

Statistical Natural Language Parsing

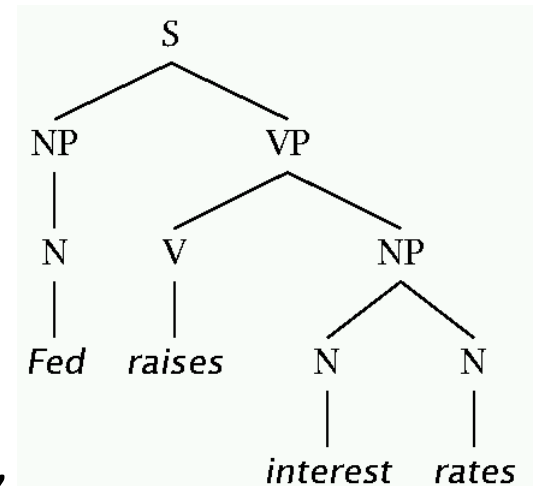
Two views of syntactic structure



Two views of linguistic structure:

1. Constituency (phrase structure)

- Phrase structure organizes words into nested constituents.
- How do we know what is a **constituent**? (Not that linguists don't argue about some cases.)
 - Distribution: a constituent behaves as a unit that can appear in different places:
 - John talked [to the children] [about drugs].
 - John talked [about drugs] [to the children].
 - *John talked drugs to the children about
 - Substitution/expansion/pro-forms:
 - I sat [on the box/right on top of the box/there].
 - Coordination, regular internal structure, no intrusion, fragments, semantics, ...

