Vine Parsing and Minimum Risk Reranking

for Speed and Precision

Markus Dreyer

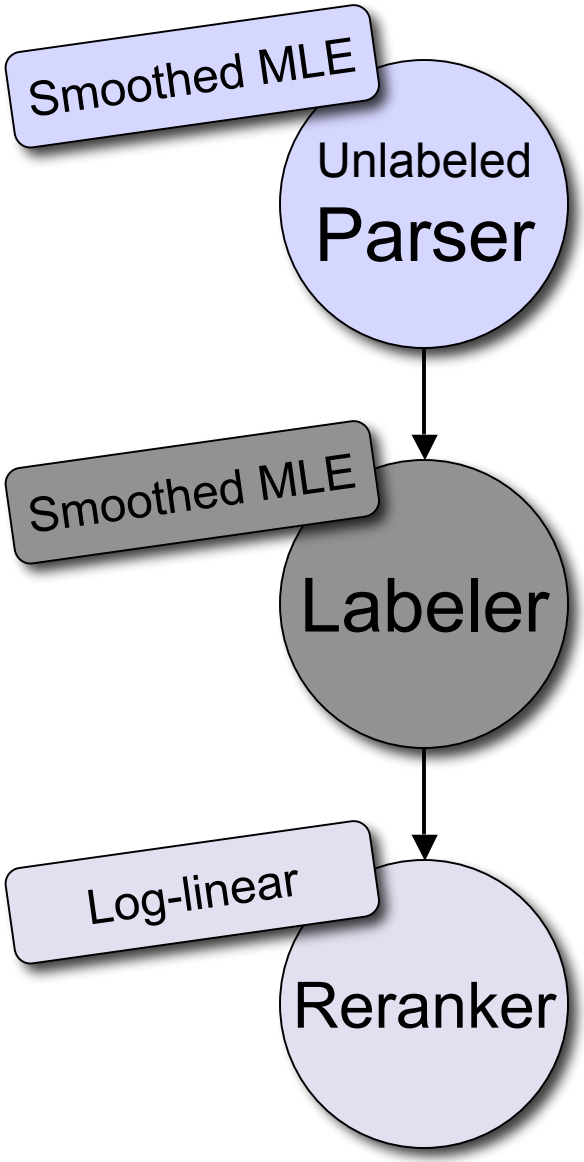
David A. Smith

Noah A. Smith

Johns Hopkins University

Design Goals

Speed
Precision

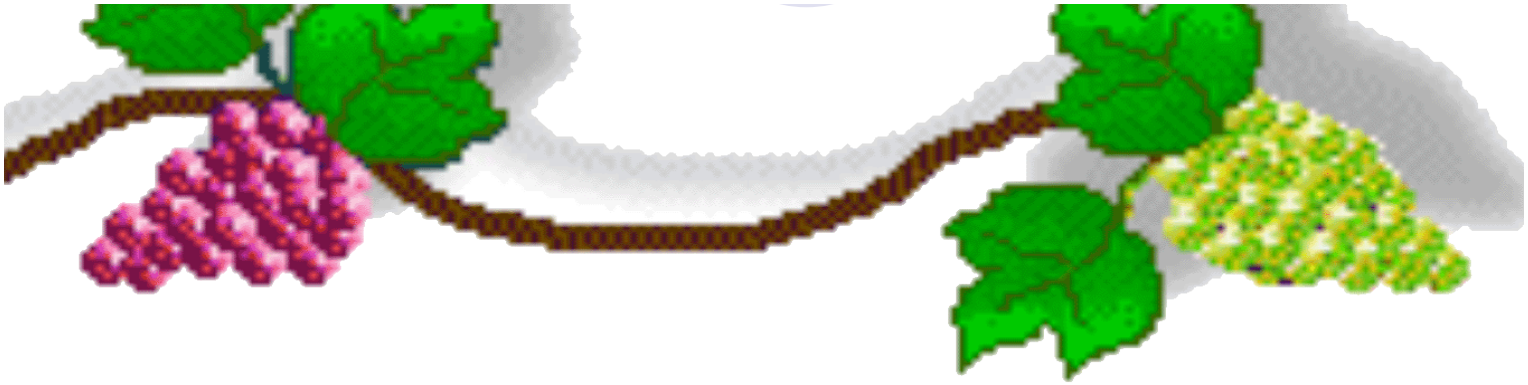


U-best unlabeled parses

U×*L*-best labeled parses

Collins (2000)

