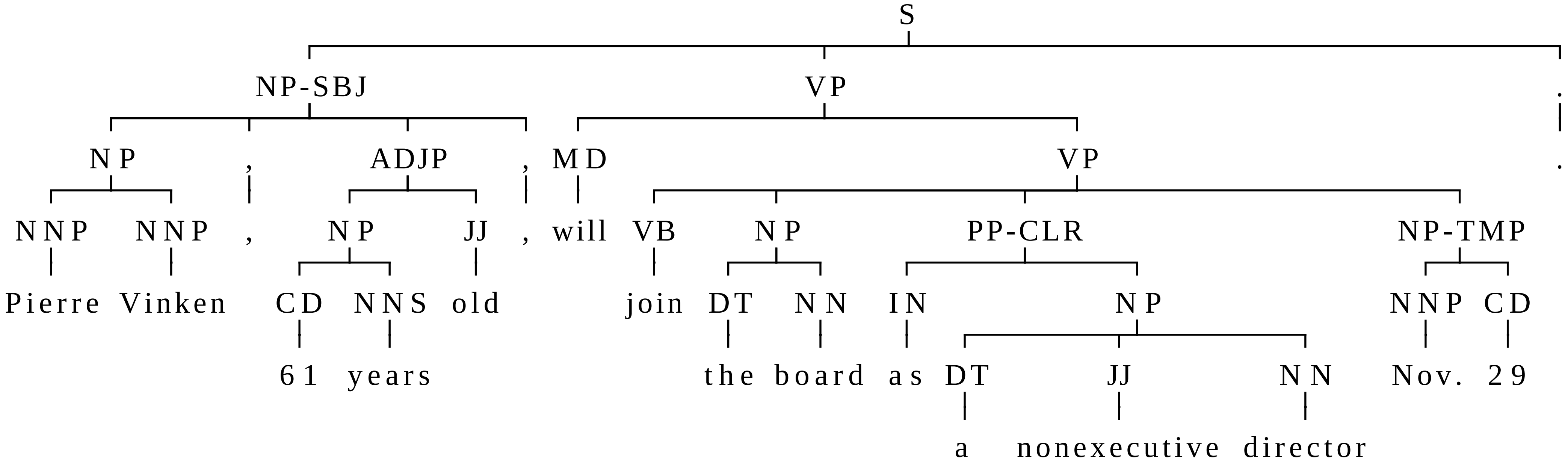
Grammar Formalisms

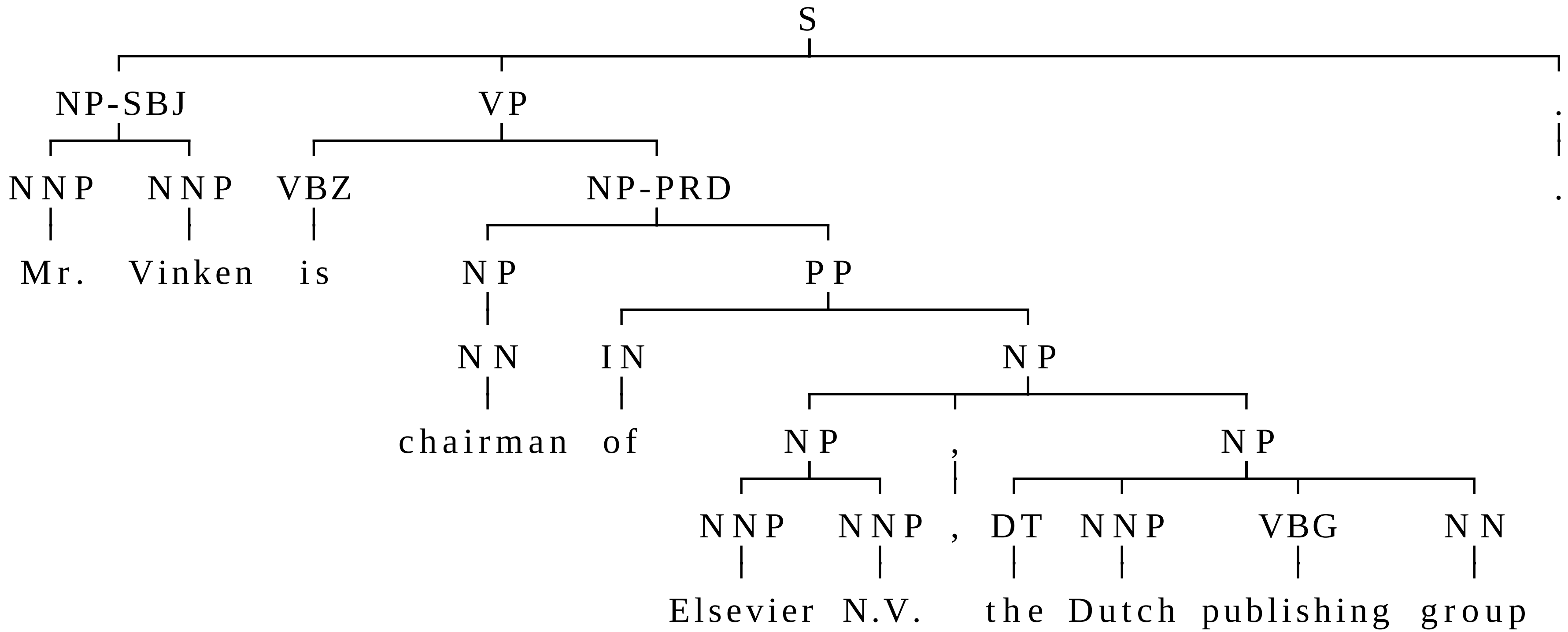
Yoav Goldberg

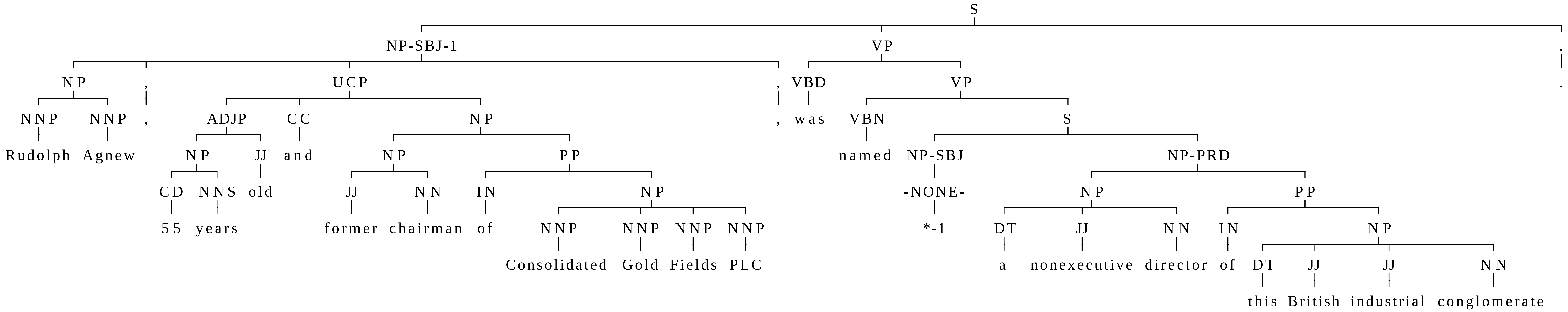
(TAG slides by Julia Hockenmaier)

