

CG136: Computational Linguistics

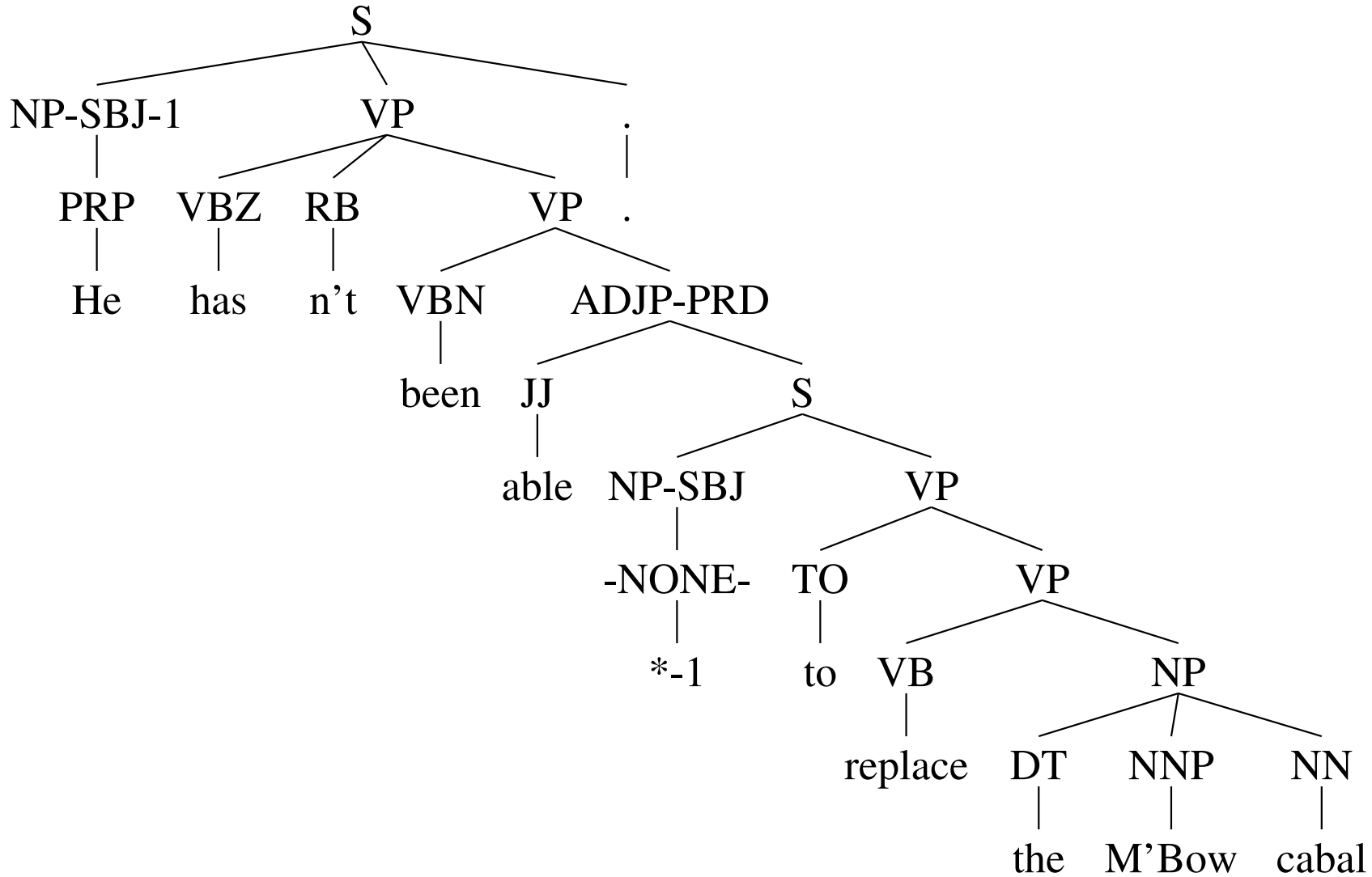
Treebanks and grammars

Mark Johnson

Brown University

April, 2006

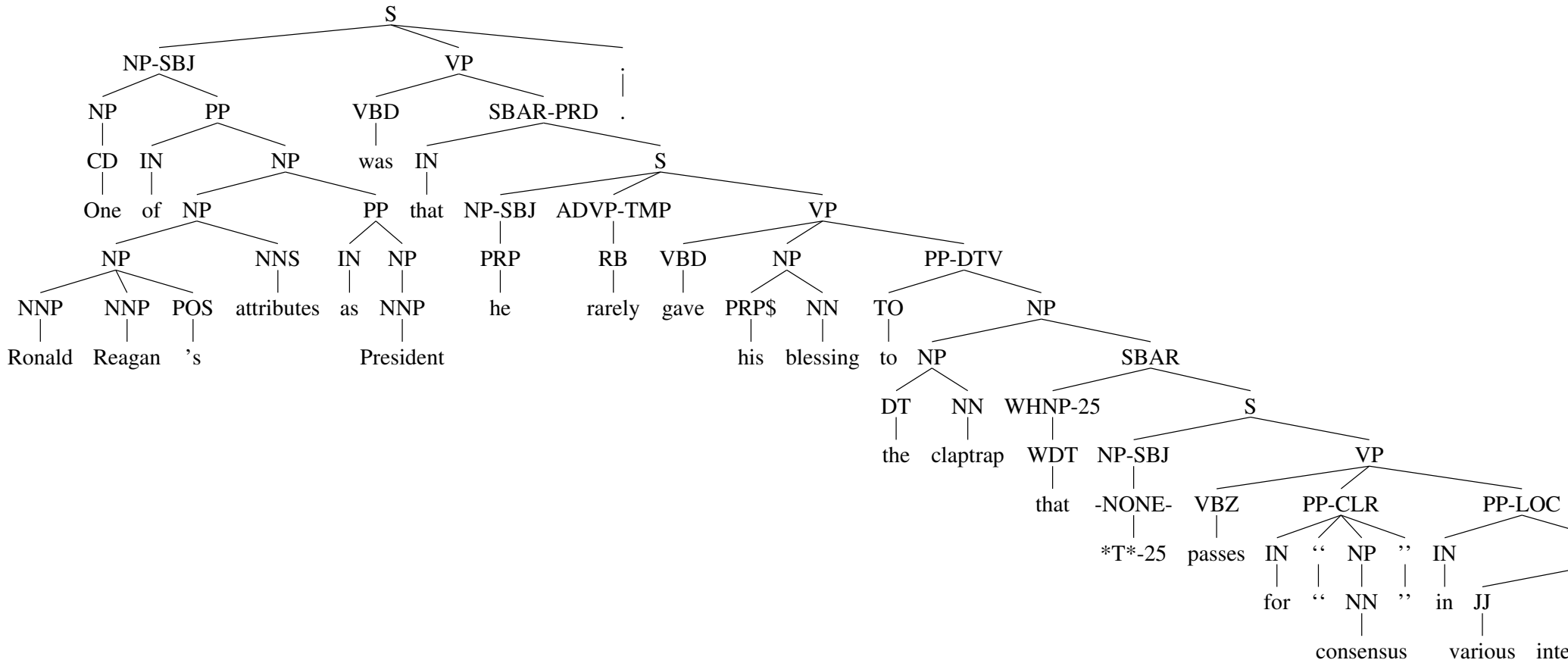
Sample WSJ treebank tree