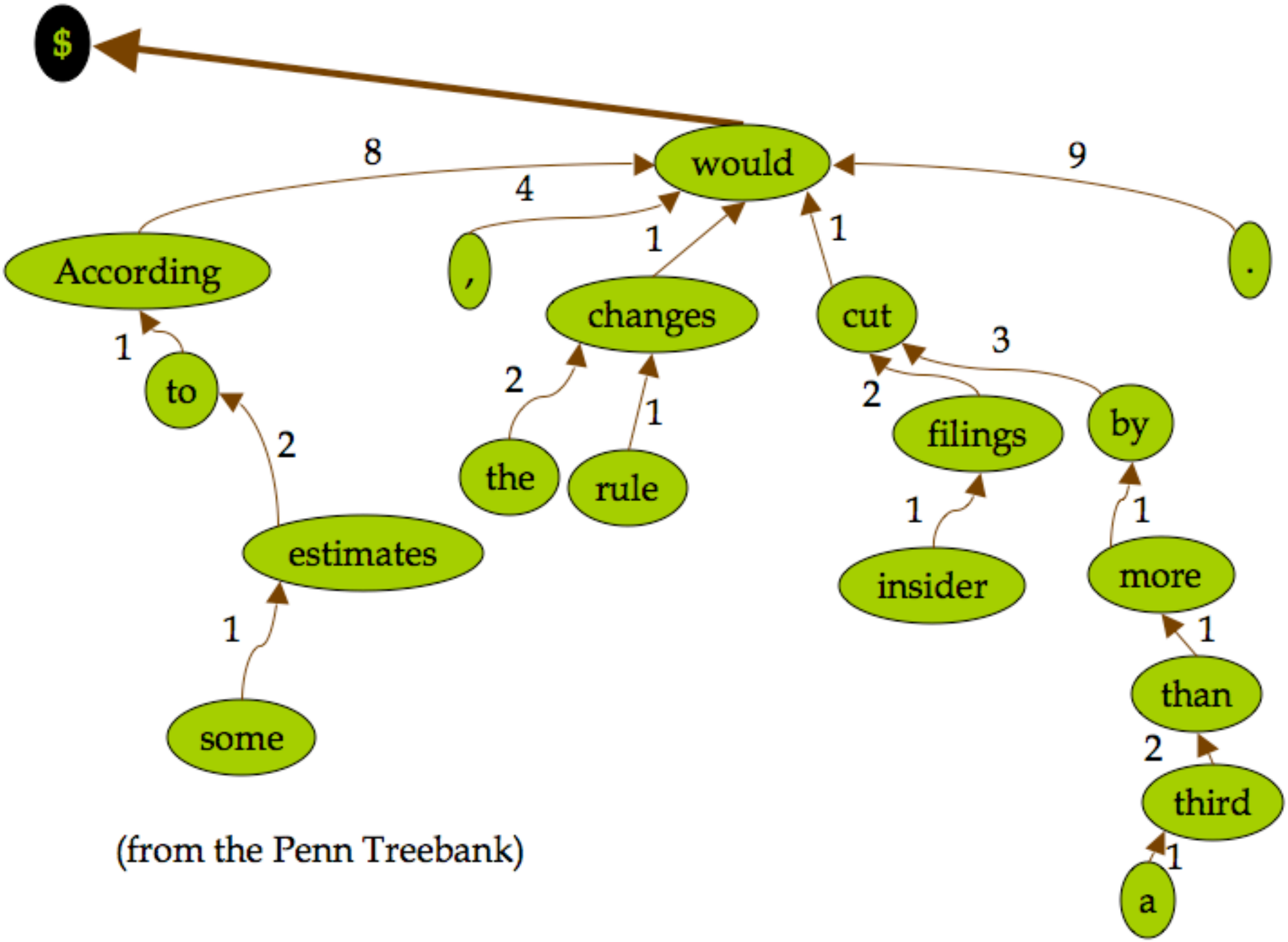
Charniak and Johnson (2005)



