

This little bit of tree structure tells us that "the guest" forms a unit called a DP (for Determiner Phrase). And, since "they" can replace that bit of structure, we say that "they" is also a DP, and therefore can replace DPs:

DP

They will eat the cookies after the party.

# Our first constituency test - substitution

We just saw our first constituency test. We call it the substitution test.

The logic of the substitution test is as follows:

If a string of words can be replaced by another word (or string of words), then we can conclude that first string of words is a constituent:

The guest will eat the cookies after the party.

They will eat the cookies after the party.

"The guest" is a constituent.

If a string of words cannot be replaced by another word (or string of words), then we cannot conclude anything at all:

The guest will eat the cookies after the party.

* They the cookies after the party.

"The guest" is a constituent.

# Why is failure inconclusive?

The only tricky bit about constituency tests is that the failure of a constituency test is inconclusive.

The reason is because constituency tests tend to have more than one requirement. For example, the substitution test has two requirements:

1. The string must be a constituent. (This is why it is a constituency test!)

2. The replacement item must be the same type of phrase.

Here is an example where (2) is violated. We already know “the guest” is a constituent. But if I try to replace it with “then” instead of “they”, the sentence is ungrammatical:

The guest will eat the cookies after the party.

* Then will eat the cookies after the party.

The reason this is ungrammatical is that “then” replaces phrases that we would probably call “prepositional phrases” or PPs. So, when a constituency test fails, we can’t conclude anything - it could be because it is not a constituent, or it could be because another requirement is not met.

# More substitutions in English

Most languages have quite a few words or phrases that can substitute for others. Here are some that work in our sentence in English:

The guest will eat the cookies after the party.

The guest will eat the cookies it.

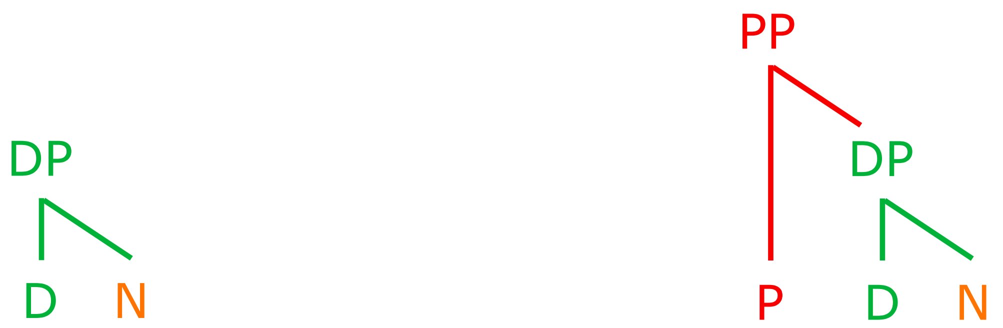
The party is a DP.

The guest will eat the cookies after the party.

The guest will eat the cookies then.

After the party is a PP.

Taken together, these two suggest the following hierarchical structure:



The guest will eat the cookies after the party.

This says that “after the party” is a constituent of type PP, and “the party” is a constituent of type DP.

# A second constituency test - stand alone

If a string of words can stand alone as an utterance, perhaps as an answer to a question, then it is a constituent:

The guest will eat the cookies after the party.

| Question | Answer | Category |
|---|---|---|
| What will the guest do? | eat cookies after the party | VP |
| When will the guest do it? | after the party | PP |
| What will the guest eat? | the cookies | DP |
| Who will eat cookies? | the student | DP |

In many languages, the stand-alone test is fairly straightforward, and can identify a variety of constituent types. The trick is to manipulate the question.

# A third constituency test - coordination

Coordination structures are created by a category of word that is sometimes called a **conjunction** and sometimes called a **coordinator**. The famous ones in English are and, or, but.

| **Constituent** | **Example** |
|---|---|
| DP | Lisa and her friends from Maryland played the game. |
| VP | The doctor opened the vial and tested the syringe. |
| PP | The bank is beside the post office and in front of the cafe. |
| AdjP | The students were very happy but also partially disappointed. |

Coordination appears to be possible if the following two conditions are met:

1. The two strings that are coordinated must both be constituents.
2. The two constituents that are coordinated must both be the same category.

# A third constituency test - coordination

It is very easy to see how coordination can be a very powerful constituency test. If a coordination construction is grammatical, you know that:

1. The two strings that are coordinated must both be constituents.
2. The two constituents that are coordinated must both be the same category.

And, crucially, it applies to the last constituent in our examples sentence involving the word “will” (the last one to be included!).

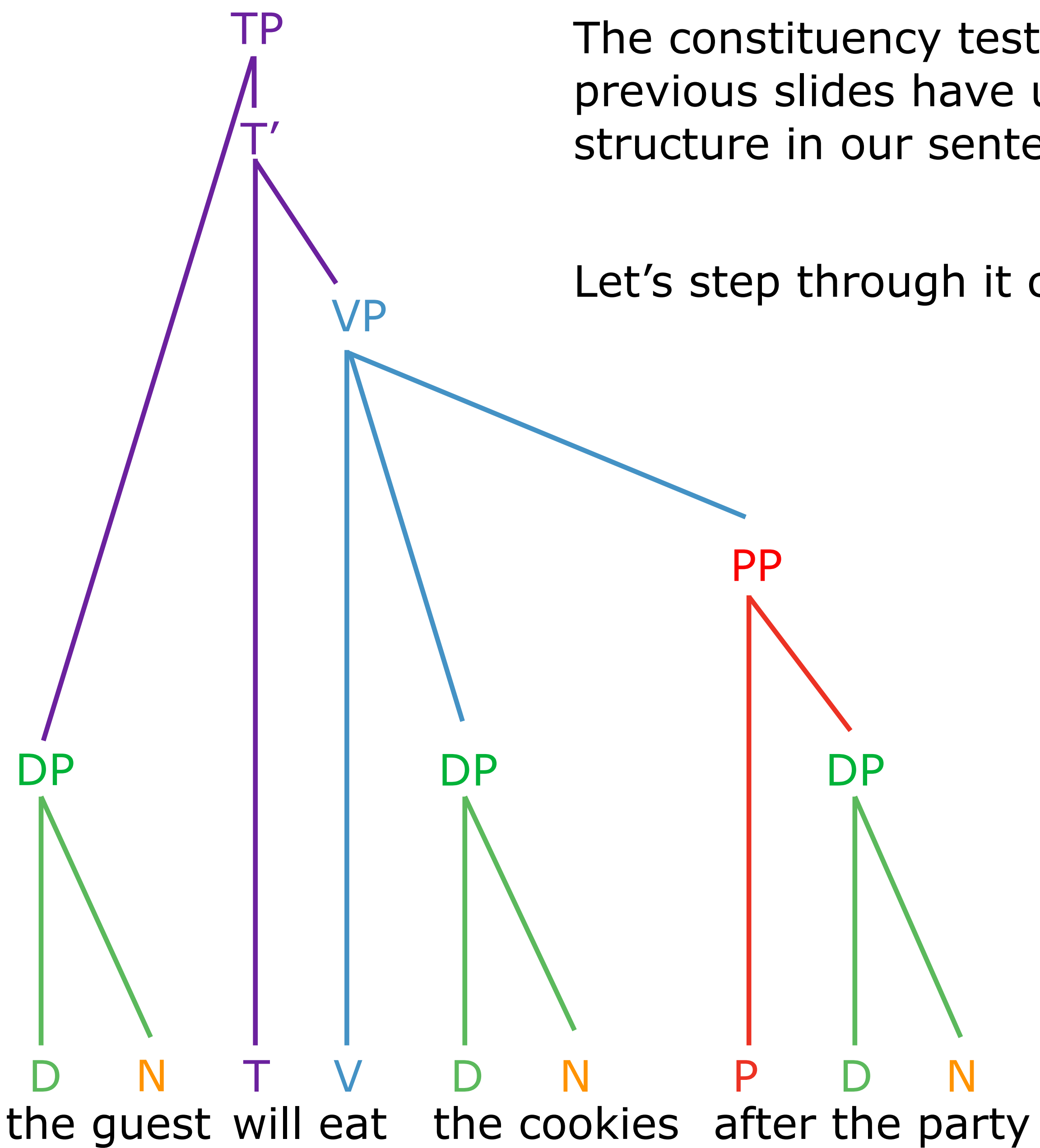
T’: The guest will eat the cookies after the party **and**
might drink the milkshake the next day.

Since this coordination is grammatical, we know that will eat the cookies after the party is a constituent (and that it is the same type as might drink the milkshake the next day).

# Other constituency tests

There are many other constituency tests. If you were to take a longer course in syntax, you would see many more. But, substitution, stand alone, and coordination are more than enough to start analyzing your native language!

