

[illegible]

Restricting the grammar form for efficient parsing



Chomsky Normal Form

- All rules are of the form $X \rightarrow YZ$ or $X \rightarrow w$
 - $X, Y, Z \in N$ and $w \in T$
- A transformation to this form doesn't change the weak generative capacity of a CFG
 - That is, it recognizes the same language
 - But maybe with different trees
- Empties and unaries are removed recursively
- n -ary rules are divided by introducing new nonterminals ($n > 2$)



A phrase structure grammar

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$VP \rightarrow V NP PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$NP \rightarrow e$

$PP \rightarrow P NP$

$N \rightarrow \textit{people}$

$N \rightarrow \textit{fish}$

$N \rightarrow \textit{tanks}$

$N \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$



Chomsky Normal Form steps

$S \rightarrow NP VP$

$S \rightarrow VP$

$VP \rightarrow V NP$

$VP \rightarrow V$

$VP \rightarrow V NP PP$

$VP \rightarrow V PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP$

$NP \rightarrow NP PP$

$NP \rightarrow PP$

$NP \rightarrow N$

$PP \rightarrow P NP$

$PP \rightarrow P$

$N \rightarrow \textit{people}$

$N \rightarrow \textit{fish}$

$N \rightarrow \textit{tanks}$

$N \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$



Chomsky Normal Form steps

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$S \rightarrow V NP$

$VP \rightarrow V$

$S \rightarrow V$

$VP \rightarrow V NP PP$

$S \rightarrow V NP PP$

$VP \rightarrow V PP$

$S \rightarrow V PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP$

$NP \rightarrow NP PP$

$NP \rightarrow PP$

$NP \rightarrow N$

$PP \rightarrow P NP$

$PP \rightarrow P$

$N \rightarrow \textit{people}$

$N \rightarrow \textit{fish}$

$N \rightarrow \textit{tanks}$

$N \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$



Chomsky Normal Form steps

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$S \rightarrow V NP$

$VP \rightarrow V$

$VP \rightarrow V NP PP$

$S \rightarrow V NP PP$

$VP \rightarrow V PP$

$S \rightarrow V PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP$

$NP \rightarrow NP PP$

$NP \rightarrow PP$

$NP \rightarrow N$

$PP \rightarrow P NP$

$PP \rightarrow P$

$N \rightarrow \textit{people}$

$N \rightarrow \textit{fish}$

$N \rightarrow \textit{tanks}$

$N \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$S \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$S \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$S \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$



Chomsky Normal Form steps

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$S \rightarrow V NP$

$VP \rightarrow V NP PP$

$S \rightarrow V NP PP$

$VP \rightarrow V PP$

$S \rightarrow V PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP$

$NP \rightarrow NP PP$

$NP \rightarrow PP$

$NP \rightarrow N$

$PP \rightarrow P NP$

$PP \rightarrow P$

$N \rightarrow \textit{people}$

$N \rightarrow \textit{fish}$

$N \rightarrow \textit{tanks}$

$N \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$S \rightarrow \textit{people}$

$VP \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$S \rightarrow \textit{fish}$

$VP \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$S \rightarrow \textit{tanks}$

$VP \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$



Chomsky Normal Form steps

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$S \rightarrow V NP$

$VP \rightarrow V NP PP$

$S \rightarrow V NP PP$

$VP \rightarrow V PP$

$S \rightarrow V PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP PP$

$NP \rightarrow P NP$

$PP \rightarrow P NP$

$NP \rightarrow \textit{people}$

$NP \rightarrow \textit{fish}$

$NP \rightarrow \textit{tanks}$

$NP \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$S \rightarrow \textit{people}$

$VP \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$S \rightarrow \textit{fish}$

$VP \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$S \rightarrow \textit{tanks}$