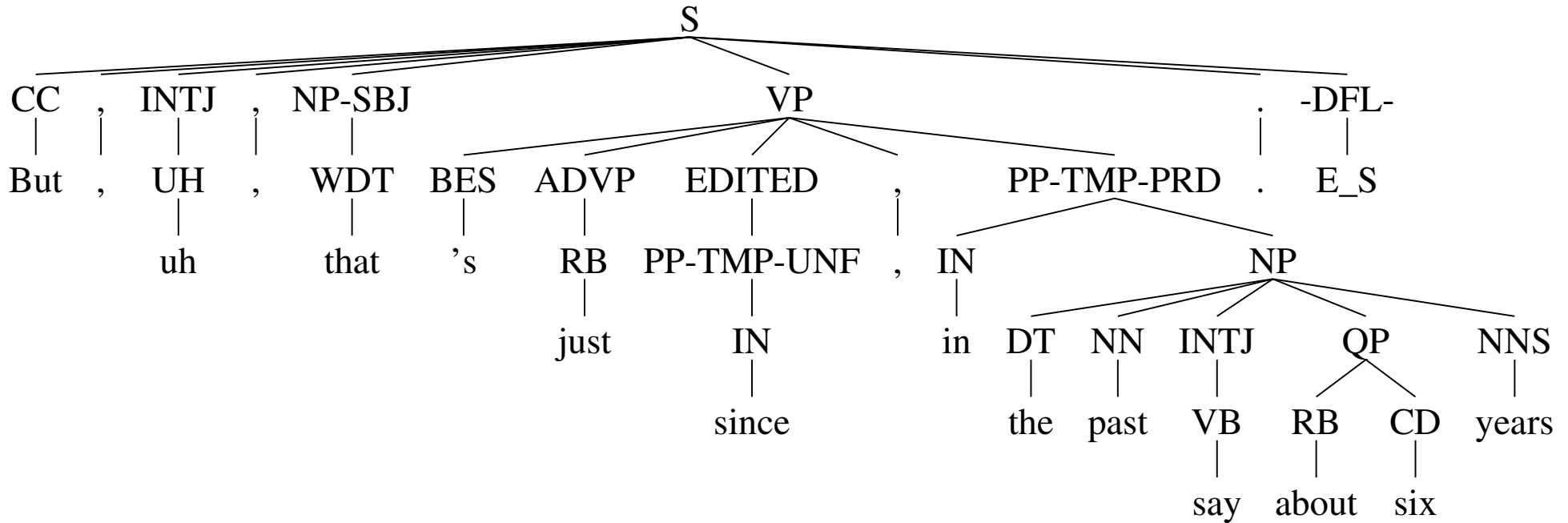
Information annotated in treebank

- Parts of speech
- Phrases and their categories
- Function-argument relationships
- “Empty nodes”, i.e., implicit arguments, argument-structure non-linearity