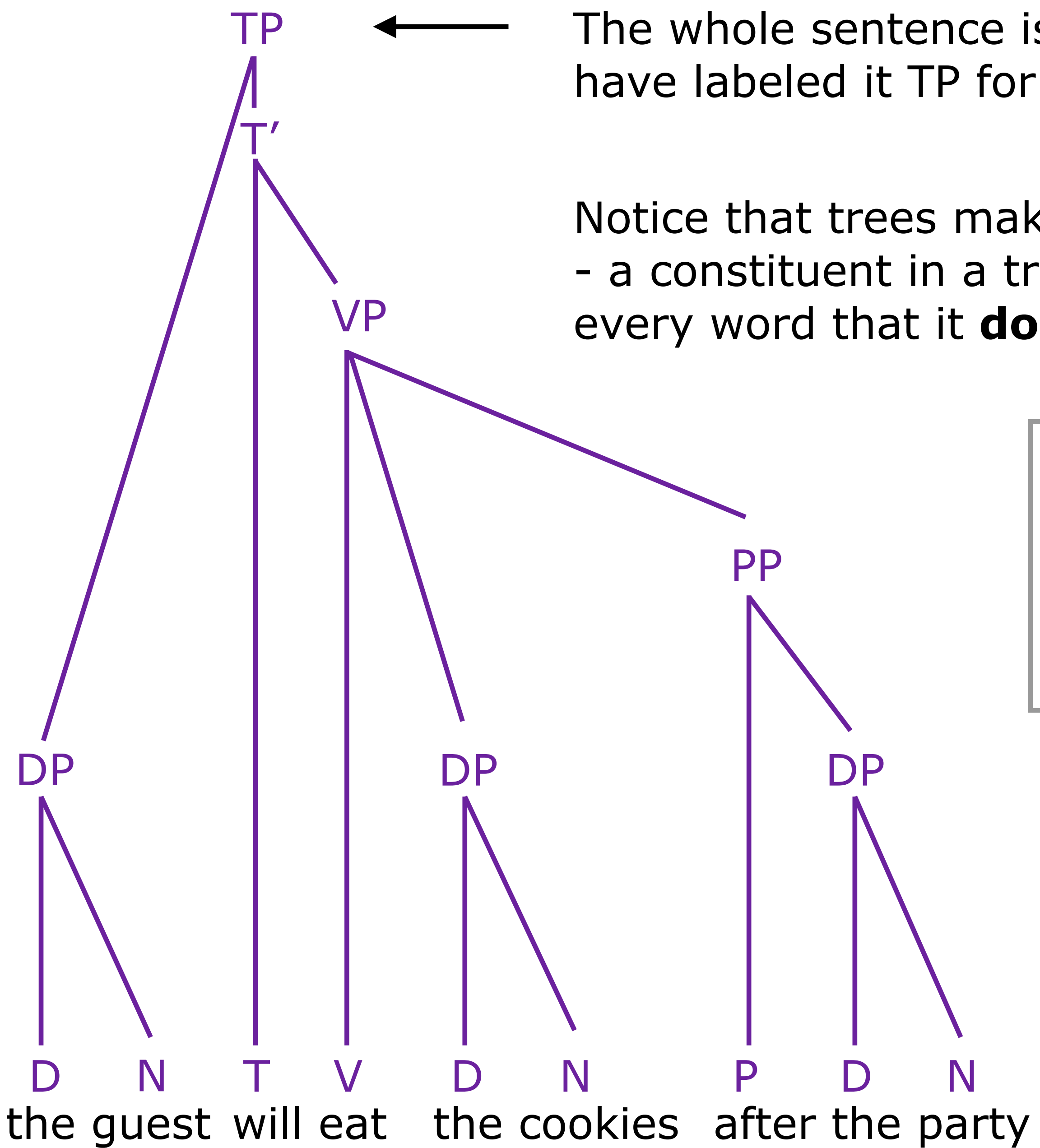
# Here is what our constituency tests tell us so far



The constituency tests that we used in the previous slides have uncovered quite a bit of structure in our sentence.

Let’s step through it constituent by constituent.

# Here is what our constituency tests tell us so far

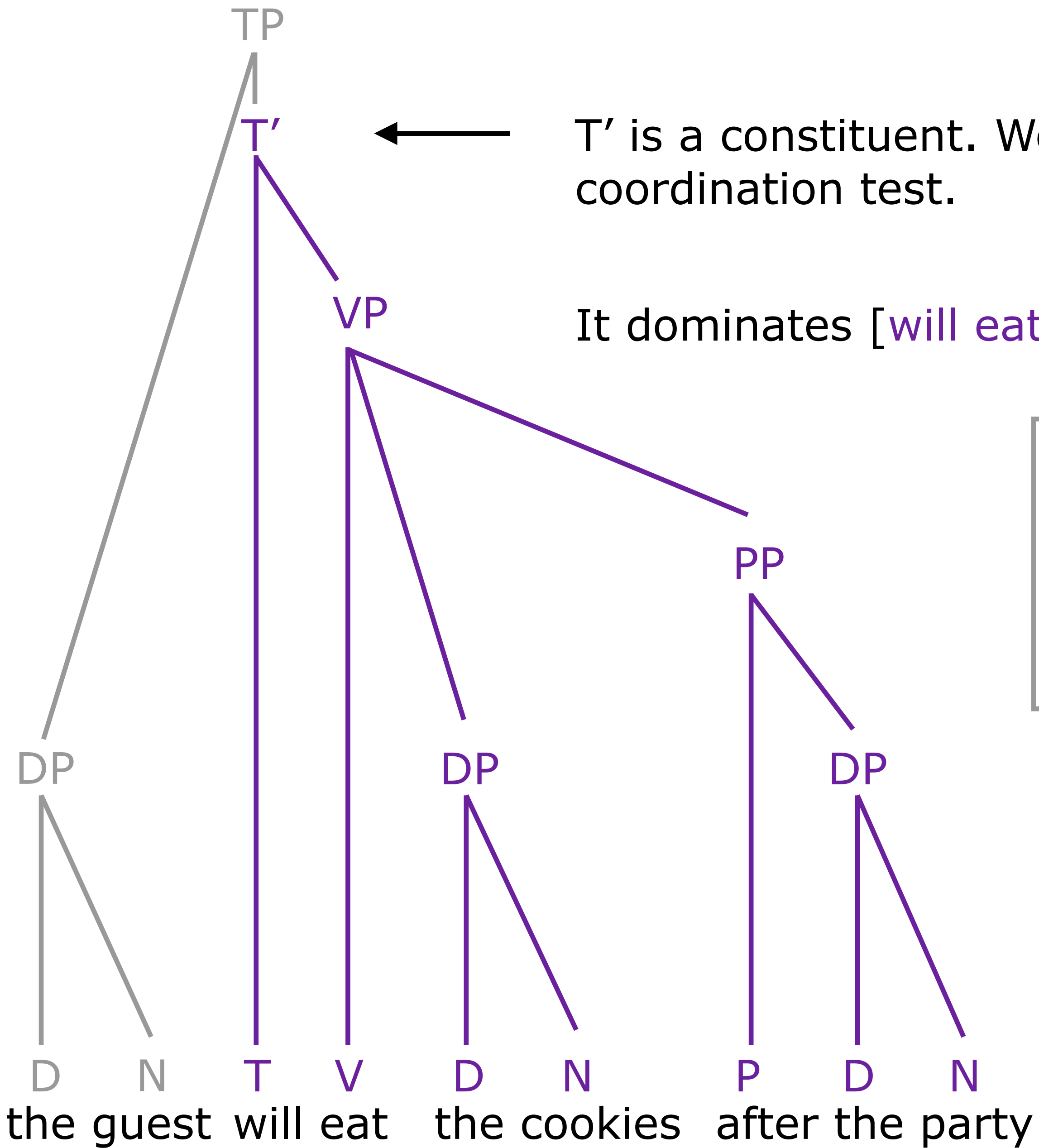


The whole sentence is obviously a constituent. I have labeled it TP for Tense Phrase.

Notice that trees make it easy to see constituents - a constituent in a tree is a node (like TP) and every word that it **dominates**.

We say that a node A **dominates** another node B if you can start at A and get to B by only going **down** the tree.

# Here is what our constituency tests tell us so far



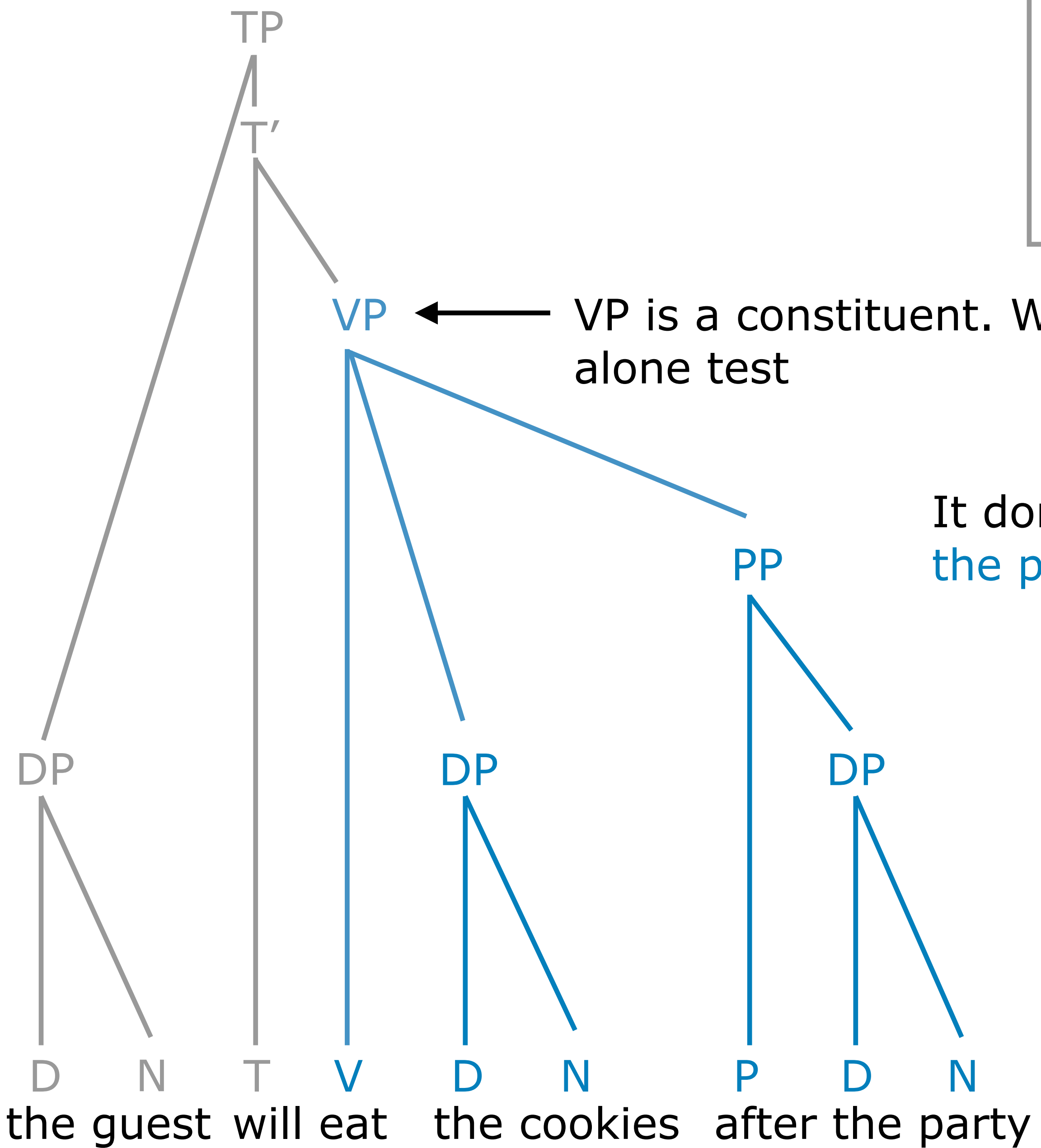
T′ is a constituent. We saw this by the coordination test.

It dominates [will eat the cookies after the party].