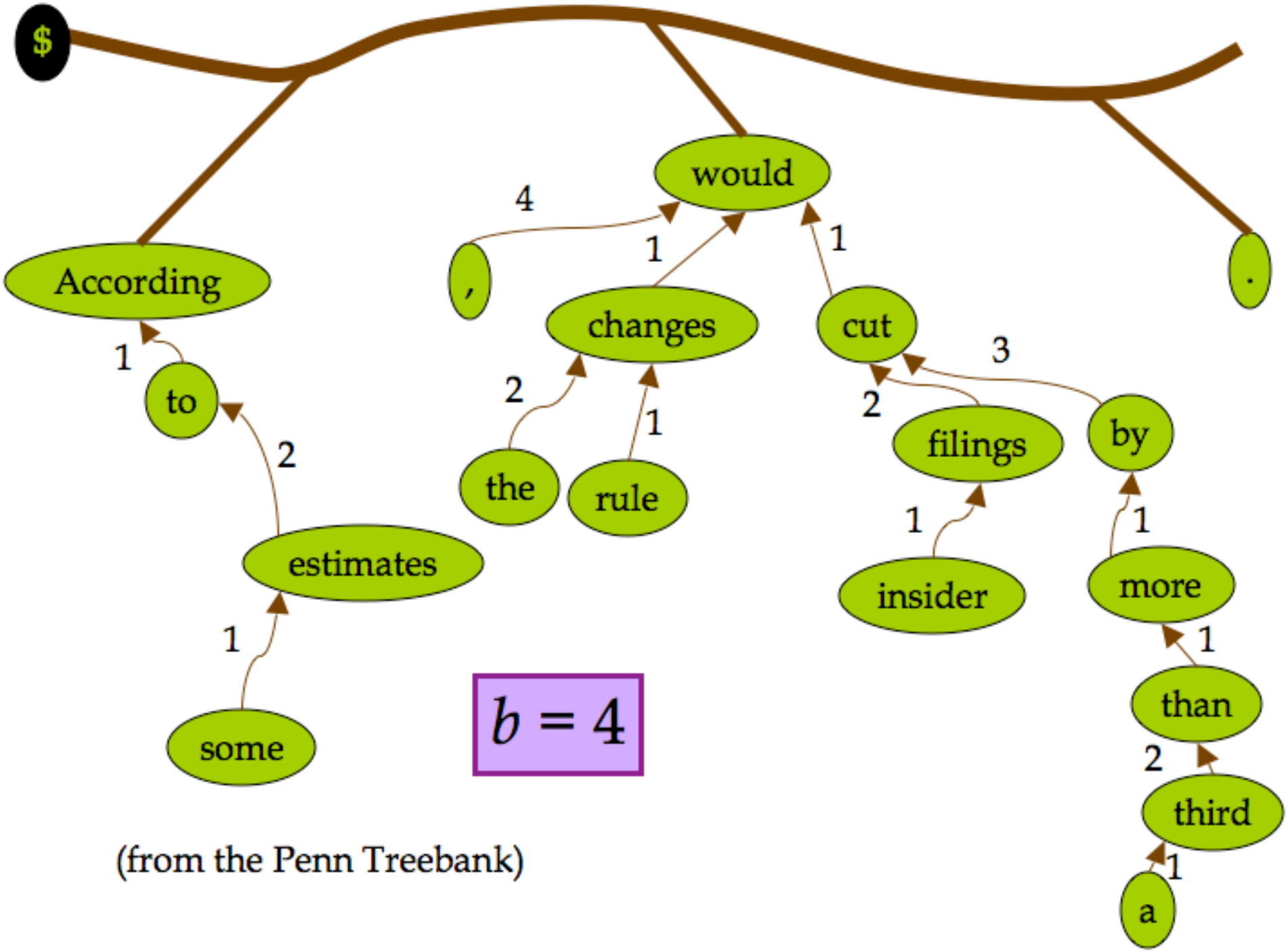
Unlabeled
Parser

Split-Head assumption
Projectivity
Vine Grammar