More usual WSJ treebank tree



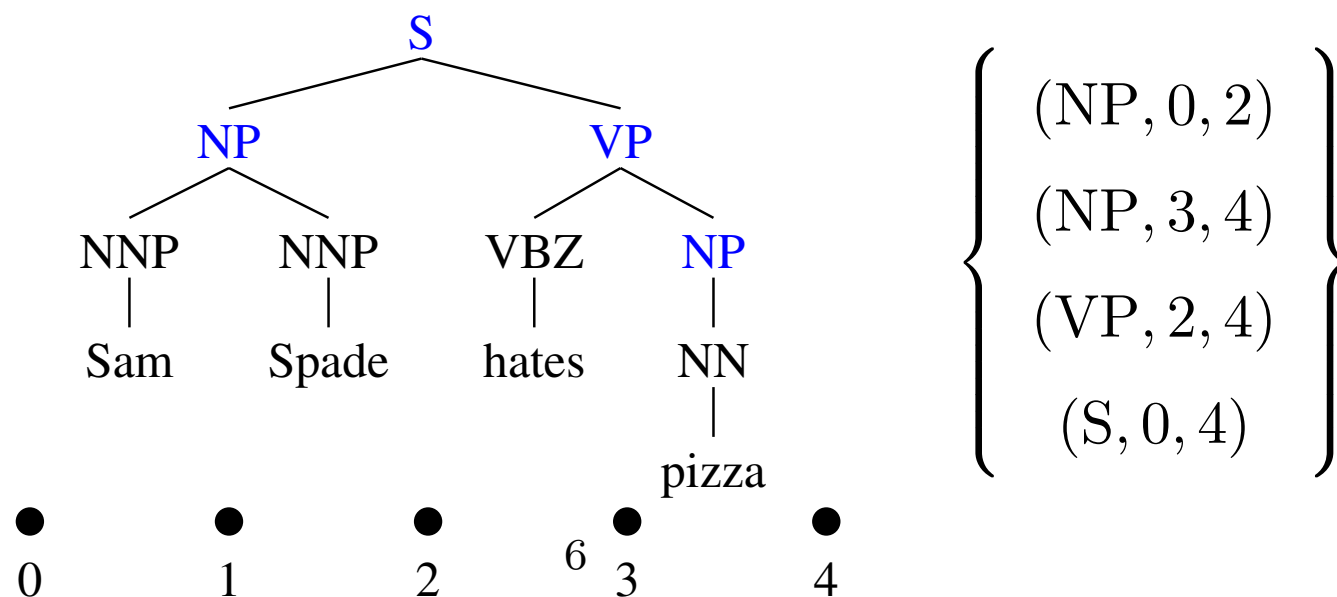
Sample Switchboard treebank tree



- Spontaneous telephone conversations
 - disagreement as to sentence boundaries, and exactly which words wer said
- Disfluencies are annotated
- Slightly different parts of speech to WSJ (ugh!)

Evaluating parser accuracy

- Divide corpus into separate *training data* (WSJ sections 2-21) and *test data* (WSJ section 23) sections
- A tree produced by a parser can be wrong in two ways:
 1. It contains nodes that shouldn't be there (*precision*)
 2. It doesn't contain nodes that should be there (*recall*)
- An edge is a triple (A, i, j) where A is a category label and i, j are left and right string positions respectively



Precision and Recall

- To compute *labeled precision and recall*, let T and G be the set of nonterminal edges in the test (i.e., parser) and gold (i.e., treebank) trees respectively