We say that a node A **dominates** another node B if you can start at A and get to B by only going **down** the tree.

# Here is what our constituency tests tell us so far

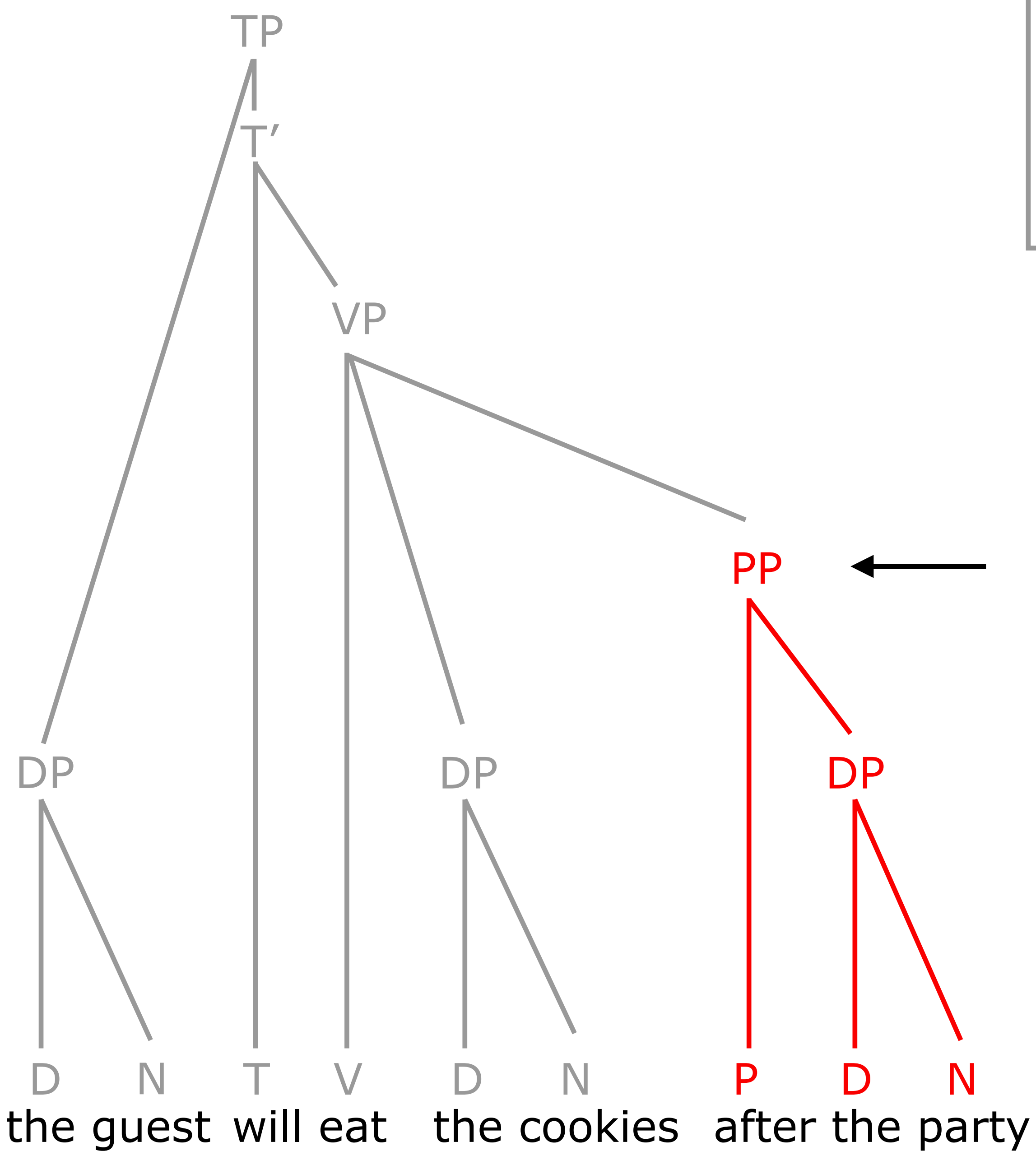
We say that a node A **dominates** another node B if you can start at A and get to B by only going **down** the tree.



VP is a constituent. We saw this by the stand alone test

It dominates [eat cookies after the party]

# Here is what our constituency tests tell us so far

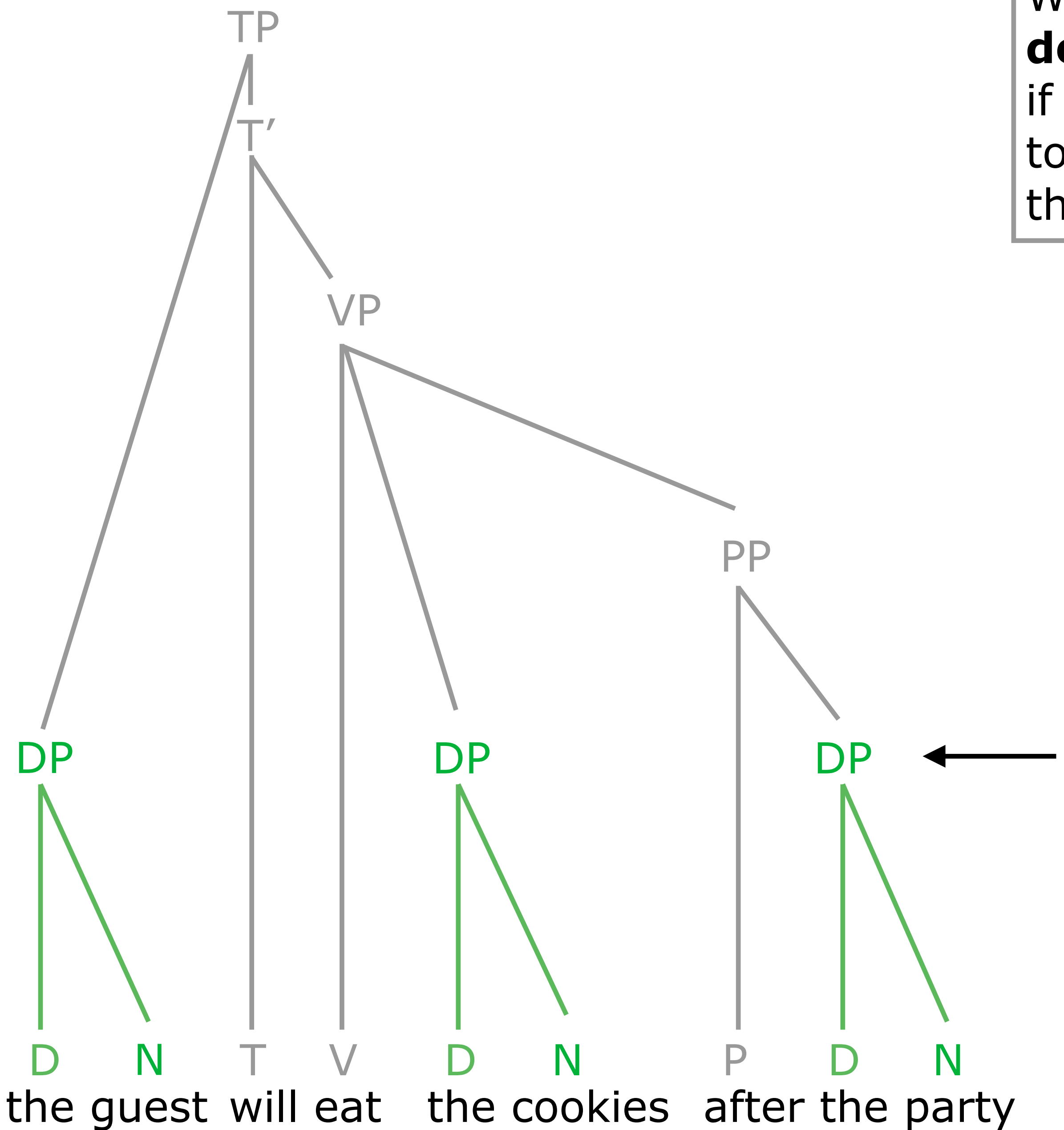


We say that a node A **dominates** another node B if you can start at A and get to B by only going **down** the tree.

PP is a constituent. We saw this by the substitution test and the stand alone test.

It dominates [after the party]

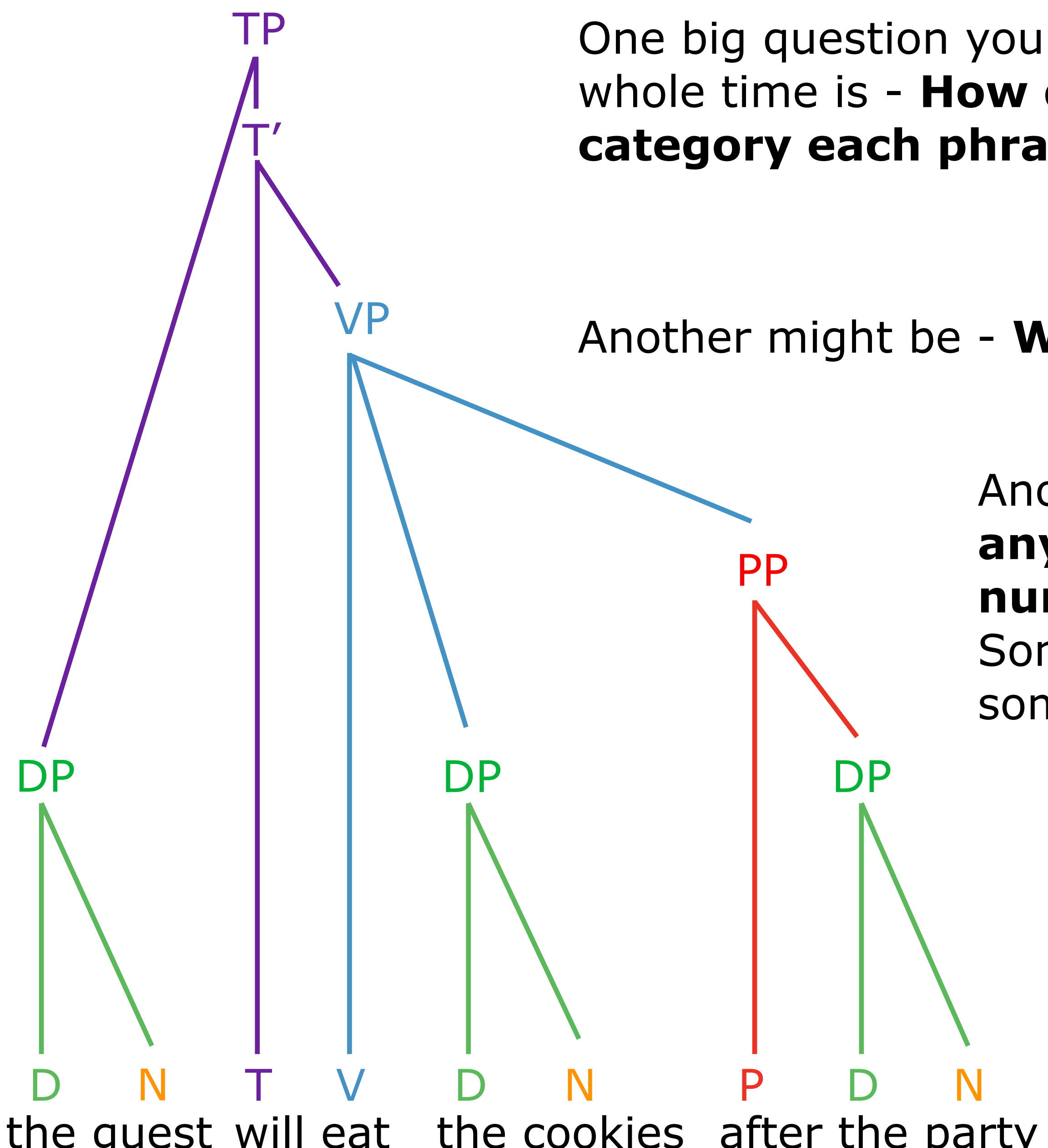
# Here is what our constituency tests tell us so far