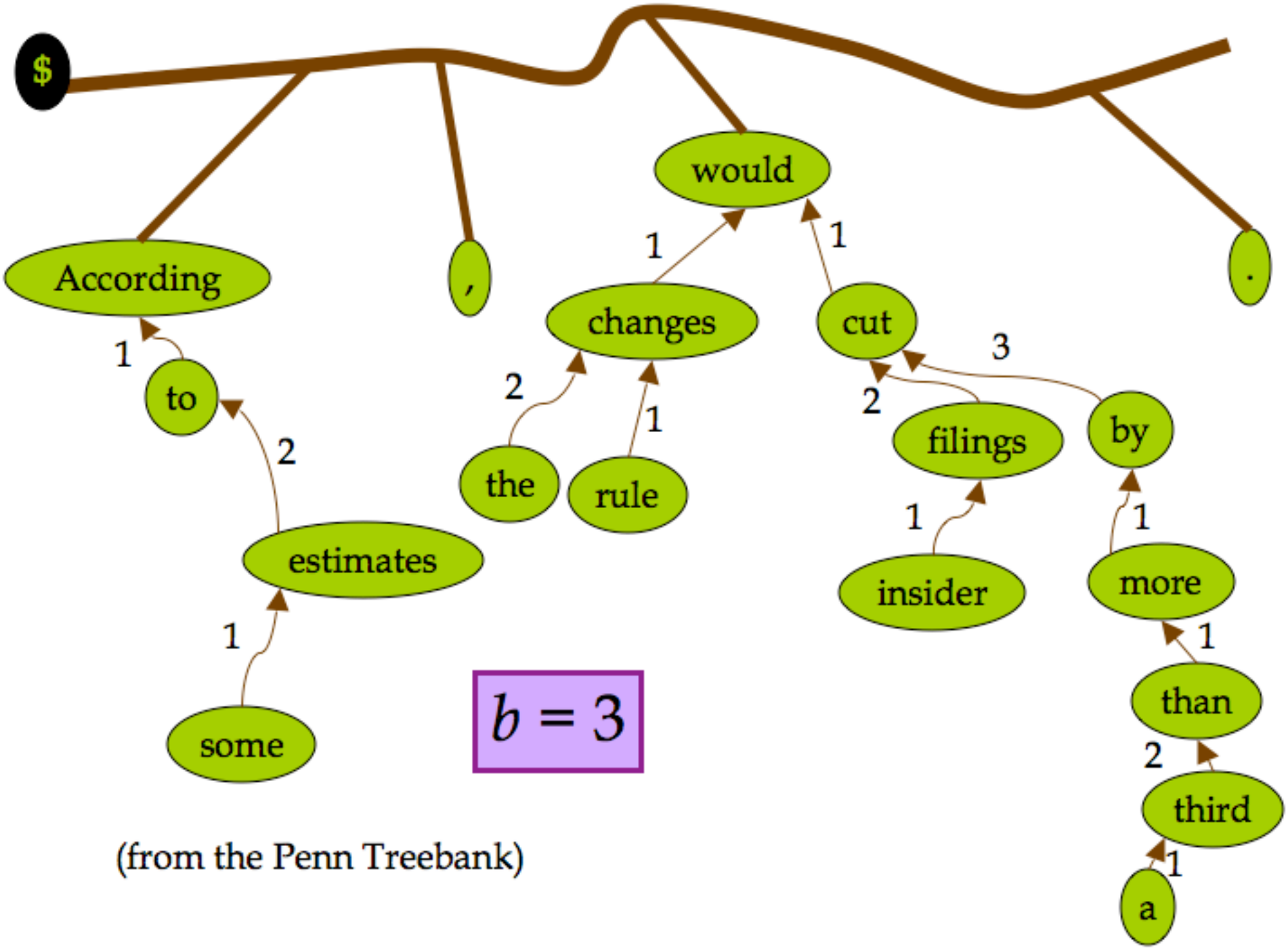
Eisner and N. Smith (2005)



(from the Penn Treebank)