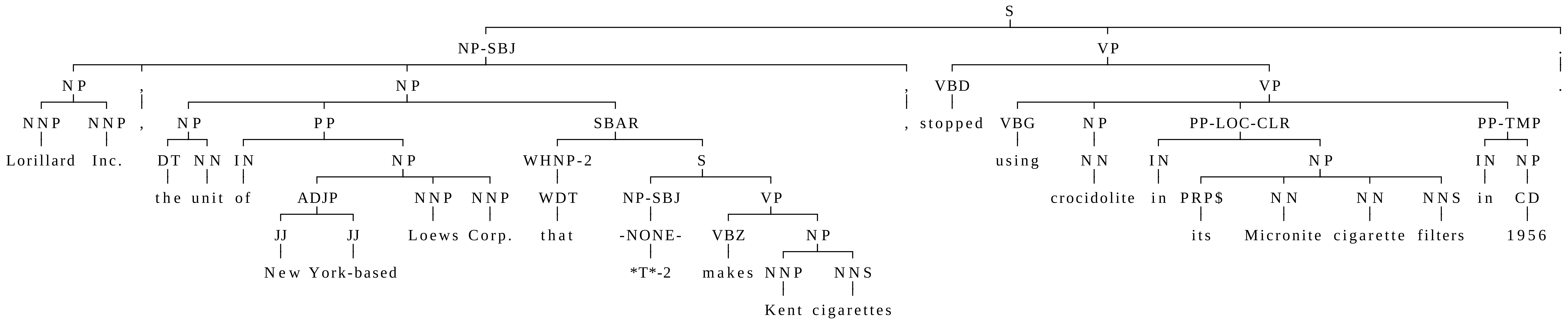
Some Examples of Trees
(from the Penn Treebank Corpus)

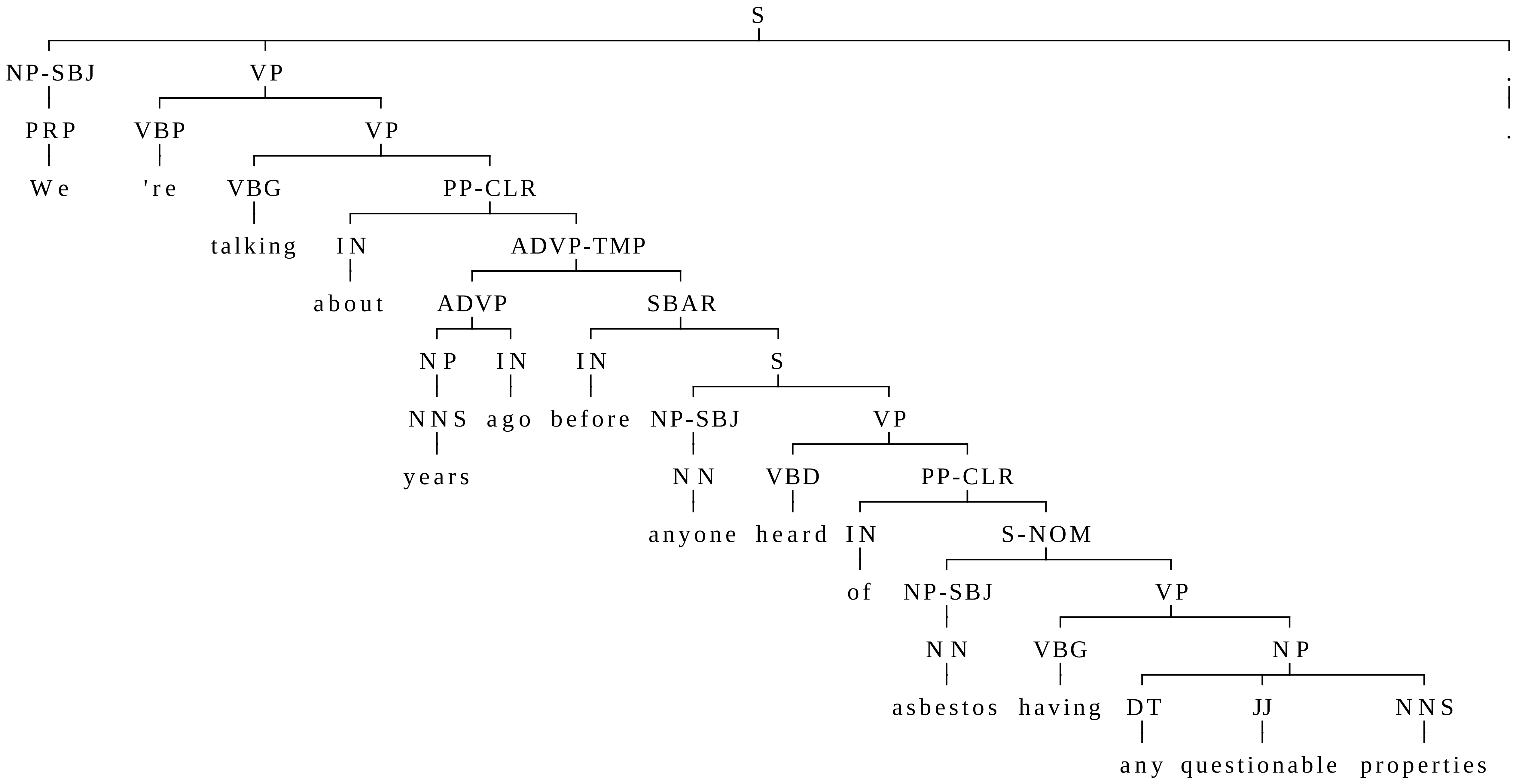


S
NP-SBJ
VP
.
NP
,
ADJP
,
MD
VP
.
NNP
NNP
,
NP
JJ
,
will
VB
NP
PP-CLR
NP-TMP
Pierre
Vinken
CD
NNS
old
join
DT
NN
IN
NP
NNP
CD
61
years
the
board
as
DT
JJ
NN
Nov.
29
a
nonexecutive
director


S
NP-SBJ
VP
.
NNP
NNP
VBZ
NP-PRD
.
Mr.
Vinken
is
NP
PP
NN
IN
NP
chairman
of
NP
,
NP
NNP
NNP
,
DT
NNP
VBG
NN
Elsevier
N.V.
the
Dutch
publishing
group

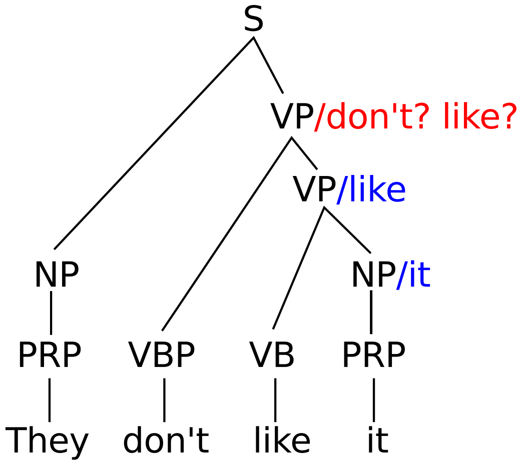

S
NP-SBJ-1
VP
.
NP
,
UCP
,
VBD
VP
.
NNP
NNP
,
ADJP
CC
NP
,
was
VBN
S
Rudolph
Agnew
NP
JJ
and
NP
PP
named
NP-SBJ
NP-PRD
CD
NNS
old
JJ
NN
IN
NP
-NONE-
NP
PP
55
years
former
chairman
of
NNP
NNP
NNP
NNP
*-1
DT
JJ
NN
IN
NP
Consolidated
Gold
Fields
PLC
a
nonexecutive
director
of
DT
JJ
JJ
NN
this
British
industrial
conglomerate