We say that a node A **dominates** another node B if you can start at A and get to B by only going **down** the tree.

Each of the DPs are constituents. We saw this by the substitution and stand alone tests.

# Some unanswered questions!



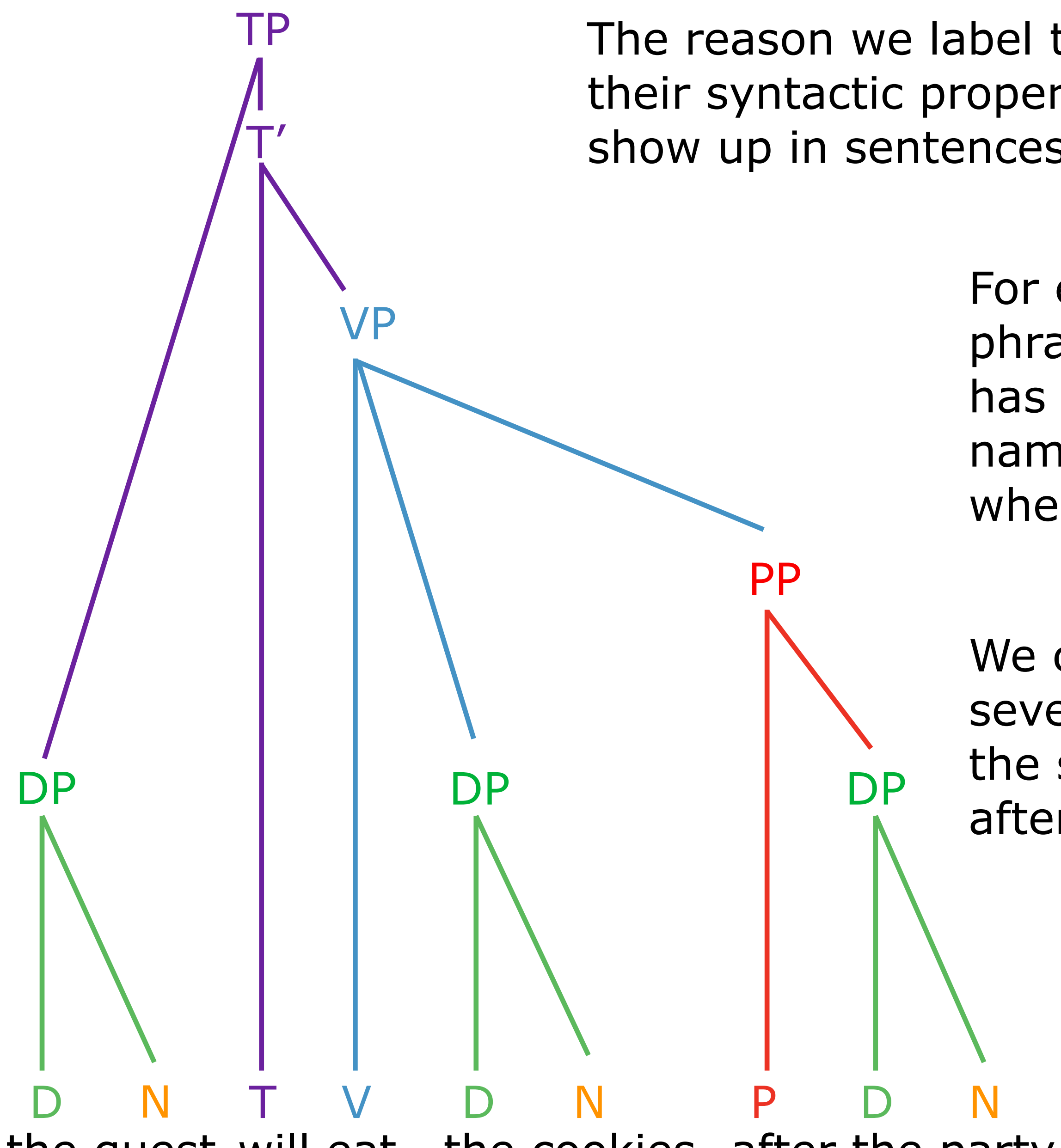
One big question you have probably had this whole time is - **How do we decide what category each phrase is?**

Another might be - **What is T’?**

Another might be - **Is there any restriction on the number of branches?** Sometimes it is two, and sometimes it is three?

# The properties of phrases

# Phrases that have the **same category** can appear in the **same syntactic position**

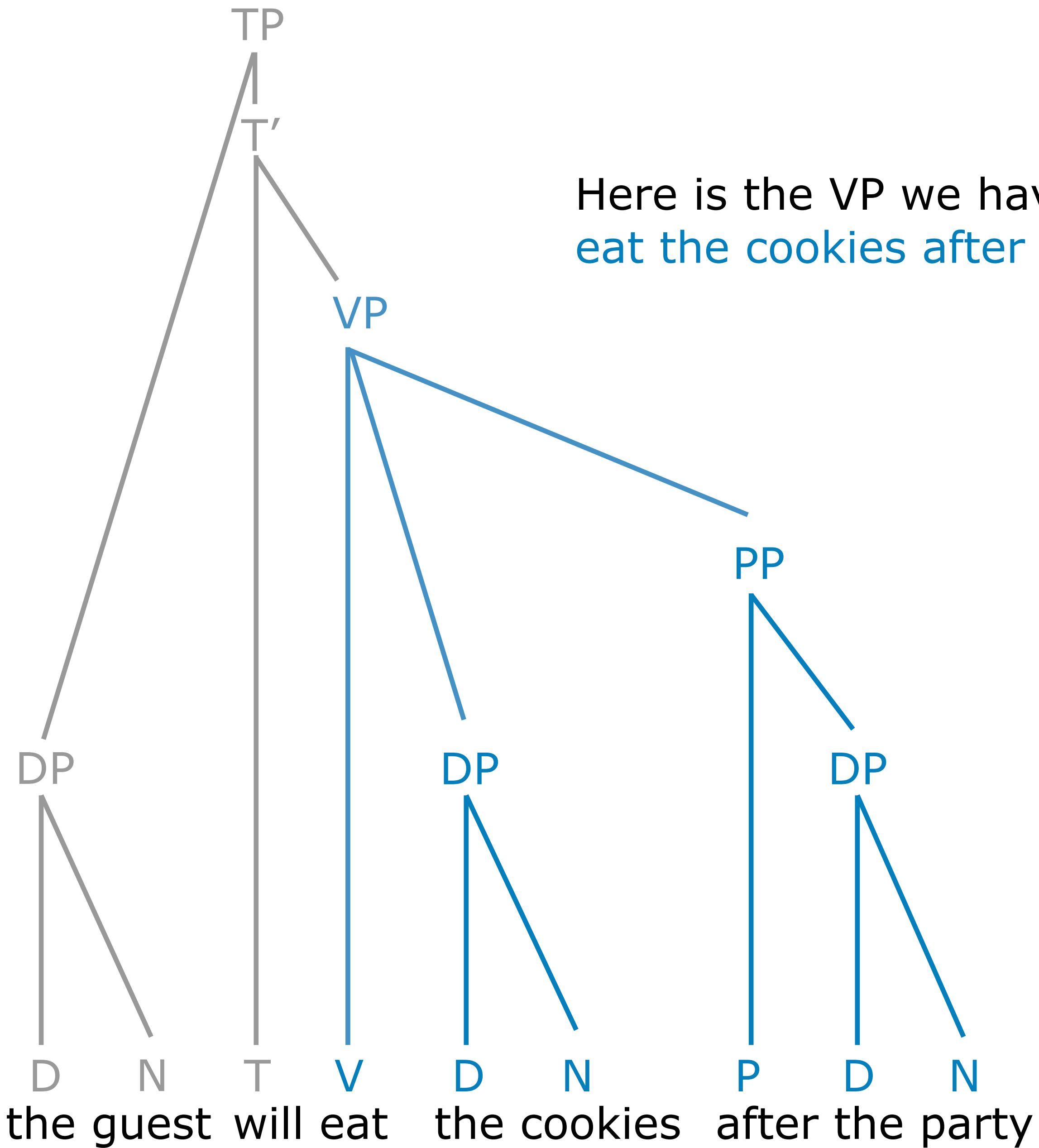


The reason we label the phrases is to indicate their syntactic properties — that is, where they show up in sentences.

For example, when we label a phrase VP, we are saying that it has certain syntactic properties - namely that it will show up where VPs show up!

