



Assignment 09: Trees and X-bar theory: Theoretical background

handed out: Monday 22nd July, 14:00
to be submitted by: Monday 29th July, 14:00

X-bar theory imposes constraints on the form of phrase structure trees. One of the core assumptions is that besides the usual syntactic categories like noun, adjective, complementizer, ..., each node in the syntax tree (where, broadly speaking, each such node makes up a so-called constituent) has a so-called projection level, which, roughly speaking, represents how grammatically "complete" a part of a sentence is at that point. X-bar theory then specifies rules on how these constituents can combine to make a well-formed tree. For example, each phrase must have a uniquely determined head, adjuncts are inserted at the level of bar projections, and some heads combine with complements to form the next projection. The exact terminology and rules are given below.

In an X-bar tree, the label of each non-terminal node encodes two kinds of information:

- the syntactic category (e.g. N for noun, P for preposition, A for adjective, ...)
- the projection level – most variants of X-bar theory assume the following three projection levels:
 - level 2: phrase (written as XP or X'' or $\bar{\bar{X}}$ or X²)
 - level 1: bar (written as X' or \bar{X} or X¹)
 - level 0: head (written as X or X⁰)

E.g., the label "AdvP" encodes that the node is of the syntactic category "adverb" and projection level 2, i.e. an adverb phrase, and "N" means that the node is of the category "noun" and projection level 0, i.e. a nominal head.

In order to be able to handle these node labels more easily in a program, we specify that the label name is a string where the first characters are the syntactic category and as the last character we write the projection level as a number; e.g. the adverb phrase "AdvP" is encoded as "Adv2" and the noun head "N" is encoded as "N0".

Some notes on notation and terminology:

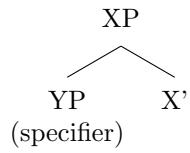
When speaking about the abstract structure of X-bar trees in general without referring to particular instances of these trees, we use variables (X, Y, ...) for the syntactic category, where a variable could stand for any category (N, A, D, ...) but same variable names means that the nodes must be of the same category; different variable names allow the categories to be different or the same.

A node is said to branch **binary** iff it has two children, and **unary** iff it has one child. A node is called a **non-terminal** (or **branching**) node iff it has children, and a **terminal** node (or **leaf**) iff it has no children. In a linguistic syntax tree, the terminal nodes are words, and the non-terminal nodes are intermediate projections encoding a combination of syntactic category and projection level as described above. The top-most node is called the **root** of the tree.

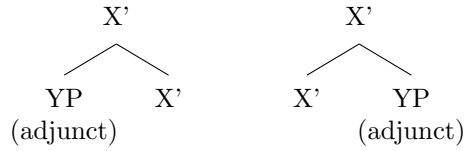
We sometimes use triangles when we don't want to spell out the full structure of a tree (e.g., because we are not interested in the precise structure of every subtree); the triangle is like a "..." stating that there is some more in between which is, however, not shown in more detail.

A constituent YP is called (examples see next page) a

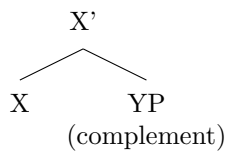
- **specifier** w.r.t. the XP iff it is daughter to XP and sister to X':



- **adjunct** or **modifier** w.r.t. the XP (or X') iff it is daughter to X' and sister to X':

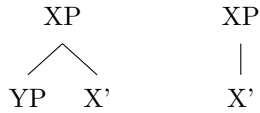


- **complement** w.r.t. the XP (or X) iff it is daughter to X' and sister to X:

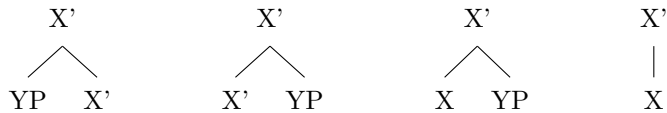


The phrase structure rules of X-bar theory are given as follows¹:

- A phrase of category X (XP) branches
 - binary into a specifier phrase of any category Y (YP) on the left and a bar level of the same category (X') on the right or
 - just the X'.



- An X' (bar projection of category X) branches
 - binary into an adjunct of any category Y (YP) on the left and new bar projection of the same category (X') on the right, or
 - binary into a new X' projection on the left and an adjunct YP on the right, or
 - binary into a head of the same category (X) and a complement phrase of any category Y (YP), or
 - unary into just the head (X).