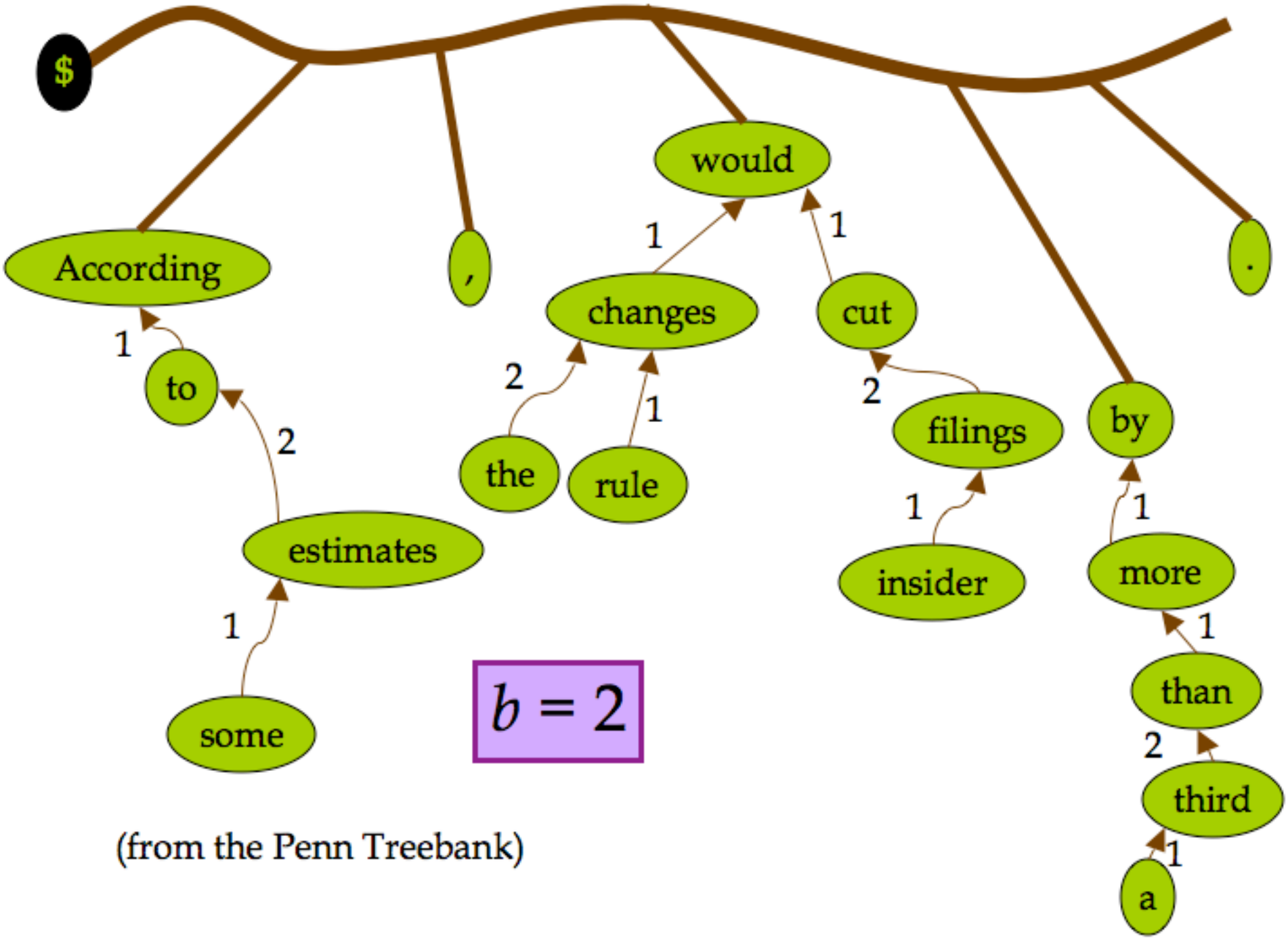


(from the Penn Treebank)