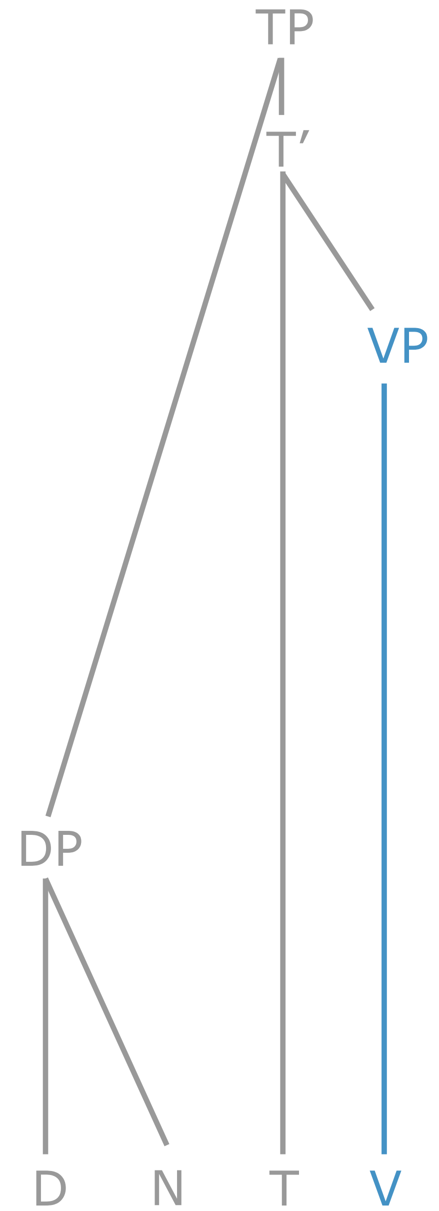
We can see this by looking at several VPs. They will all be in the same syntactic location — after “the guest will”.

# Phrases that have the **same category** can appear in the **same syntactic position**



Here is the VP we have been working with. It is eat the cookies after the party

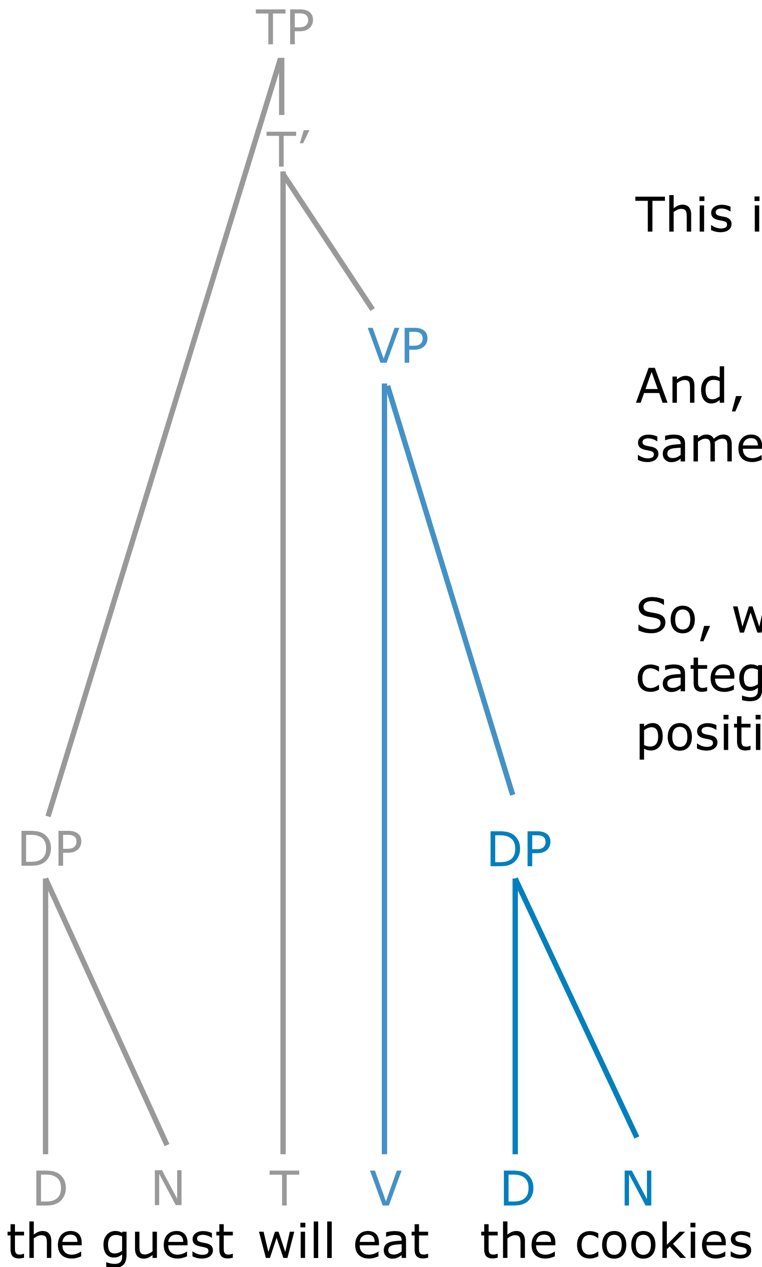
# Phrases that have the **same category** can appear in the **same syntactic position**



This is all a VP. It is just eat.

But, crucially, notice that it appears in exactly the same syntactic position - after “The guest will”.

# Phrases that have the **same category** can appear in the **same syntactic position**



This is all a VP. It is eat the cookies.

And, crucially, notice that it appears in exactly the same syntactic position - after “The guest will”.

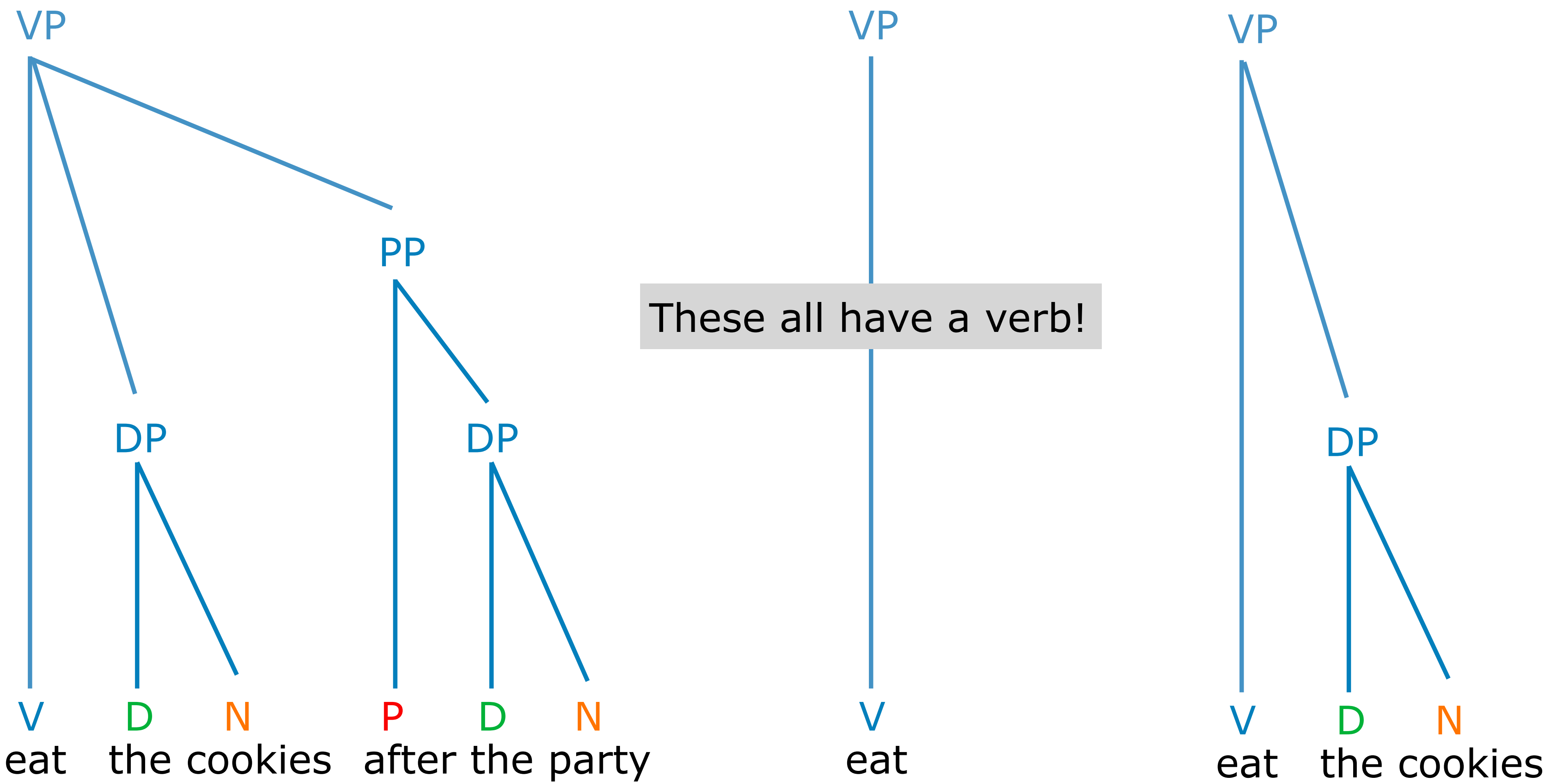
So, we can see that these are all the same category because they all appear in the same position.

But why did we call them VP?