$VP \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$

$PP \rightarrow \textit{with}$



Chomsky Normal Form steps

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$S \rightarrow V NP$

$VP \rightarrow V @VP_V$

$@VP_V \rightarrow NP PP$

$S \rightarrow V @S_V$

$@S_V \rightarrow NP PP$

$VP \rightarrow V PP$

$S \rightarrow V PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP PP$

$NP \rightarrow P NP$

$PP \rightarrow P NP$

$NP \rightarrow \textit{people}$

$NP \rightarrow \textit{fish}$

$NP \rightarrow \textit{tanks}$

$NP \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$S \rightarrow \textit{people}$

$VP \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$S \rightarrow \textit{fish}$

$VP \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$S \rightarrow \textit{tanks}$

$VP \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$

$PP \rightarrow \textit{with}$



A phrase structure grammar

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$VP \rightarrow V NP PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP PP$

$NP \rightarrow N$

$NP \rightarrow e$

$PP \rightarrow P NP$

$N \rightarrow \textit{people}$

$N \rightarrow \textit{fish}$

$N \rightarrow \textit{tanks}$

$N \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$



Chomsky Normal Form steps

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$S \rightarrow V NP$

$VP \rightarrow V @VP_V$

$@VP_V \rightarrow NP PP$

$S \rightarrow V @S_V$

$@S_V \rightarrow NP PP$

$VP \rightarrow V PP$

$S \rightarrow V PP$

$NP \rightarrow NP NP$

$NP \rightarrow NP PP$

$NP \rightarrow P NP$

$PP \rightarrow P NP$

$NP \rightarrow \textit{people}$

$NP \rightarrow \textit{fish}$

$NP \rightarrow \textit{tanks}$

$NP \rightarrow \textit{rods}$

$V \rightarrow \textit{people}$

$S \rightarrow \textit{people}$

$VP \rightarrow \textit{people}$

$V \rightarrow \textit{fish}$

$S \rightarrow \textit{fish}$

$VP \rightarrow \textit{fish}$

$V \rightarrow \textit{tanks}$

$S \rightarrow \textit{tanks}$

$VP \rightarrow \textit{tanks}$

$P \rightarrow \textit{with}$

$PP \rightarrow \textit{with}$

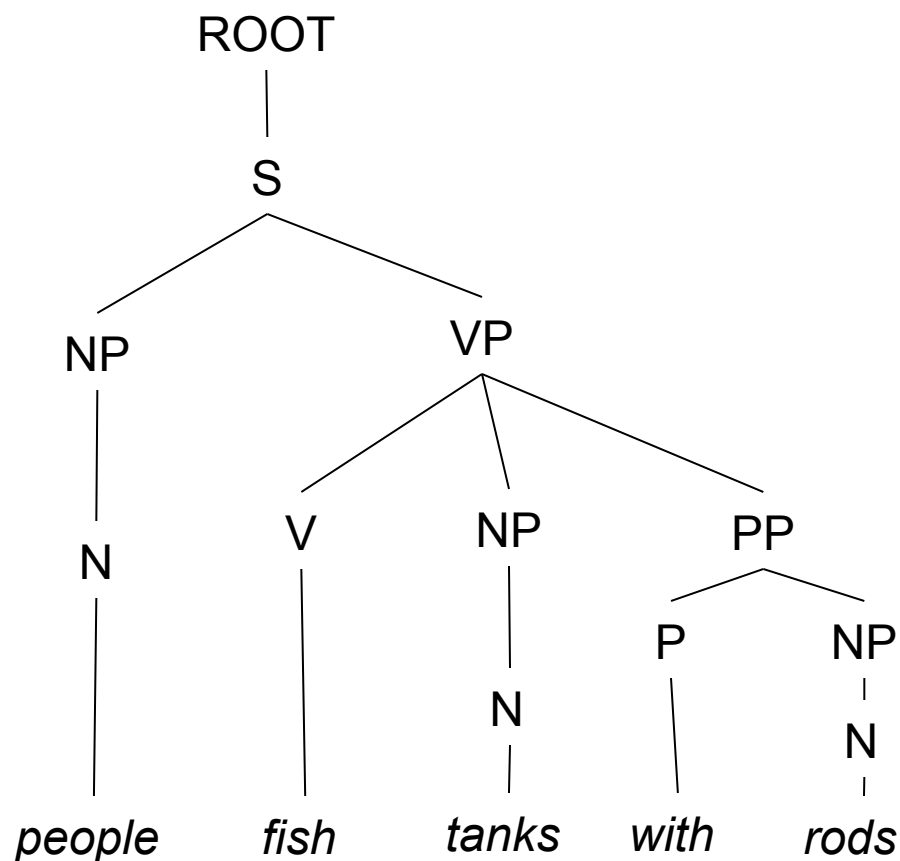


Chomsky Normal Form

- You should think of this as a transformation for efficient parsing
- With some extra book-keeping in symbol names, you can even reconstruct the same trees with a detransform
- In practice full Chomsky Normal Form is a pain
 - Reconstructing n-aries is easy
 - Reconstructing unaries/empties is trickier
- **Binarization** is crucial for cubic time CFG parsing
- The rest isn't necessary; it just makes the algorithms cleaner and a bit quicker

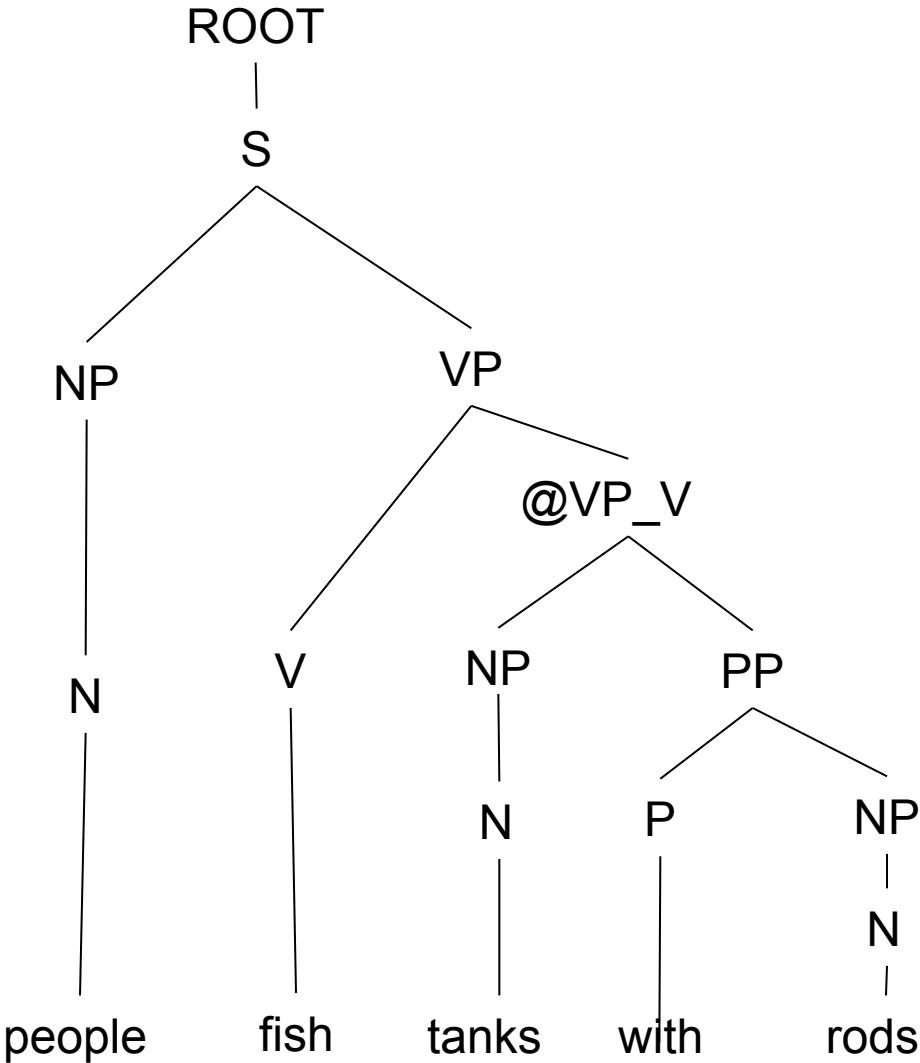


An example: before binarization...