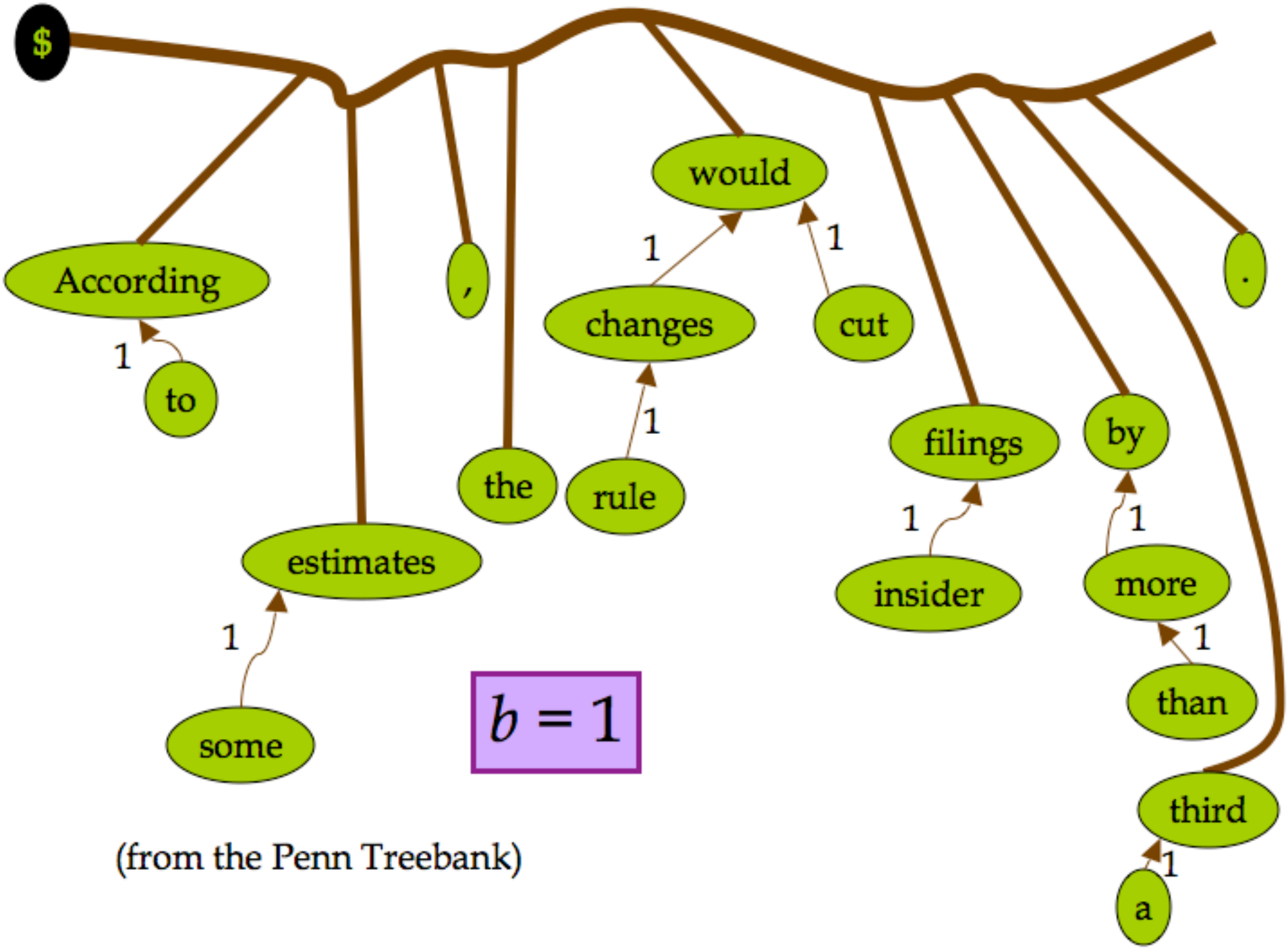


$b = 3$

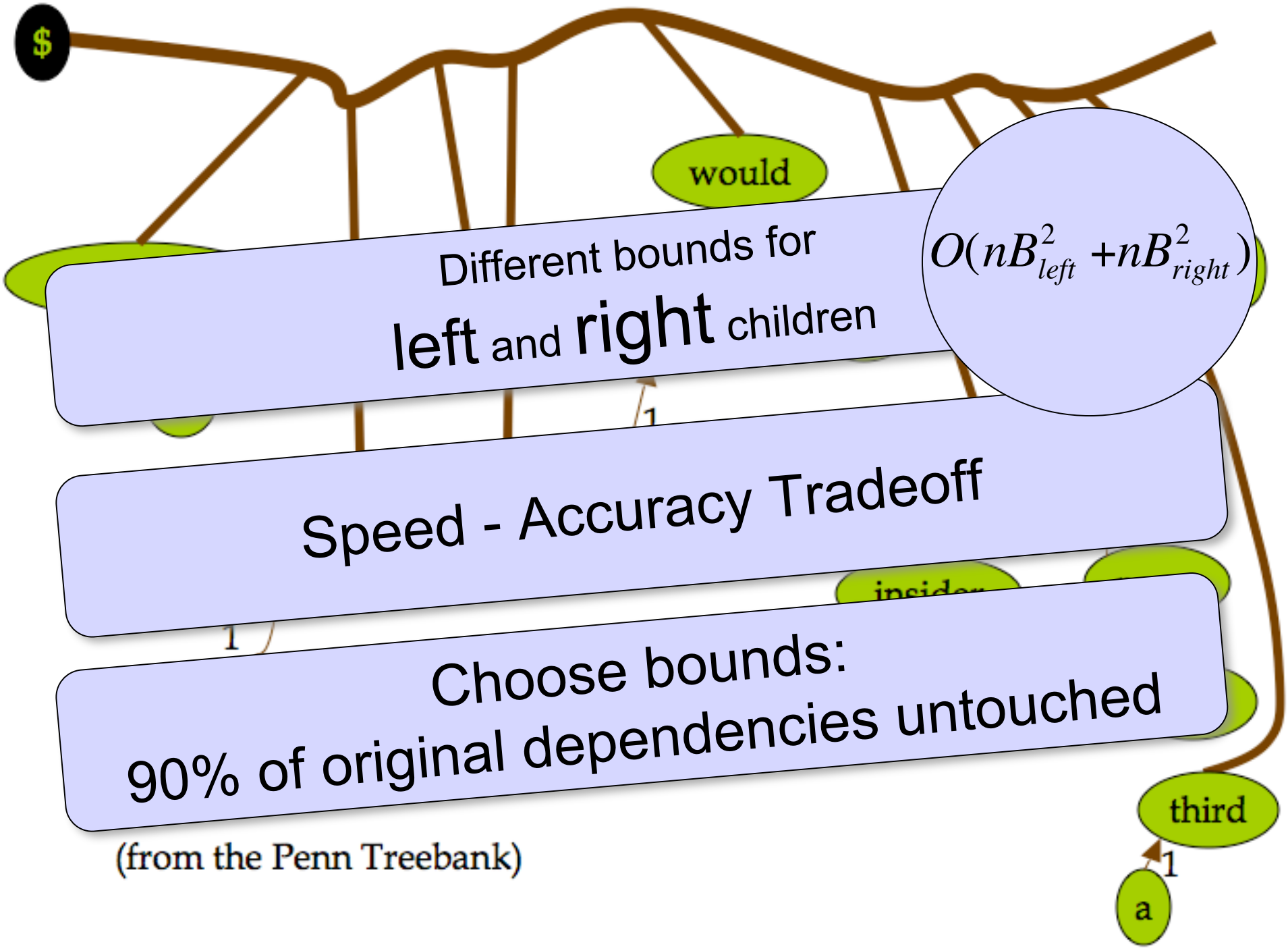
(from the Penn Treebank)



(from the Penn Treebank)



(from the Penn Treebank)



Different bounds for
left and right children

$$O(nB_{left}^2 + nB_{right}^2)$$

Speed - Accuracy Tradeoff

Choose bounds:
90% of original dependencies untouched

(from the Penn Treebank)

Reranker

Minimum Risk Training

Deterministic annealing

D. Smith and Eisner (2006)

Median 67.6

Labeled