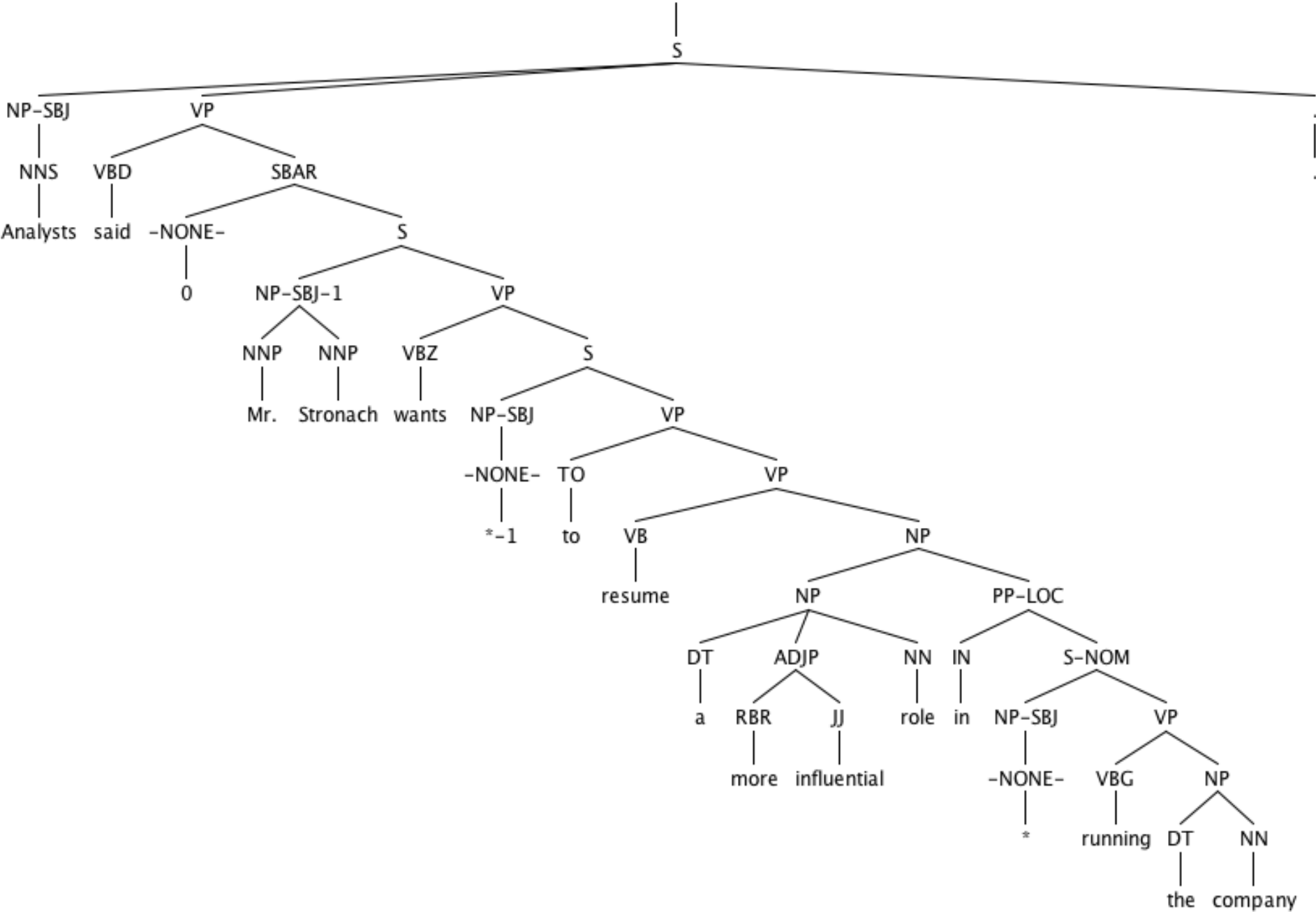




Two views of linguistic structure:

1. Constituency (phrase structure)

- Phrase structure organizes words into nested constituents.
- How do we know what is a **constituent**? (Not that linguists don't argue about some cases.)
 - Distribution: a constituent behaves as a unit that can appear in different places:
 - John talked [to the children] [about drugs].
 - John talked [about drugs] [to the children].
 - *John talked drugs to the children about
 - Substitution/expansion/pro-forms:
 - I sat [on the box/right on top of the box/there].
 - Coordination, regular internal structure, no intrusion, fragments, semantics, ...





Headed phrase structure

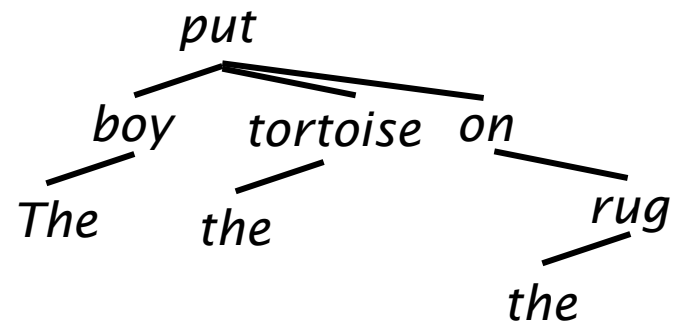
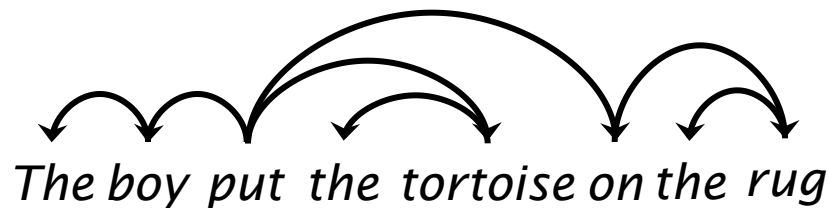
- $VP \rightarrow \dots VB^* \dots$
- $NP \rightarrow \dots NN^* \dots$
- $ADJP \rightarrow \dots JJ^* \dots$
- $ADVP \rightarrow \dots RB^* \dots$
- $SBAR(Q) \rightarrow S | SINV | SQ \rightarrow \dots NP VP \dots$
- Plus minor phrase types:
 - QP (quantifier phrase in NP), CONJP (multi word constructions: *as well as*), INTJ (interjections), etc.



Two views of linguistic structure:

2. Dependency structure

- Dependency structure shows which words depend on (modify or are arguments of) which other words.





Two views of linguistic structure:

2. Dependency structure

- Dependency structure shows which words depend on (modify or are arguments of) which other words.

The boy put the tortoise on the rug

Statistical Natural Language Parsing

Two views of syntactic structure