# Because phrases have **heads**

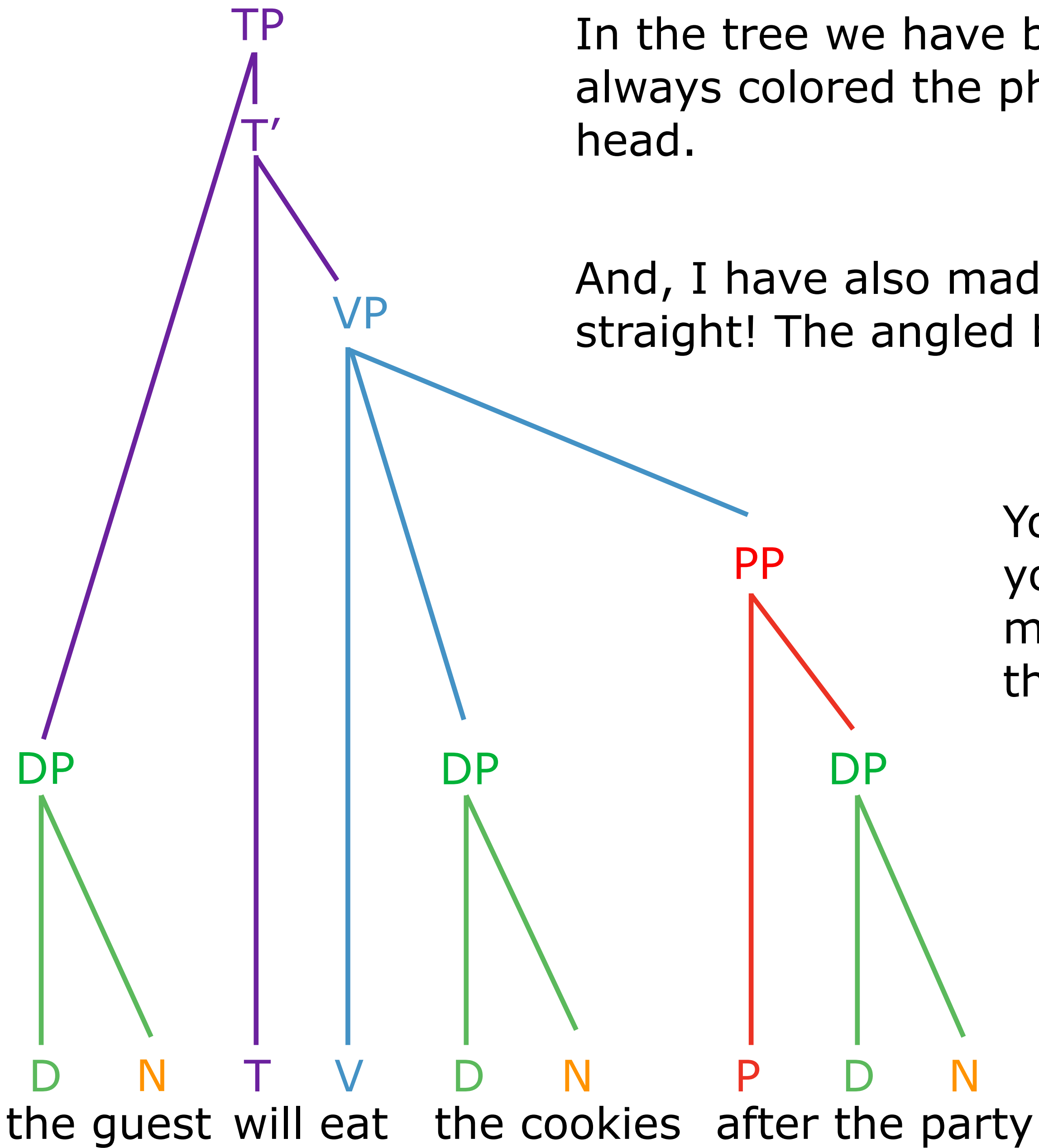
The **head** of a phrase determines its syntactic properties. It is a metaphor — like the way your head controls your properties.

Heads can usually be identified by the fact that they are required for the phrase. If the head is not there, then that phrase would be a different type!



These all have a verb!

# I have already been showing you the heads of the phrases!



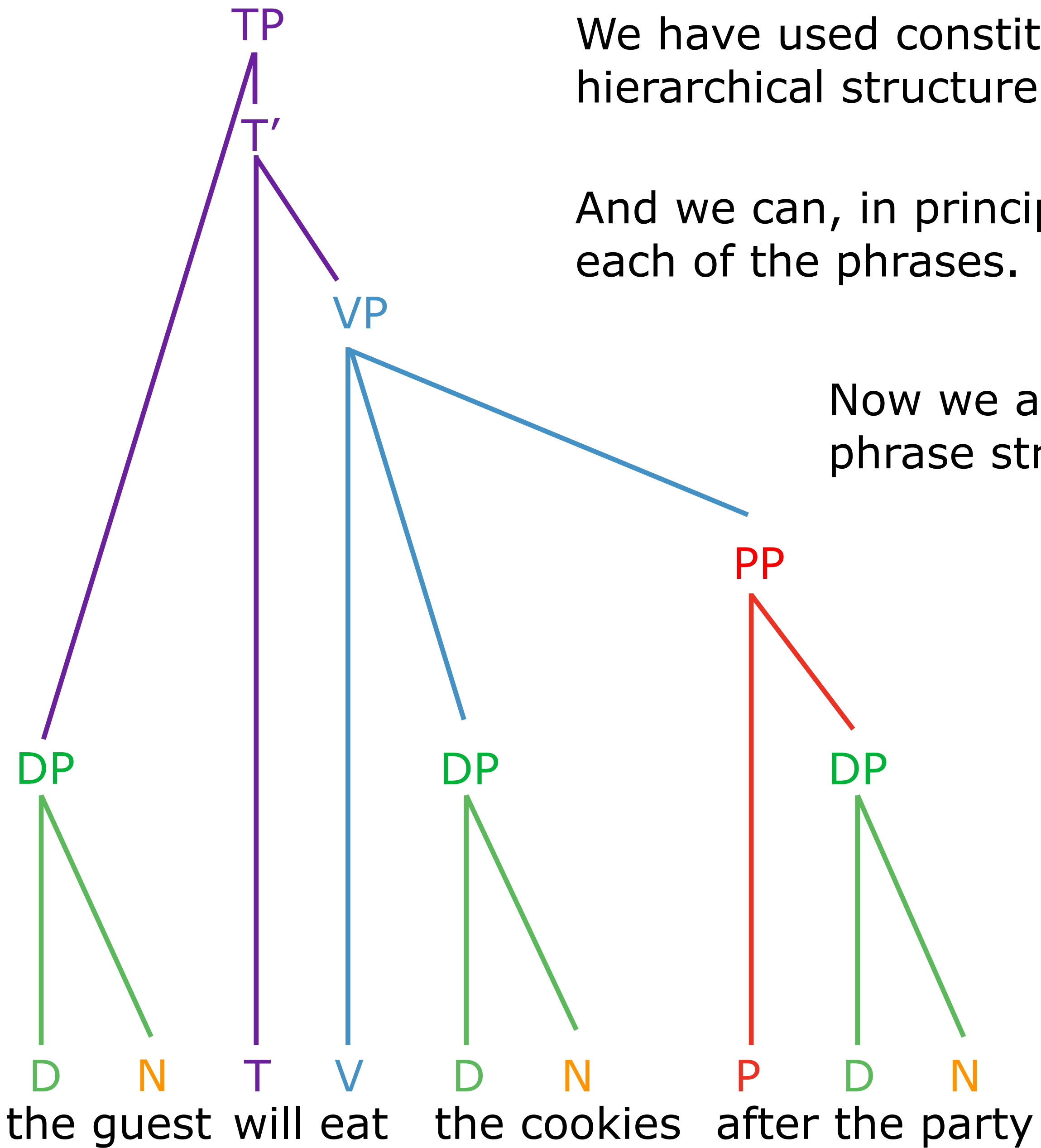
In the tree we have been constructing, I have always colored the phrase the same color as its head.

And, I have also made the branch from the head straight! The angled branches are not heads!

You don’t have to do that in your trees. This is just to make things easier for us in the lectures!

# Phrase Structure Rules

# Final step: Phrase structure rules



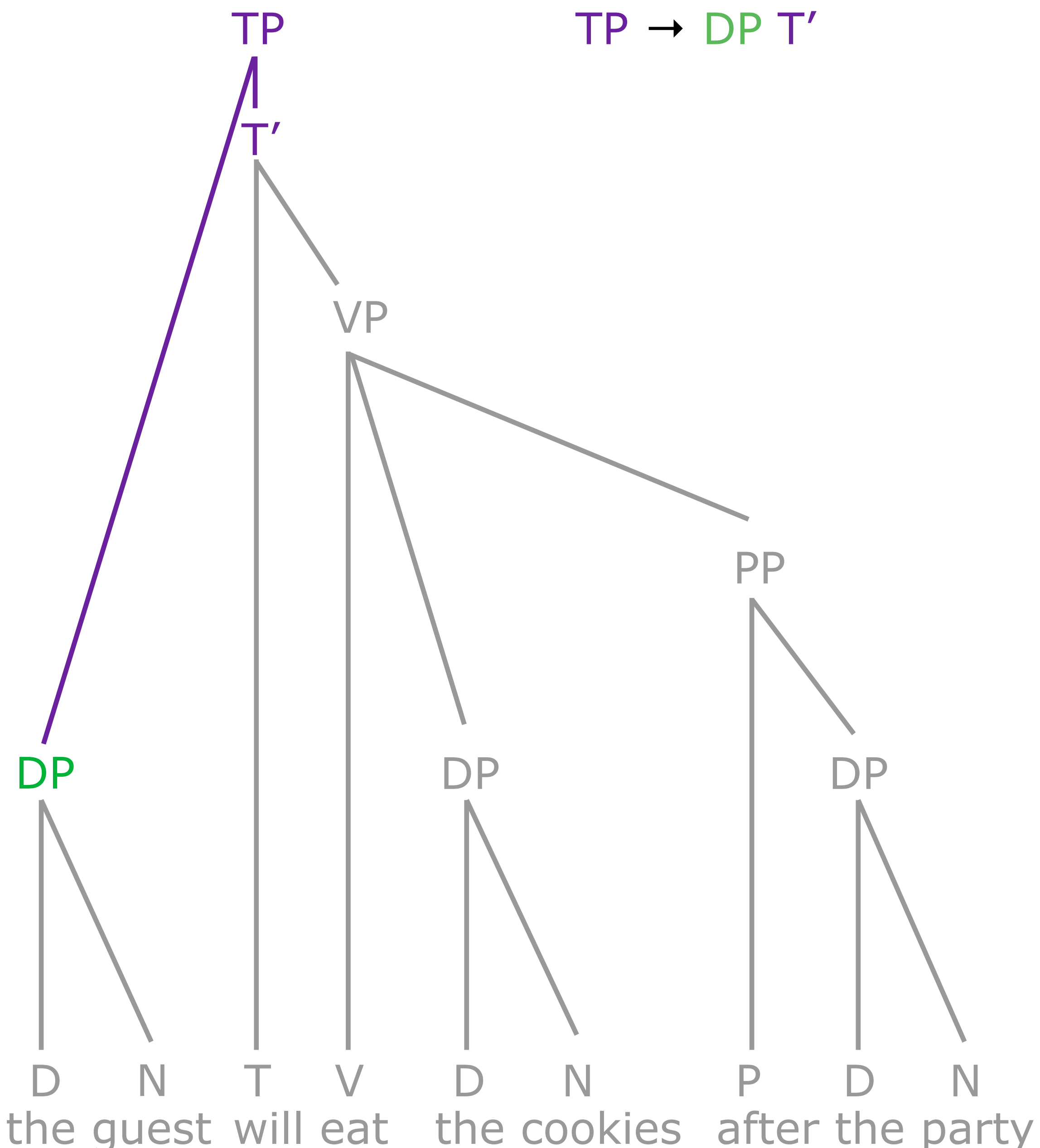
We have used constituency tests to figure out the hierarchical structure of the sentence.