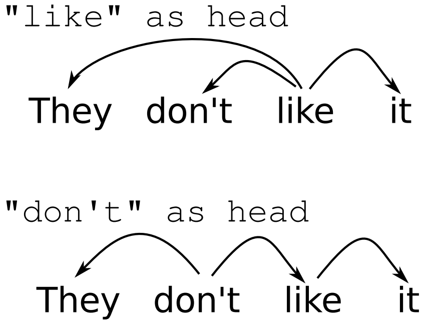

S
NP-SBJ
VP
.
NP
,
NP
,
VBD
VP
NNP
NNP
NP
PP
SBAR
stopped
VBG
NP
PP-LOC-CLR
PP-TMP
Lorillard
Inc.
DT
NN
IN
NP
WHNP-2
S
using
NN
IN
NP
IN
NP
the
unit
of
ADJP
NNP
NNP
WDT
NP-SBJ
VP
crocidolite
in
PRP$
NN
NN
NNS
in
CD
JJ
JJ
Loews
Corp.
that
-NONE-
VBZ
NP
its
Micronite
cigarette
filters
1956
New
York-based
T-2
makes
NNP
NNS
Kent
cigarettes


S
NP-SBJ
VP
.
PRP
VBP
VP
.
We
're
VBG
PP-CLR
talking
IN
ADVP-TMP
about
ADVP
SBAR
NP
IN
IN
S
NNS
ago
before
NP-SBJ
VP
years
NN
VBD
PP-CLR
anyone
heard
IN
S-NOM
of
NP-SBJ
VP
NN
VBG
NP
asbestos
having
DT
JJ
NNS
any
questionable
properties

Constituents → Heads → Dependencies

The choice of heads determines the dependency structure





Tree-Adjoining Grammar

Defining a grammar formalism

- **The conventional way**
  - Define simple elementary objects (e.g. words)
  - Define various operations to combine these objects.
  - Introduce new operations to deal with more complex structures.

- **The TAG way: Complicate locally, simplify globally**
  - Define complex elementary objects (e.g. trees) that capture crucial linguistic properties.
  - Define simple, general operations to combine these objects.
  - *What kind of predictions does this system make?*

- **[A. Joshi, *Starting with complex primitives pays off*, Cognitive Science (2004)]**

(Lexicalized) Tree-Adjoining Grammar

- **TAG is a tree-rewriting formalism:**
  - TAG defines operations (**substitution**, **adjunction**) on trees.
  - The **elementary objects** in TAG are trees (not strings)
- **TAG is lexicalized:**
  - Each elementary tree is **anchored** to a lexical item (word)
  - "**Extended domain of locality**":
    The elementary tree contains all arguments of the anchor.
  - TAG requires a linguistic theory which specifies the shape of these elementary trees.
- **TAG is mildly context-sensitive:**
  - can capture Dutch cross-serial dependencies
  - but is still efficiently parseable

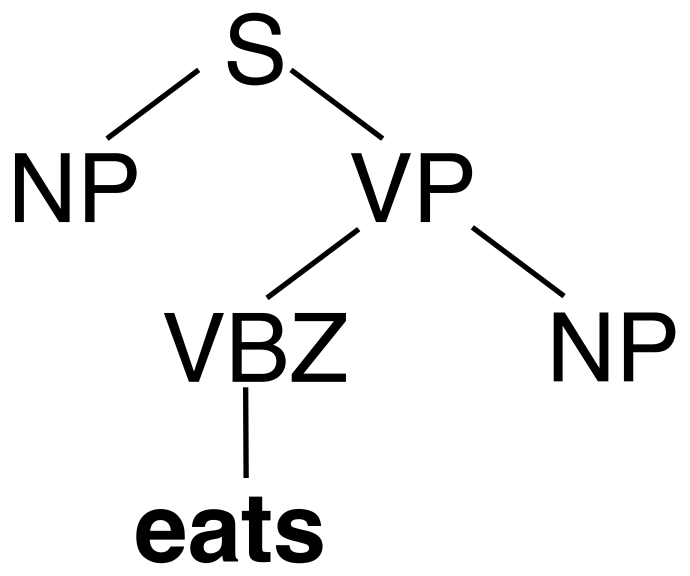
AK Joshi and Y Schabes (1996) Tree Adjoining Grammars.

Domain of locality

- In a CFG, the domain of locality is confined to a single rule.
- Each local tree is independent.

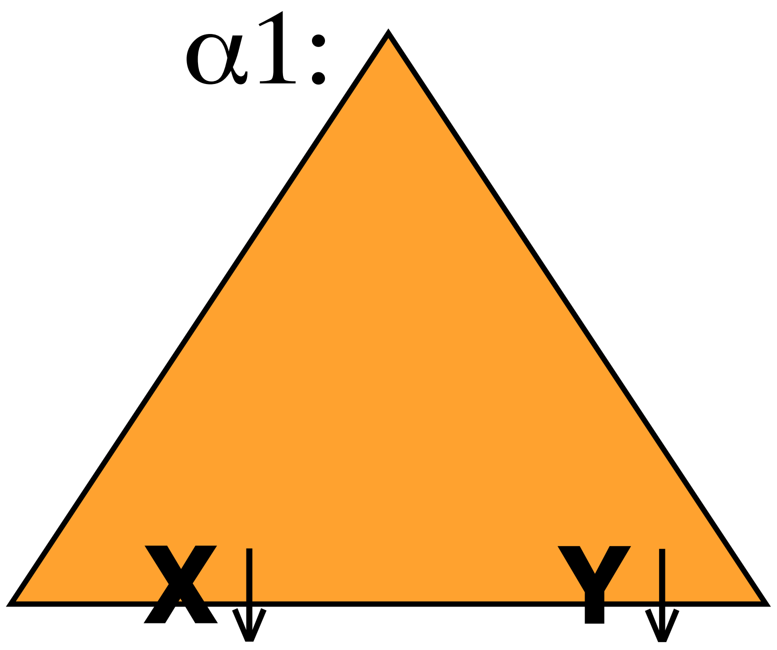
Extended domain of locality

- We want to capture all arguments of a word in a single elementary object.
- We also want to retain certain syntactic structures (e.g. VPs).
- Our elementary objects are tree fragments:

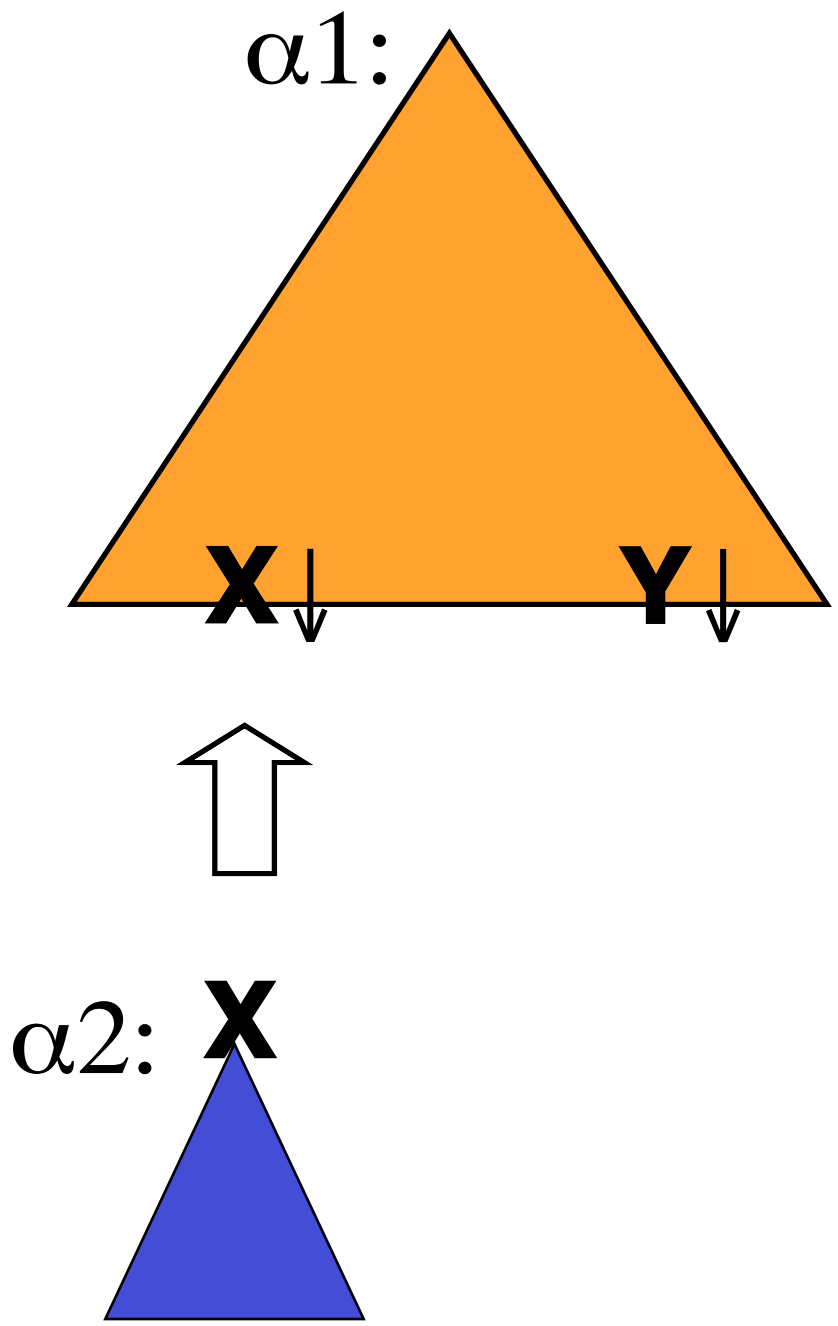


TAG substitution (arguments)

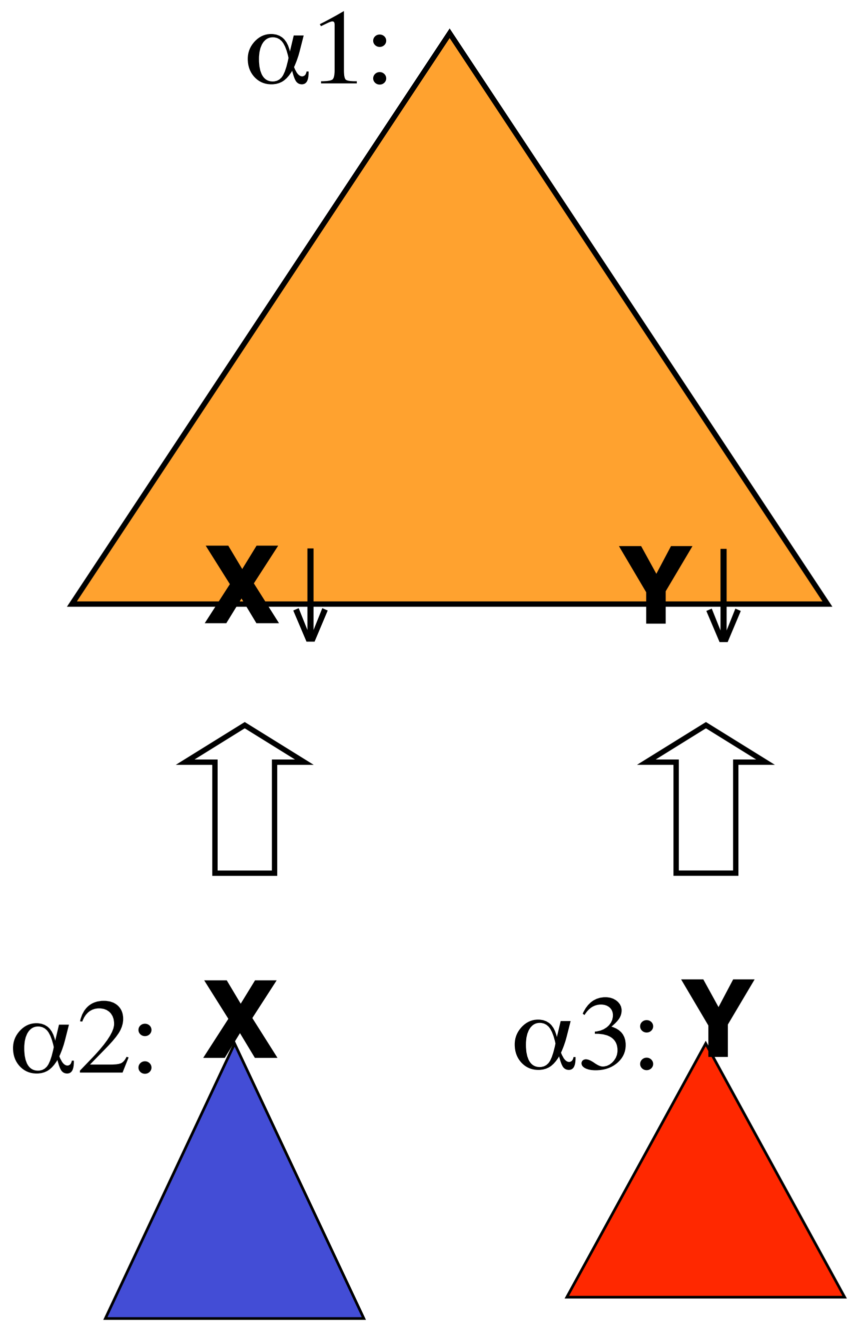
TAG substitution (arguments)