$$\text{Precision} = \frac{|T| \cap |G|}{|T|}$$

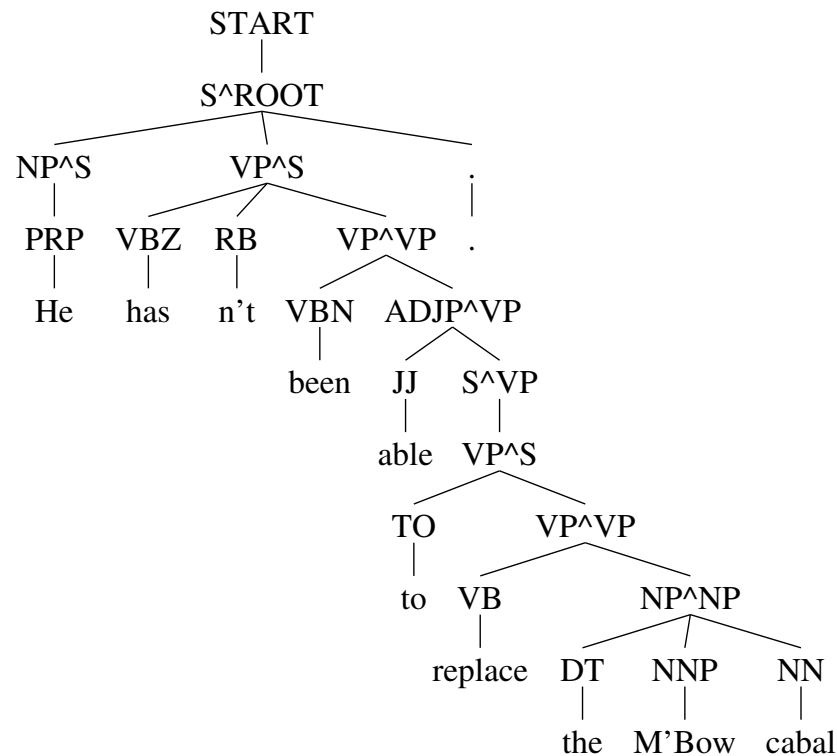
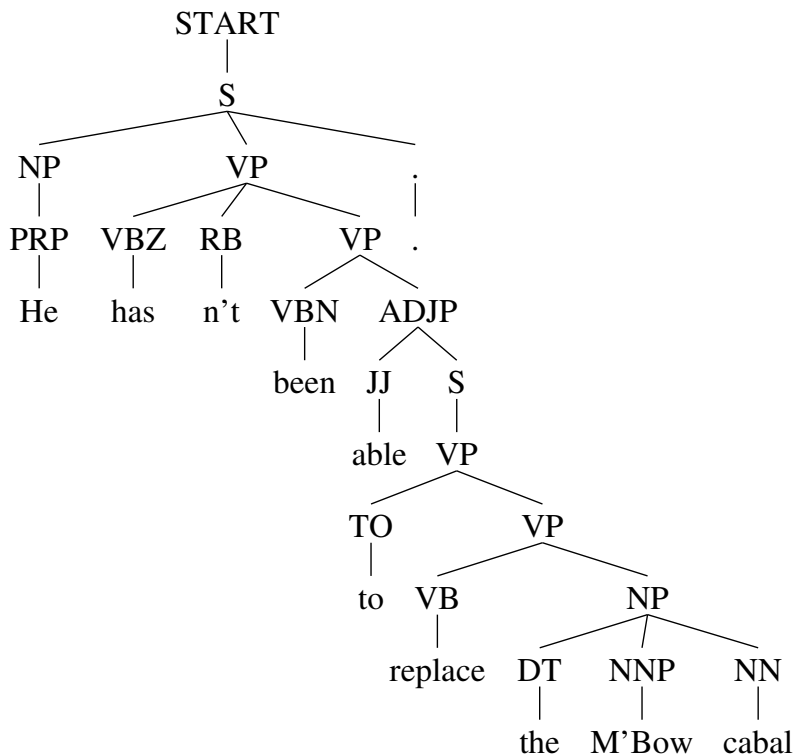
$$\text{Recall} = \frac{|T| \cap |G|}{|G|}$$

- F-score is the *geometric mean* of precision and recall

$$f\text{-score} = \frac{1}{0.5 \frac{1}{\text{Precision}} + 0.5 \frac{1}{\text{Recall}}} = \frac{2 |T| \cap |G|}{|T| + |G|}$$

- Ignoring lexical items (i.e., treating parts of speech as terminals) the treebank grammar contains 14,962 productions, and its MAP parses achieve precision = 0.742 and recall = 0.699