And we can, in principle, identify the heads of each of the phrases.

Now we are ready to read off the phrase structure rules from this tree:

TP → DP T’<br>
T’ → T VP<br>
VP → V DP PP<br>
PP → P DP<br>
DP → D NP

# Final step: Phrase structure rules

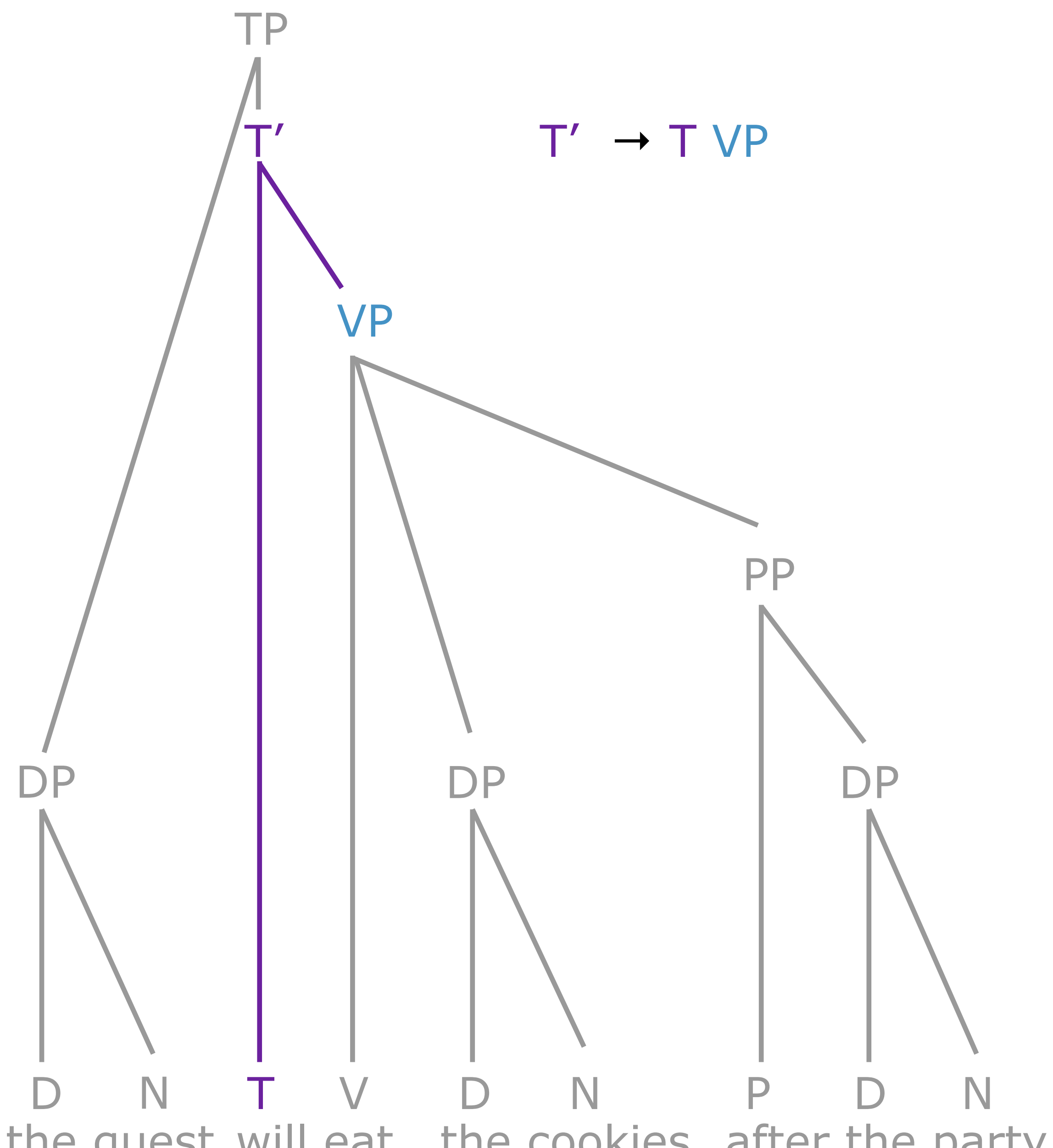


TP → DP T′

**The rules for the complete tree**

TP → DP T′

# Final step: Phrase structure rules



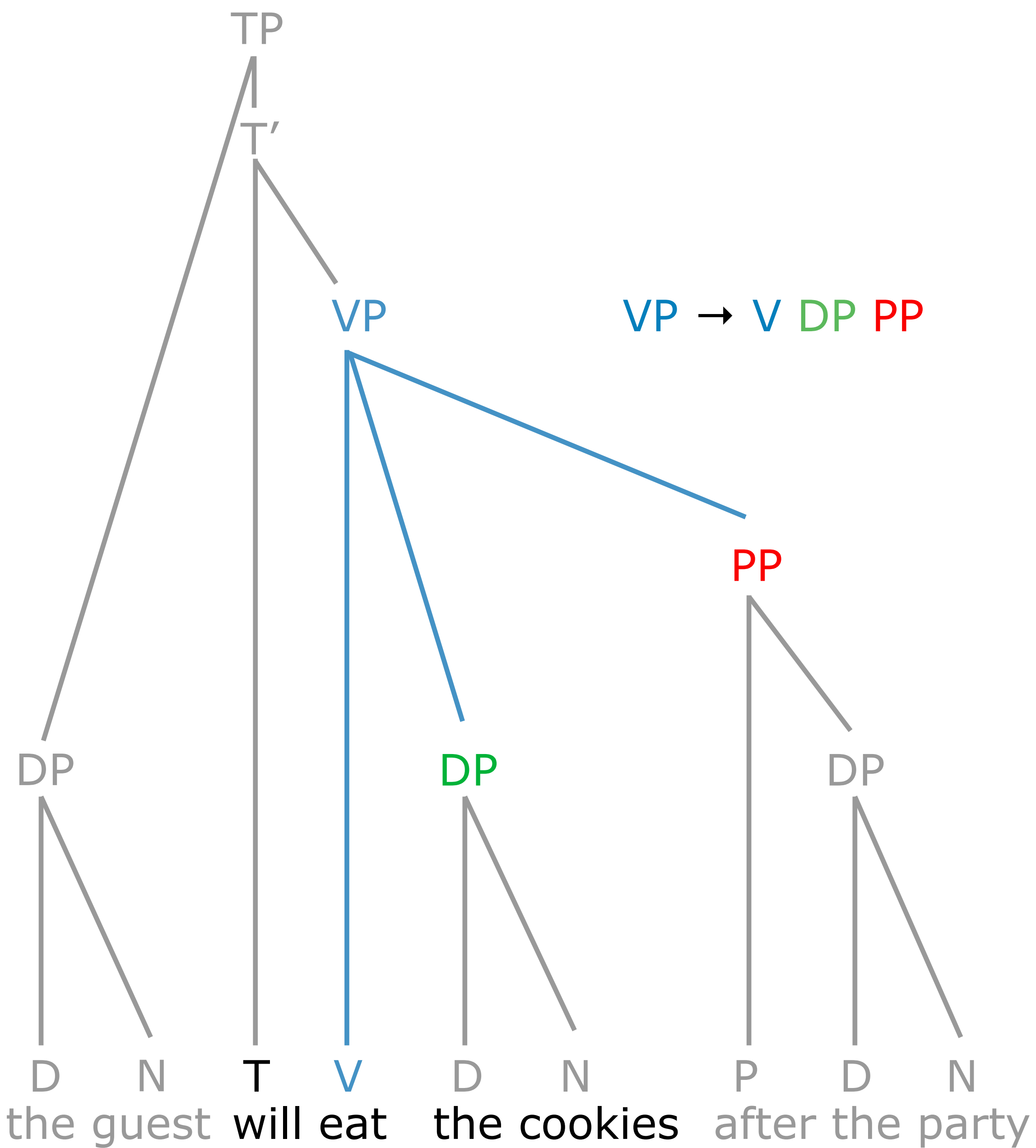
T′ → T VP

**The rules for the complete tree**

TP → DP T′

T′ → T VP

# Final step: Phrase structure rules



VP → V DP PP

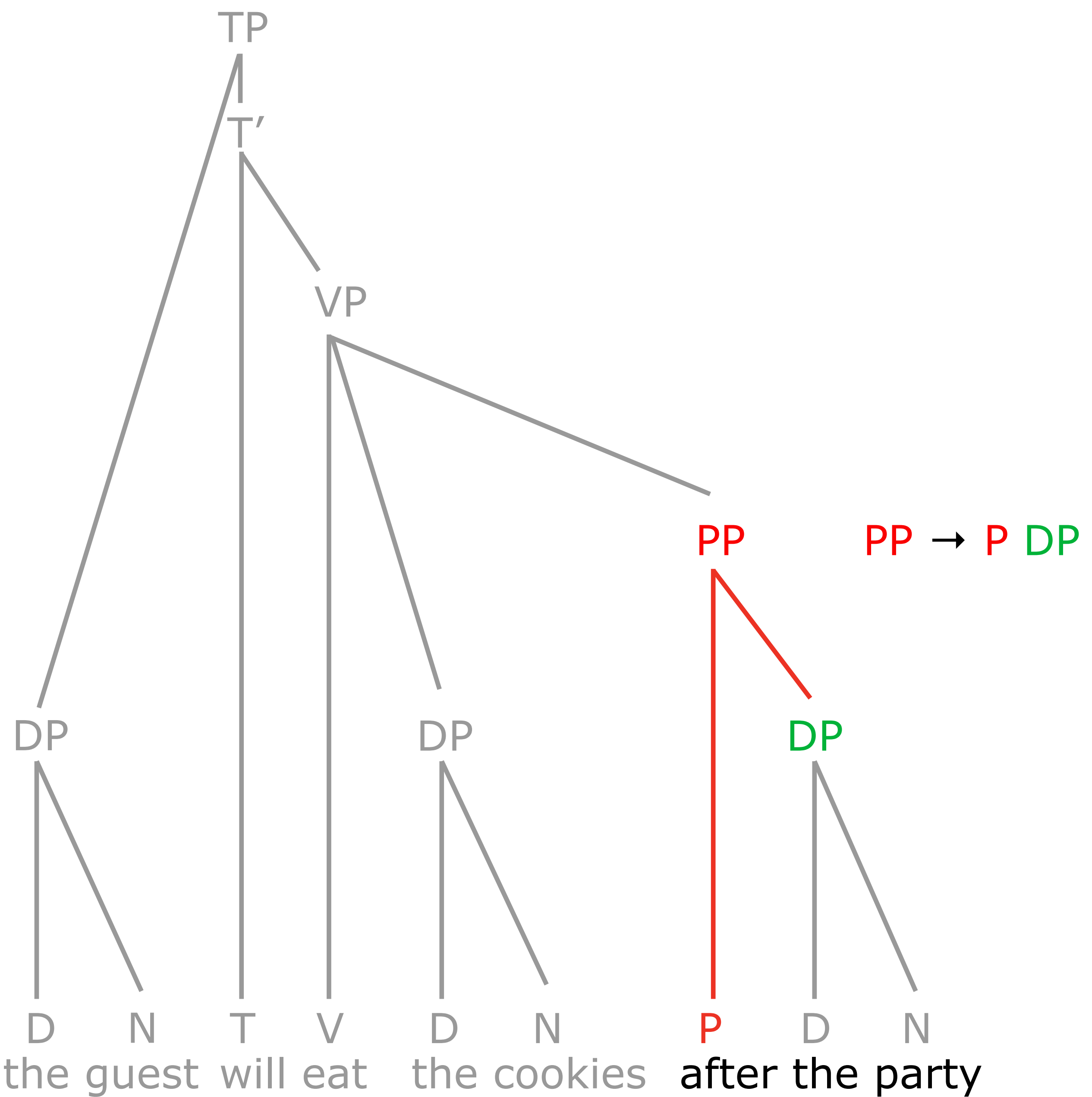
**The rules for the complete tree**

TP → DP T′

T′ → T VP

VP → V DP PP