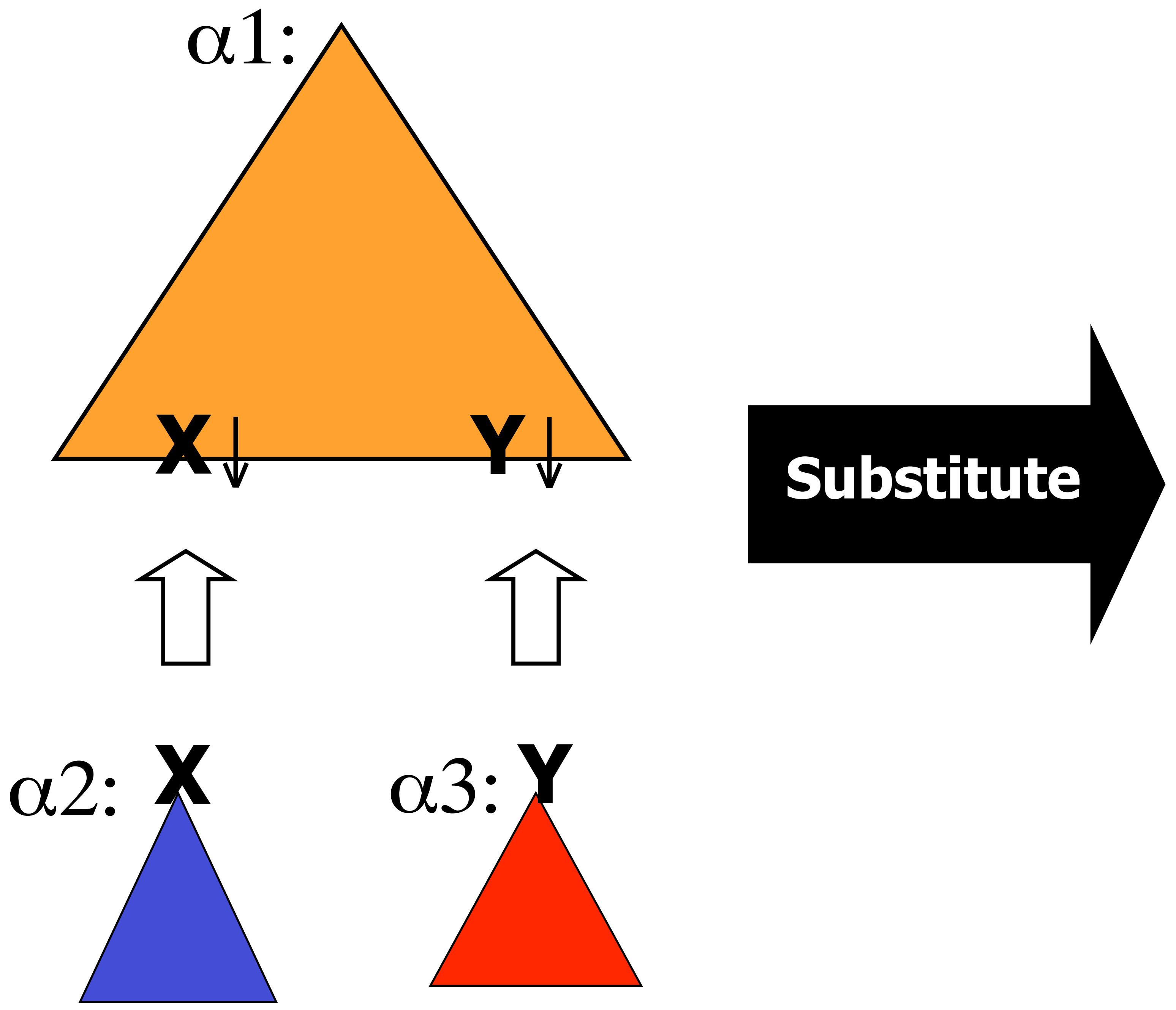
TAG substitution (arguments)



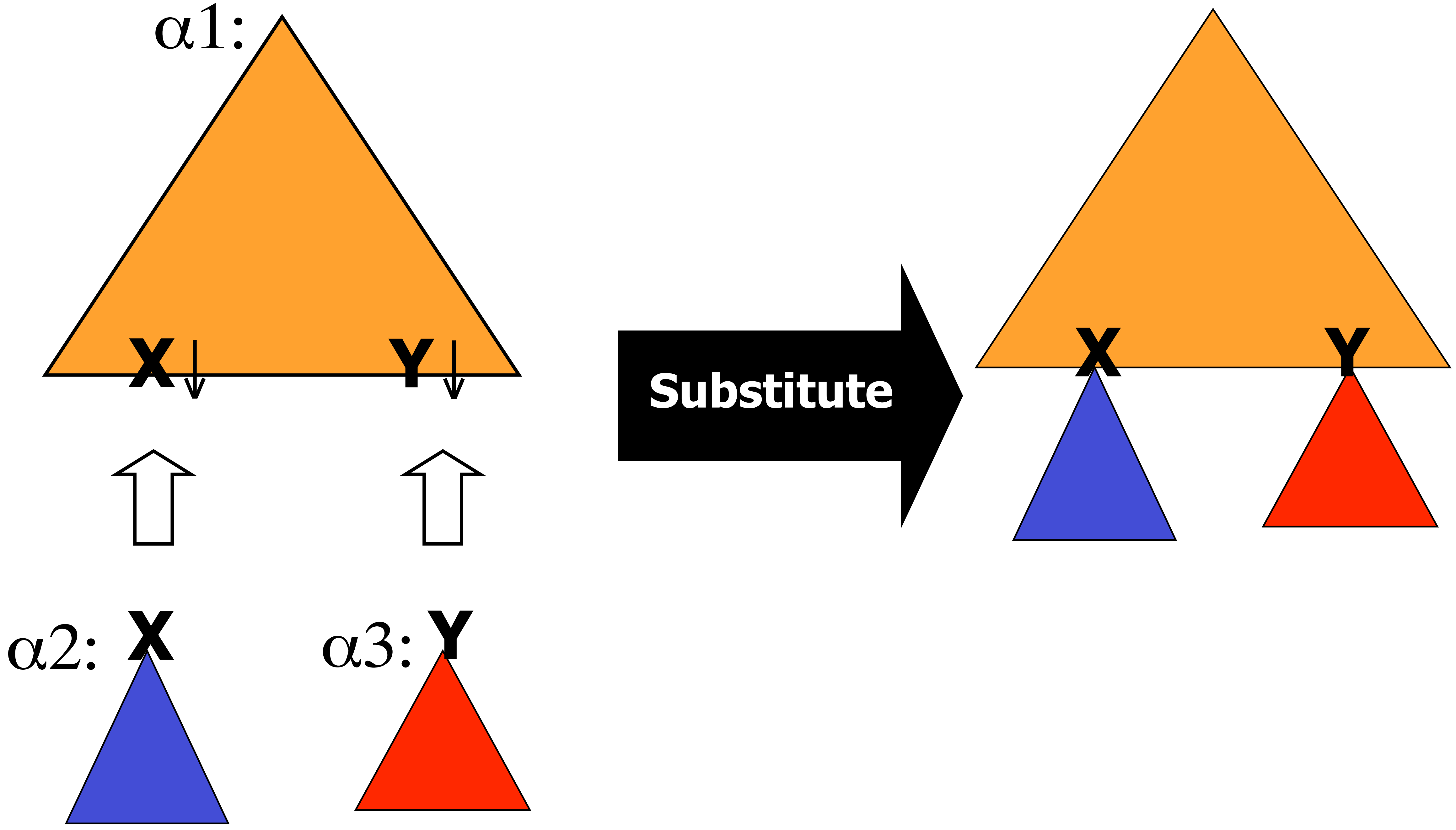
TAG substitution (arguments)