Parent-annotation tree transforms



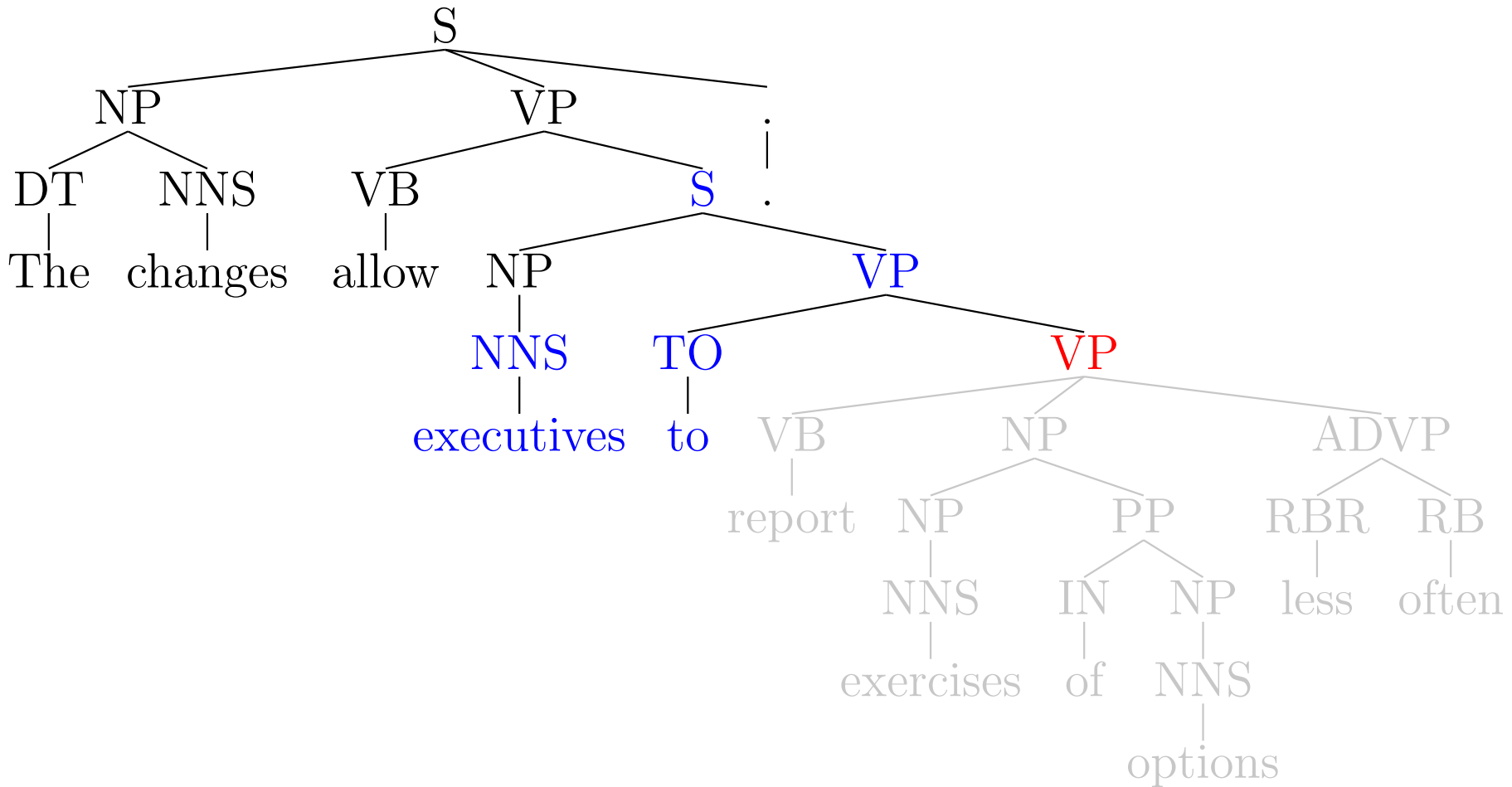
Results using parts-of-speech as terminals:

	# of rules	Precision	Recall	$-\log P(\text{test})$
Original	14,962	0.742	0.700	126,461
Parent-annotated	22,773	0.801	0.787	115,718