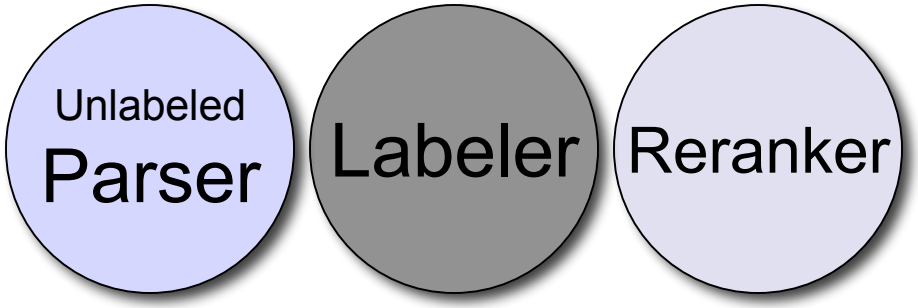
Japanese (82.9), Portuguese (75.3), Bulgarian (74.8), Chinese (71.6), German (71.0)

Median 77.5

Unlabeled

Japanese (86.0), Portuguese (82.4), Bulgarian (82.0), Swedish (79.5), Chinese (77.6)

Summary



- Parsing constraints
- Linear-time inference and decoding
- Minimum Risk reranking
- High precision, mediocre recall

Future Work

- Better estimation
- Better labeler (label bigrams)
- More fine-grained parsing constraints (length bounds given head)