- A head X branches unary into a terminal word z:



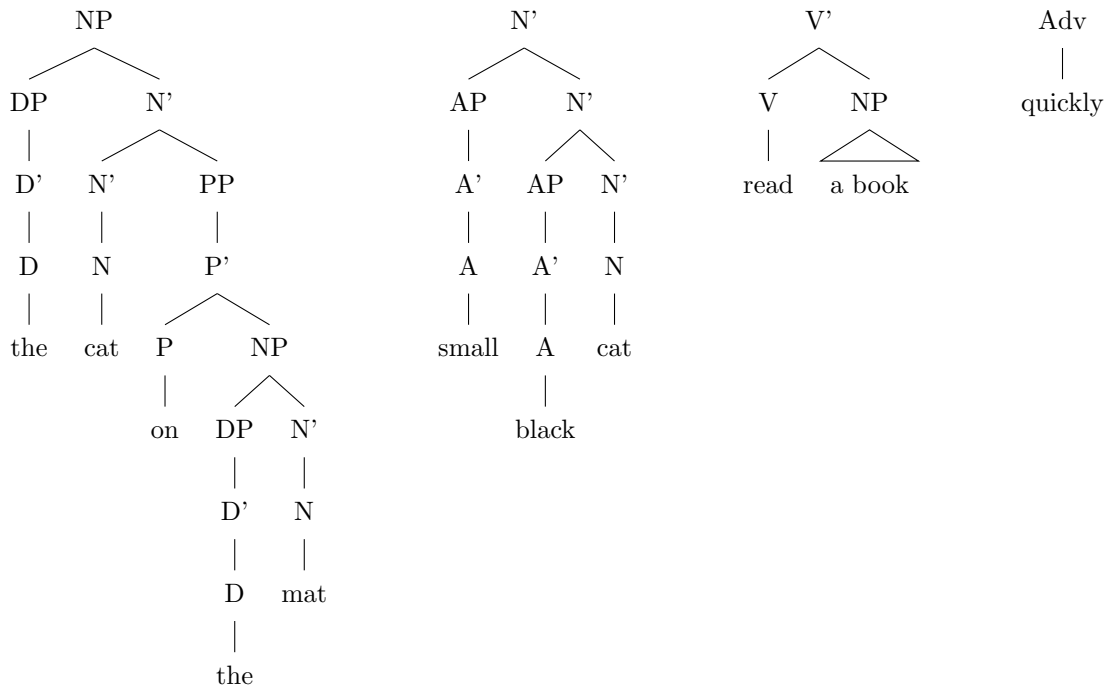
- A terminal node (a word) is just that:

z

¹For reasons of simplicity, we assume that there occur no specifiers on the right of an X' or complements on the left of a head, and that there are no further constraints on the combinations of the categories (for example, what complement phrases verb heads can go with).

Some examples:

- The following trees are well-formed according to the rules of X-bar theory:



More precisely:

The first tree is a phrase of category N with a DP as a specifier (daughter to NP, sister to N'), a PP as an adjunct (daughter to N', sister to N') on the right and no complement (since the N head has no sister). The DP specifier itself has no specifiers, adjuncts or complements; the PP complement has no specifier or adjunct, but an NP complement; the latter has a DP specifier (which itself has no specifiers, adjuncts or complements) and no adjuncts or complements.

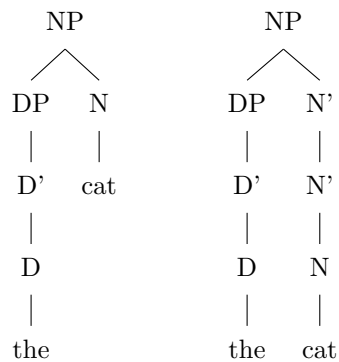
The second tree is a bar of category N with no complement and two AP adjuncts on the left, which themselves both have no specifiers, adjuncts or complements. We can not talk about the specifiers of the N' because we don't have the full NP.

The third tree is a V' with an NP complement (daughter to V', sister to the V head). We were too lazy to spell out what the NP looks like internally so we used a triangle to abbreviate, but of course the NP does have an internal structure, it's just not shown in the picture.

The fourth tree is an adverb head.

The constituents of the tree (e.g. the specifier DP and the adjunct PP in the first tree) are themselves well-formed phrases of X-bar theory. This needs to be checked recursively; if some part of the tree is not well-formed, the tree as a whole isn't either.

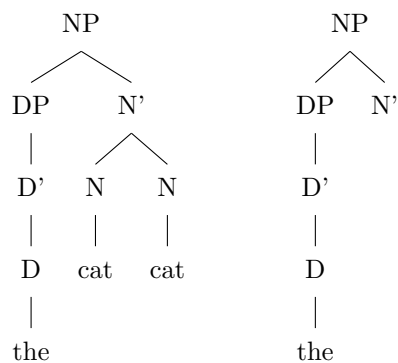
- The following trees are not well-formed because the first one has no bar level (NP branches directly into N) and the second one has two (N' branches into another N' although there is no adjunct or complement present):



- The following tree is not well-formed because inside the DP, the syntactic category of the phrase and the bar don't match (DP branching into a P'):



- The following trees are not well-formed because the first one has two heads (N' branching into two Ns) and the second one has no head at all:



You will find similar and more examples in `test_ex.09.py`.

Total points: 8