# Final step: Phrase structure rules



**The rules for the complete tree**

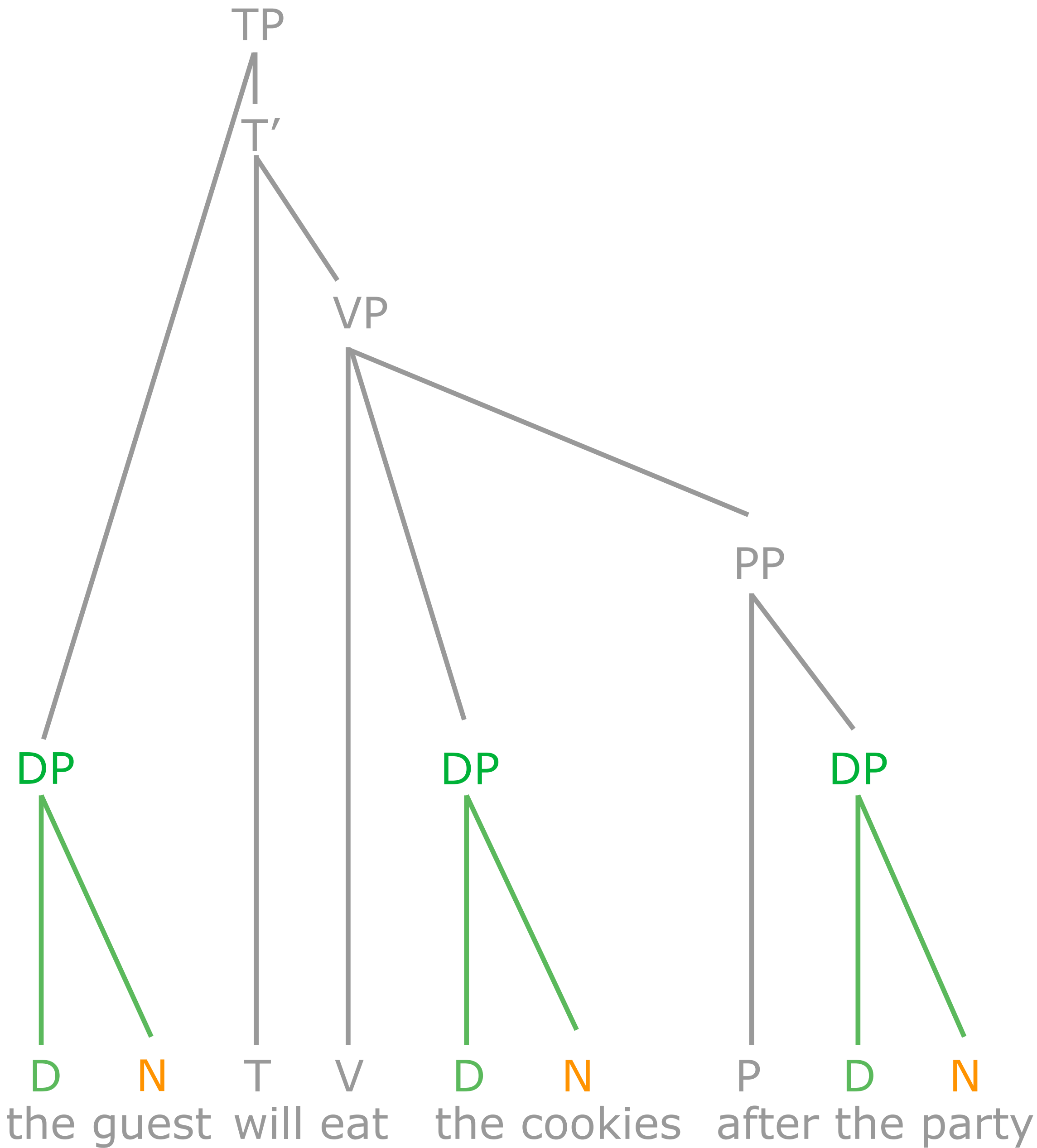
TP → DP T′

T′ → T VP

VP → V DP PP

PP → P DP

# Final step: Phrase structure rules



DP → D NP

**The rules for the complete tree**

TP → DP T′

T′ → T VP

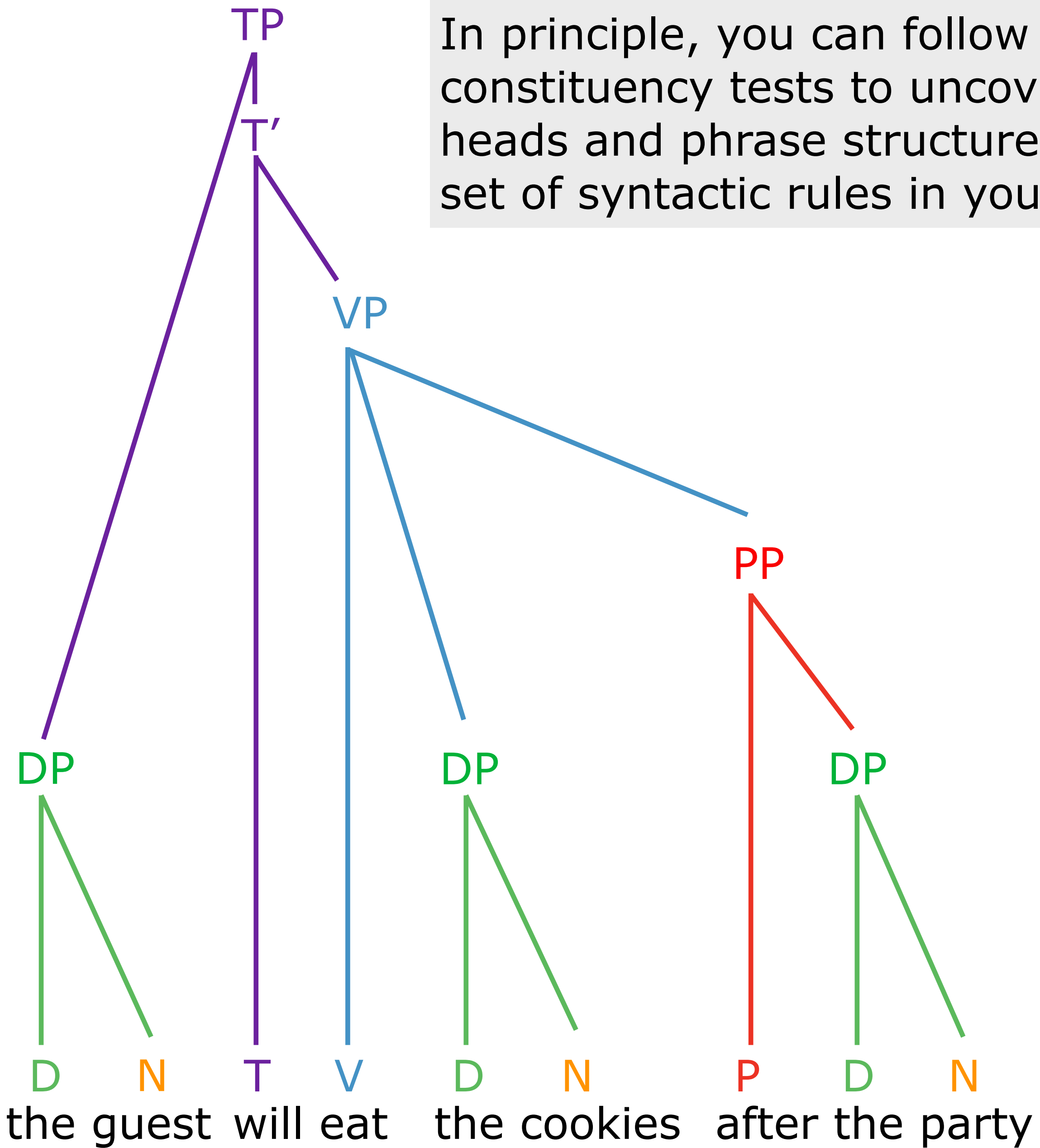
VP → V DP PP

PP → P DP

DP → D NP

# Putting it all together

In principle, you can follow this procedure - using constituency tests to uncover structure, postulating heads and phrase structure rules - to uncover the full set of syntactic rules in your language!



TP → DP T’

T’ → T VP

VP → V DP PP

PP → P DP

DP → D NP

# For next time:
# Building a theory of the rules!