Modern statistical parsers

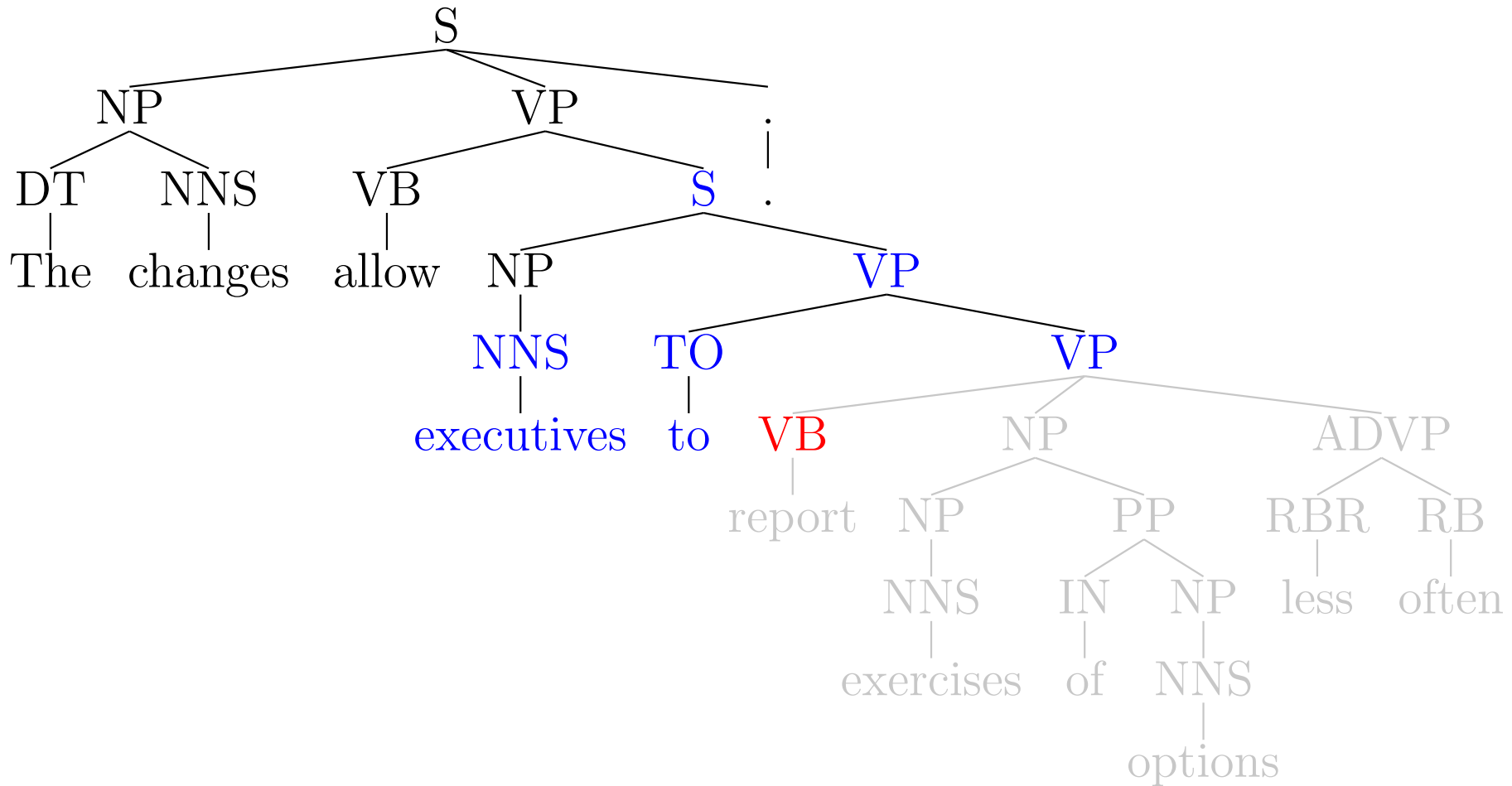
- The Charniak and Collins parsers can be viewed as PCFGs estimated from tree-transformed training data
- Both parsers condition on the lexical context (i.e., words) as well as phrasal categories
- Most combinations of features are never observed in training data \Rightarrow MLE is either 0 or undefined
- Smoothing is absolutely essential! (Performed by an EM procedure using heldout data)

Parsing model from Charniak 2001



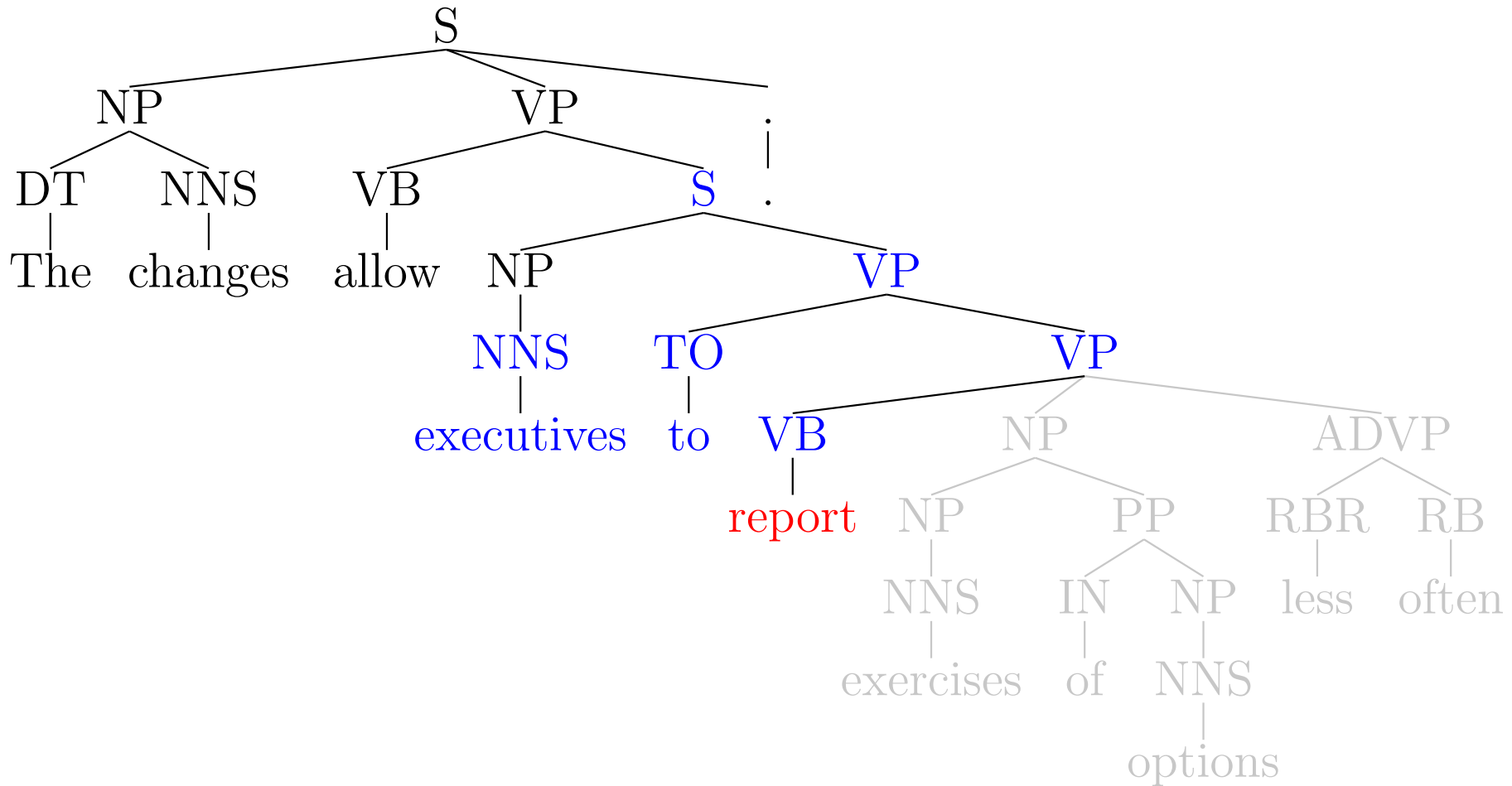
- Predicted node is shown in red
- Conditioning nodes are shown in blue