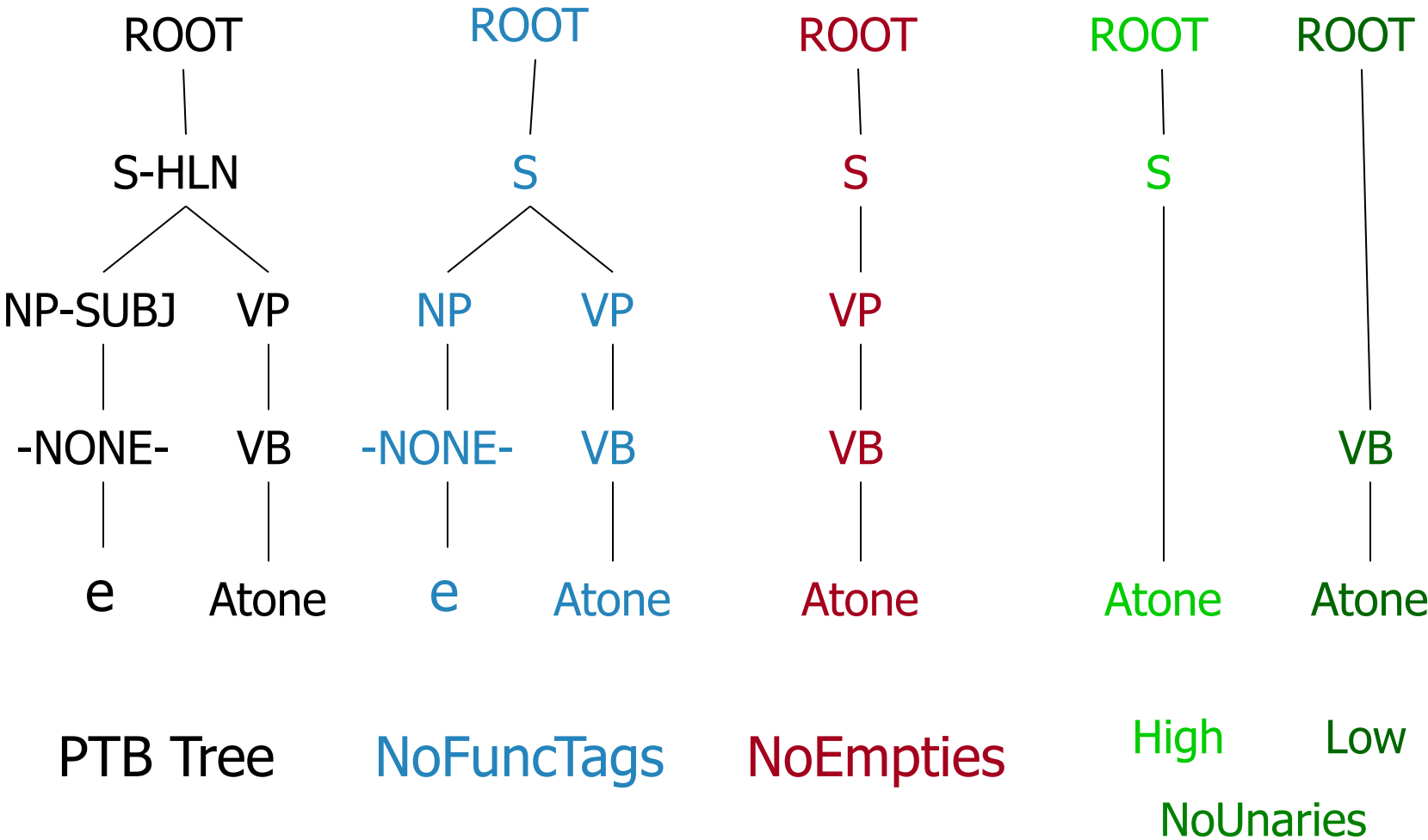


After binarization...