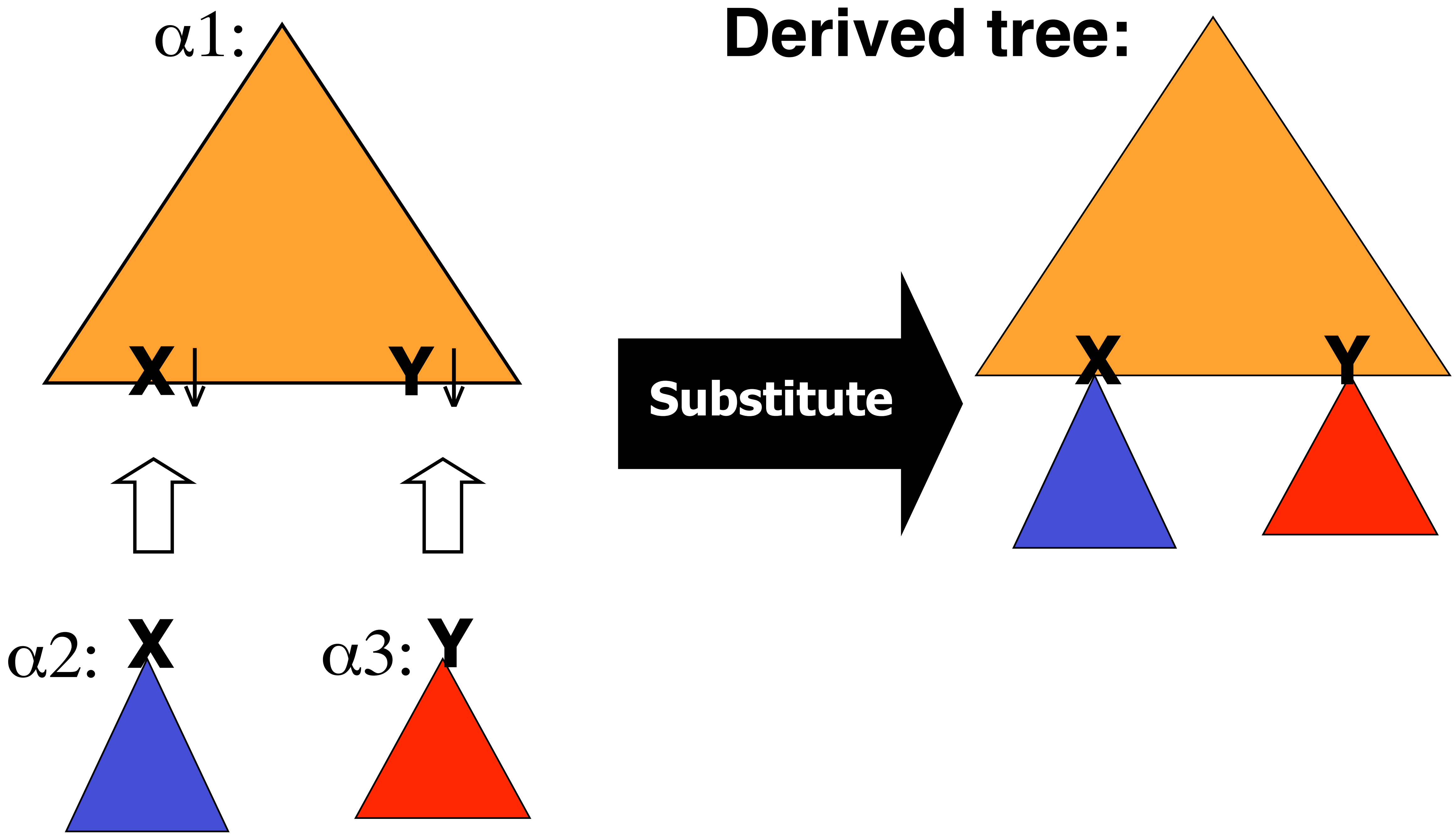
TAG substitution (arguments)



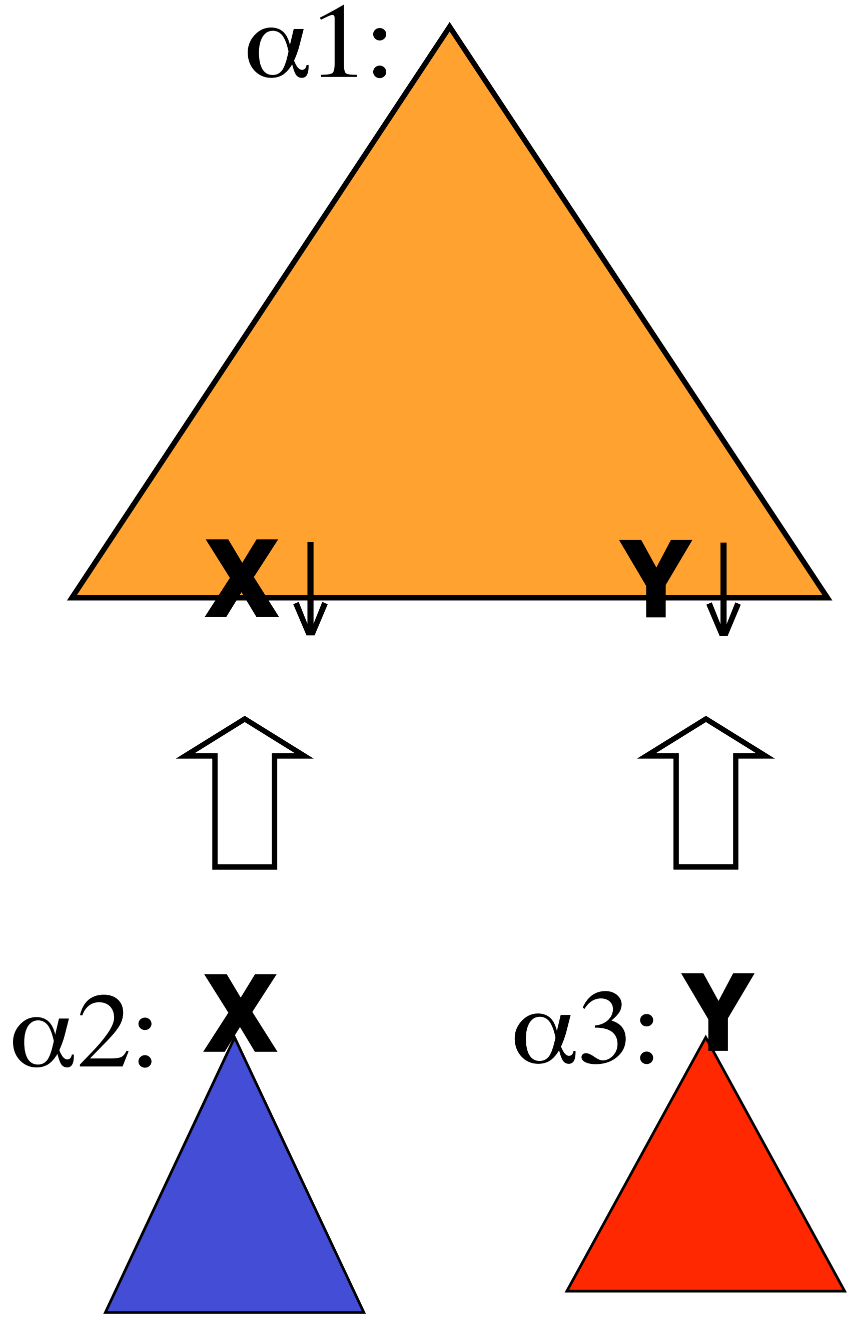
TAG substitution (arguments)