Parsing model from Charniak 2001



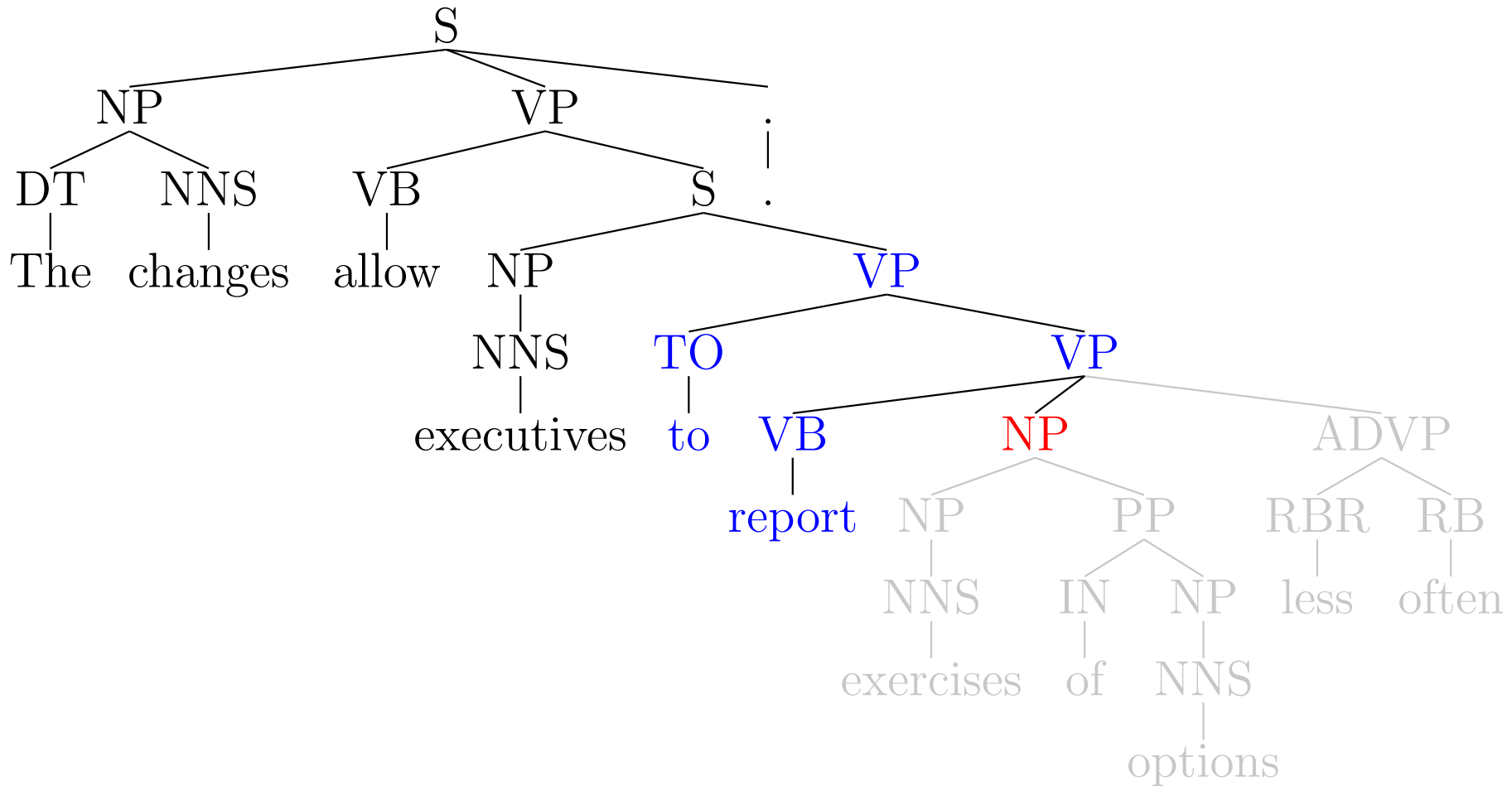
- Predicted node is shown in red
- Conditioning nodes are shown in blue