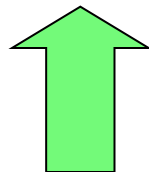




Treebank: empties and unaries

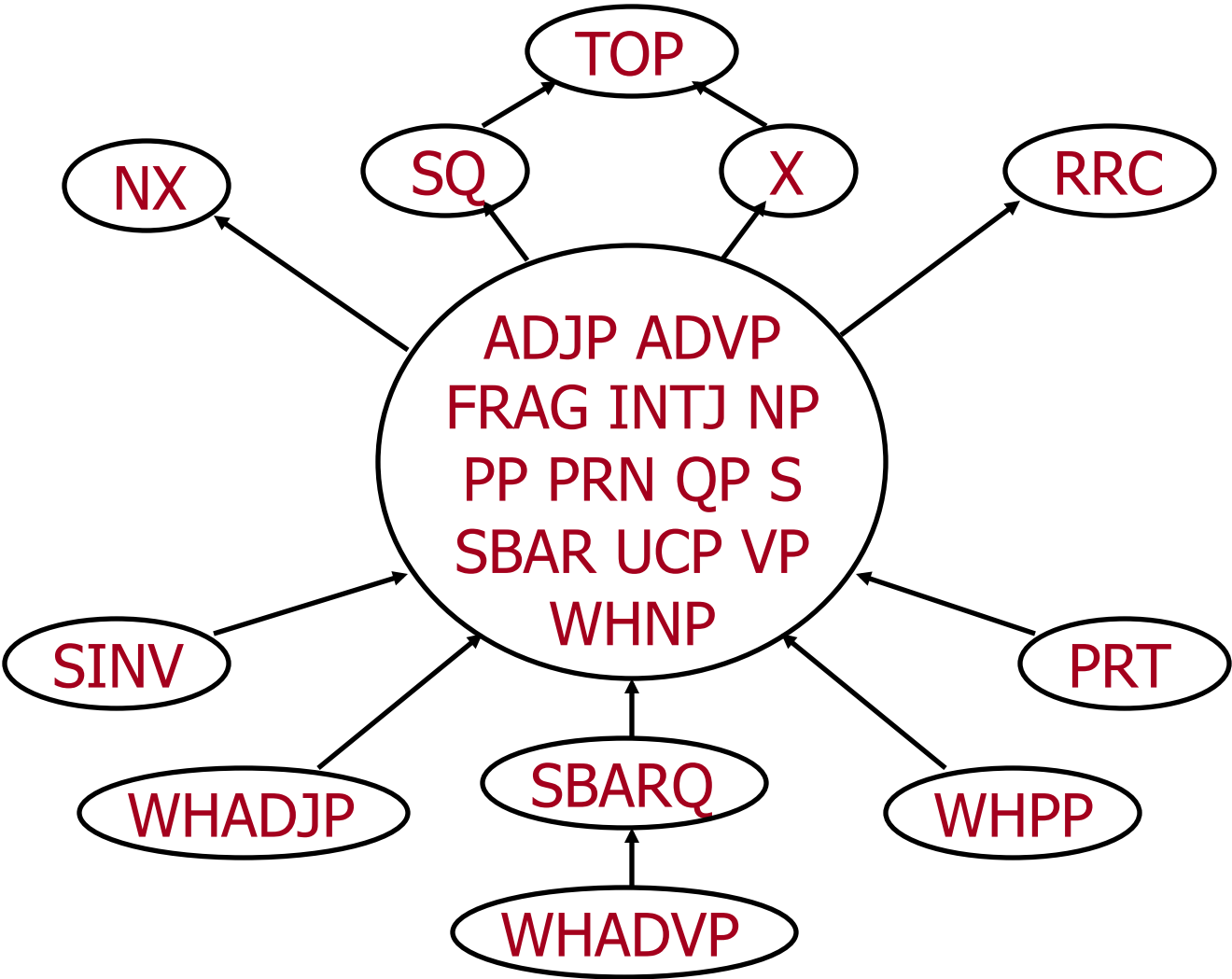


Unary rules: alchemy in the land of treebanks





Same-Span Reachability



NoEmpties

- LST
- CONJP
- NAC

Restricting the grammar form for efficient parsing