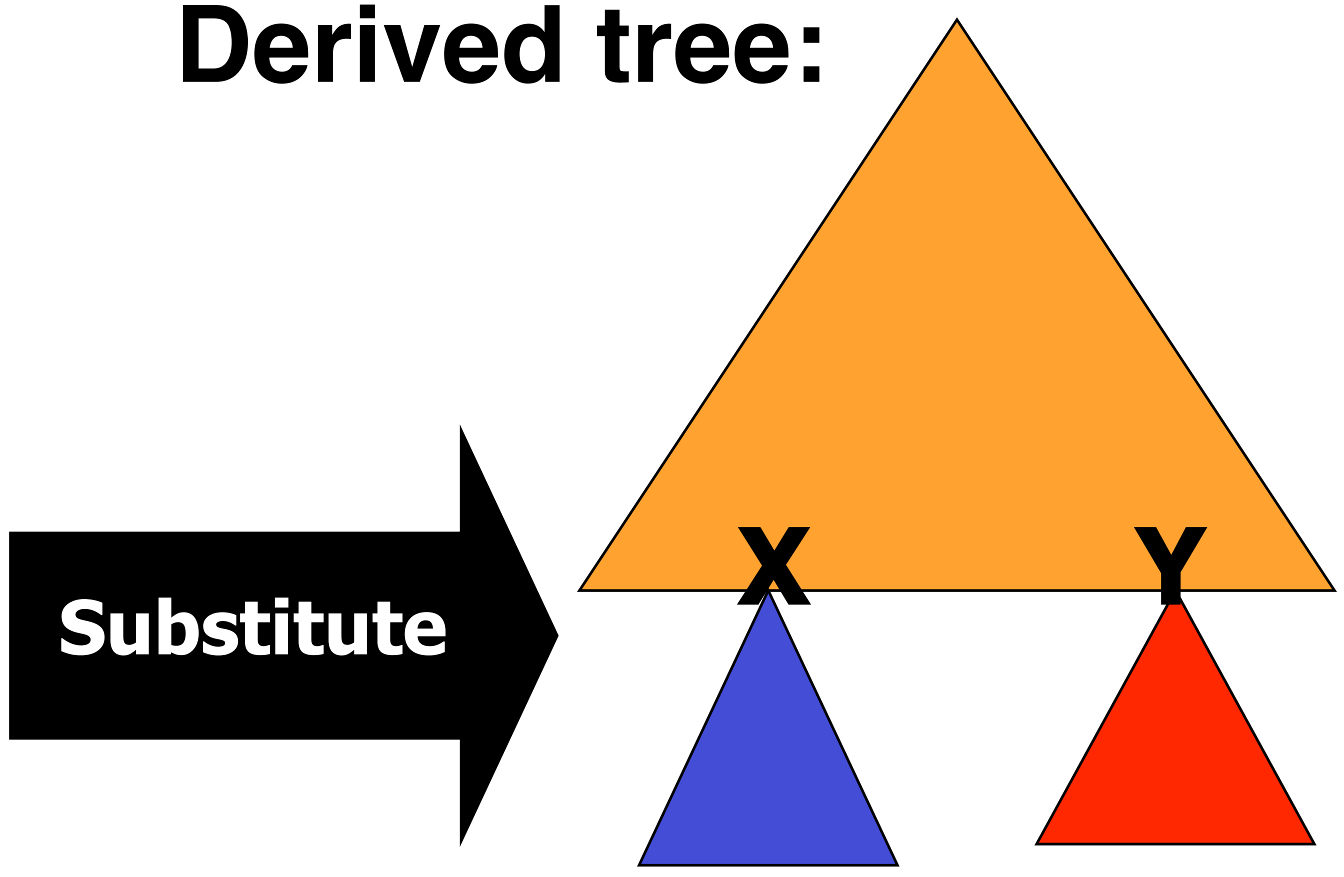
TAG substitution (arguments)



TAG substitution (arguments)

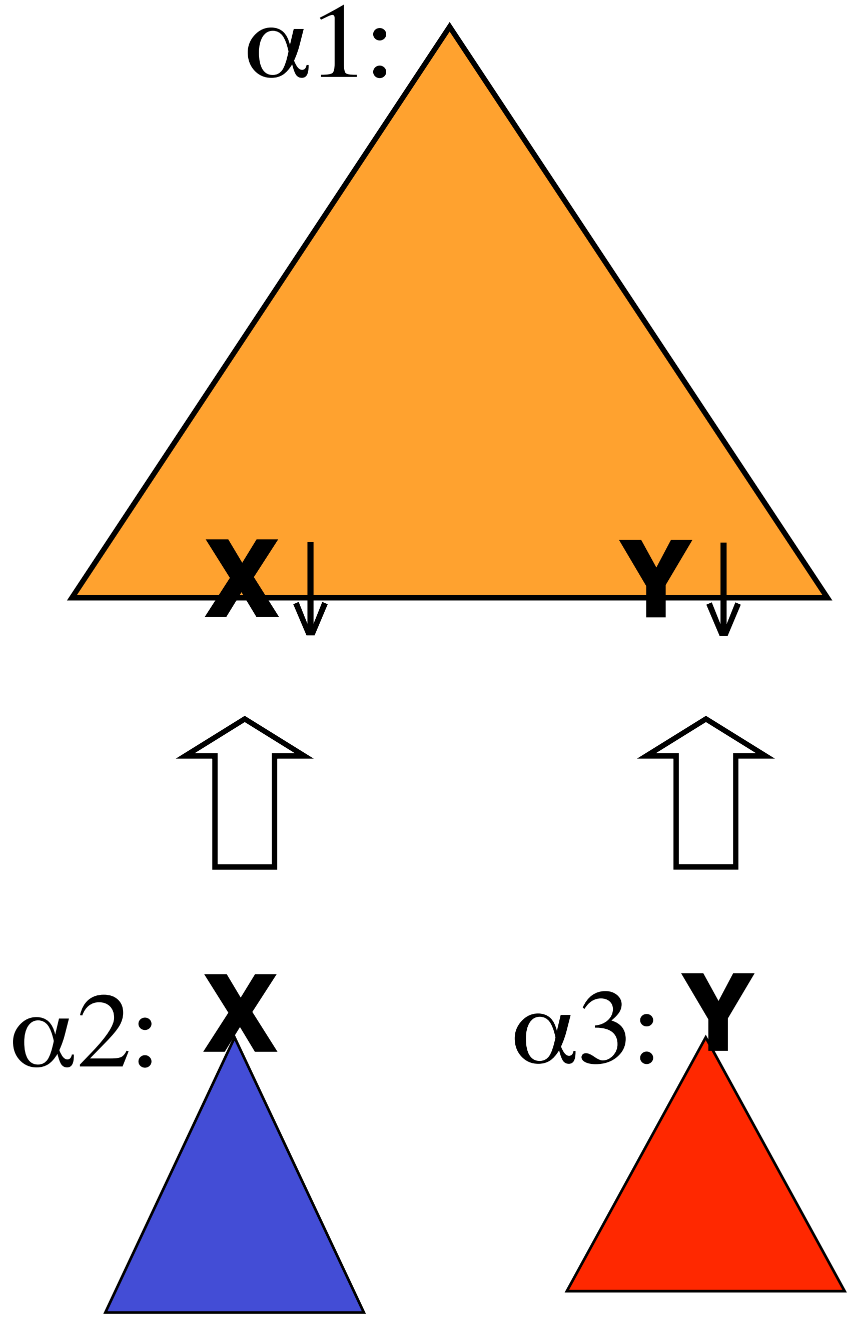


Derived tree:

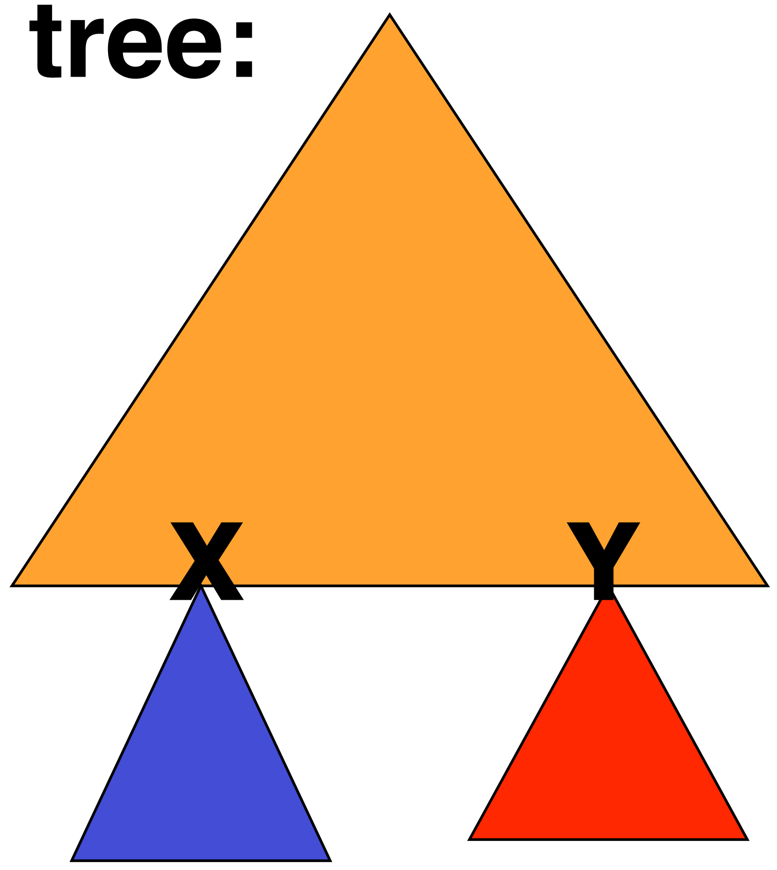


Derivation tree:

TAG substitution (arguments)



Derived tree:



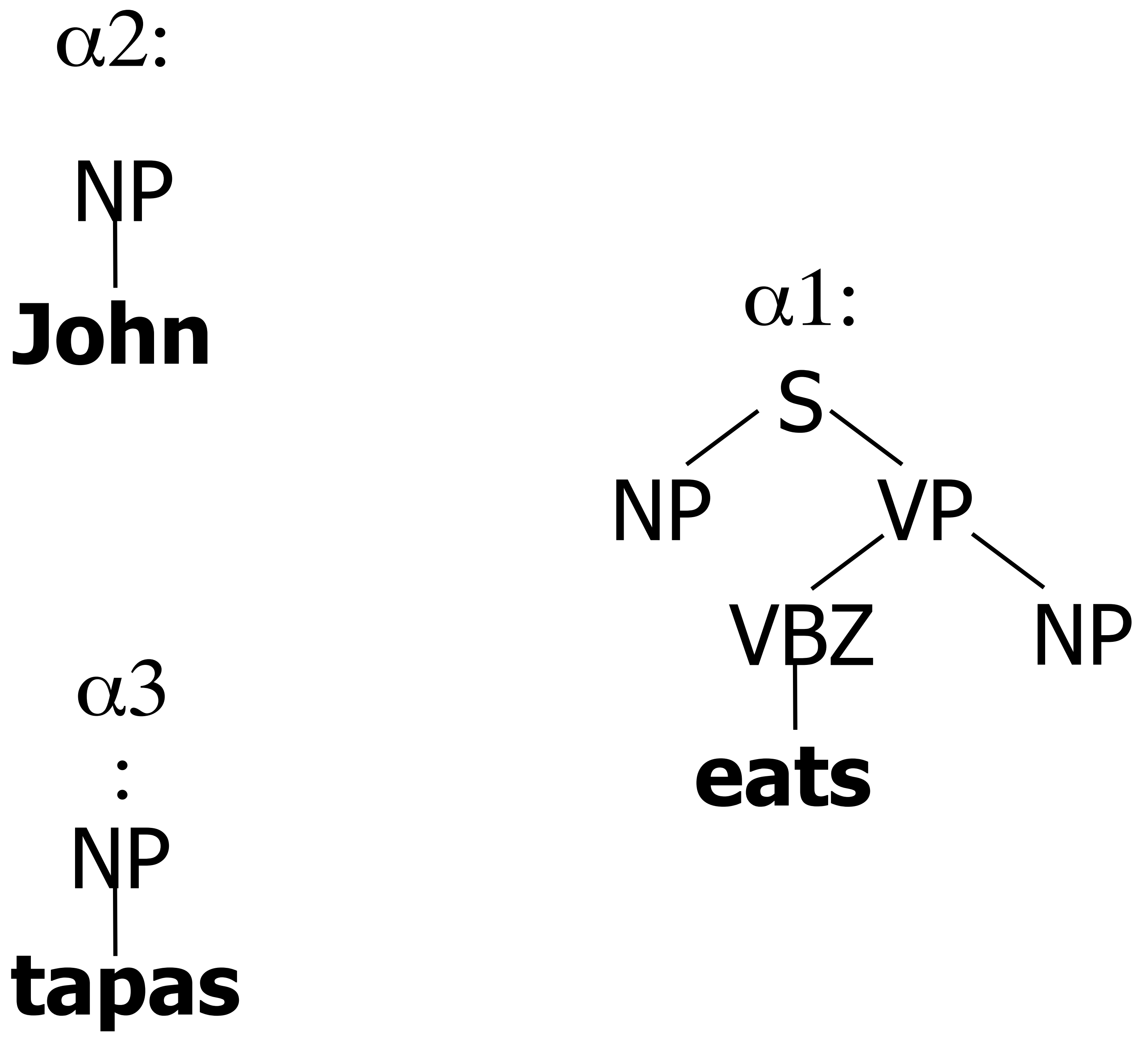
Derivation tree:

$$\begin{tikzcd}[column sep=small] & \alpha1 \arrow[dl, no head] \arrow[dr, no head] & \\ \alpha2 & & \alpha3 \end{tikzcd}$$

Tree-Substitution Grammar

- **TAG without adjunction**
  = Tree-substitution grammar.
  - elementary objects = trees.
  - recursive operation: substitution
- **Substitution alone does not give us anything beyond context-free grammar.**

A small TSG lexicon


α2:
NP
John
α1:
S
NP
VP
VBZ
NP
eats
α3
:
NP
tapas