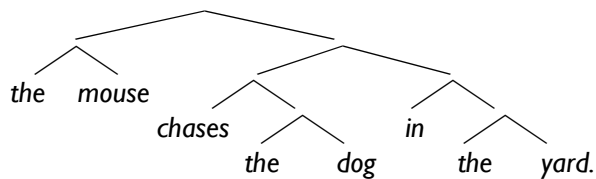*Constituency tests*

Constituents of a sentence are chunks of that sentence that constitute a linguistically relevant part of that sentence.

Limiting cases: Every sentence is a constituent, every word is a constituent.

What we want to find out: What's in between?

The tree structure below represents the constituency structure of the English sentence "The mouse chases the dog in the yard."

In a constituency structure tree, every node corresponds to a constituent of the sentence, and every constituent of the sentence is represented by a node. Therefore this tree tells us that

-   every word is a constituent (terminal nodes)
-   the whole sentence is one (root node)

but also what larger constituents there are:

-   *the mouse*
-   *the dog*
-   *chases the dog*
-   *the yard*
-   *in the yard*
-   *chases the dog in the yard*

It also tells us that for instance "*mouse chases*" is not a constituent, and neither is *"in the".*

How do we find out what's a constituent?

Linguists argue for the constituent structure of a sentence based on the behavior that certain groups of words have. For this class, we'll use two characteristics that have been taken as tests for constituency.

Substitution test

If a chunk of words can be substituted with one of the following pro-forms, and the result is a grammatical sentence with the same meaning as the original one, then the chunk is a constituent.

pronouns   he, she, it, they, …
do so / do it / do that
there, then
such
one (with a contrastive meaning, as for instance in "this boy vs. that one")

Example: The mouse chased the dog *there.*
where *there* replaces *in the yard.*
→ *in the yard* is a constituent of the sentence *The mouse chased the dog in the yard*

Answer to questions

If a chunk of words can stand alone as an answer to a question, it's a constituent of the corresponding full sentence answer.

Example: What did the mouse do? Chase the dog in the yard.
                   The mouse chased the dog in the yard.
→ *chase the dog in the yard* is a constituent of the sentence *The mouse chased the dog in the yard.*

Constituent names

We also talked about having names for constituents. Those names are related to the *heads* of the constituents. A constituent headed by a *noun* is called a *noun phrase* (or *NP*), etc. In the substitution test, the different pro-forms substitute for different constituents:

pro-form	can stand in for a
he, she, it, they, …	NP (noun phrase)
do so / do it / do that	VP (verb phrase)
there, then	PP (prepositional phrase) or AdvP (adverb phrase)
such	AP (adjective phrase)
one	N' (read as "N-bar"; it's a constituent that contains a noun, its arguments and modifiers, but not the determiner)