Parsing model from Charniak 2001



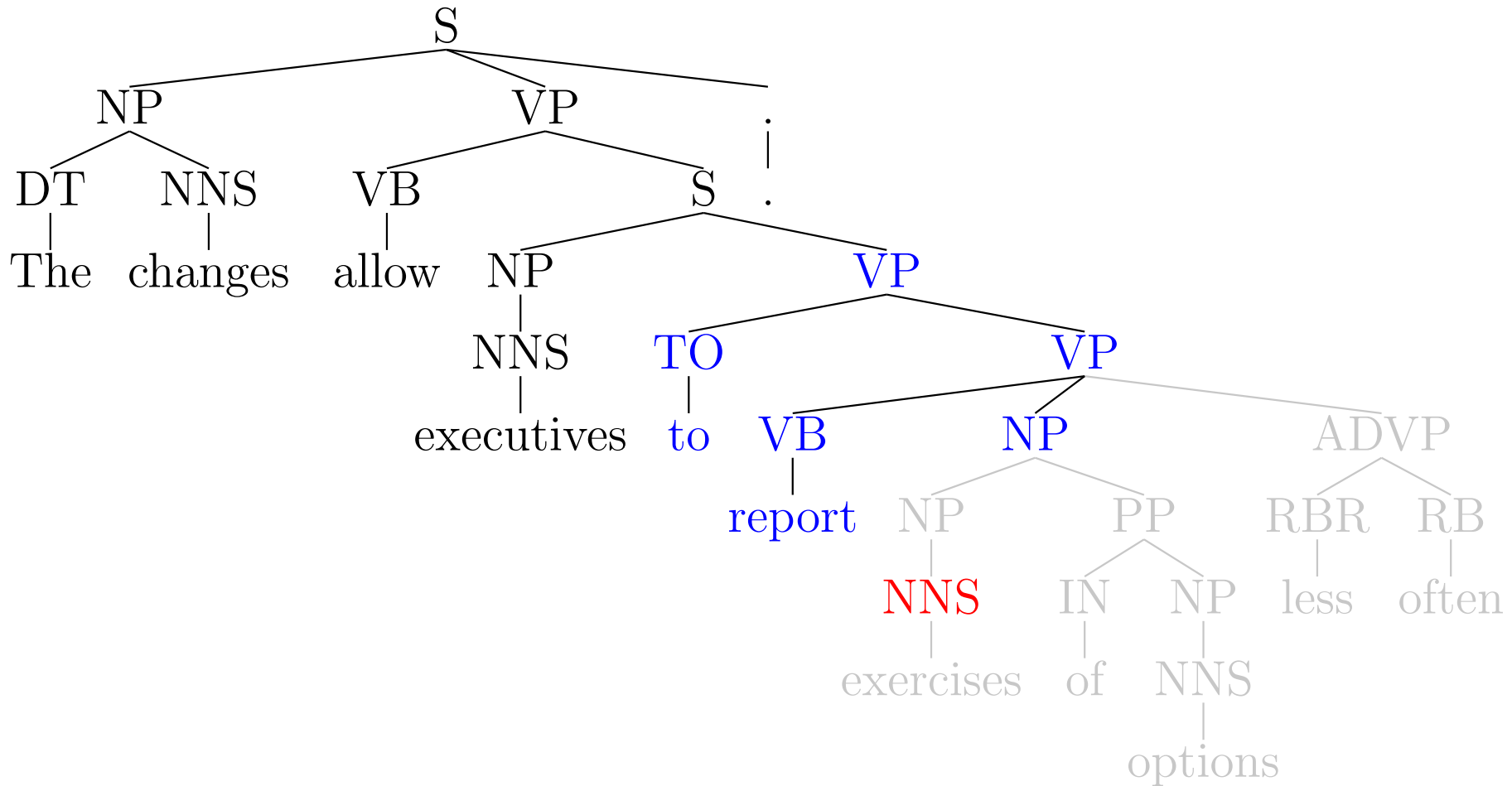
- Predicted node is shown in red
- Conditioning nodes are shown in blue

Parsing model from Charniak 2001



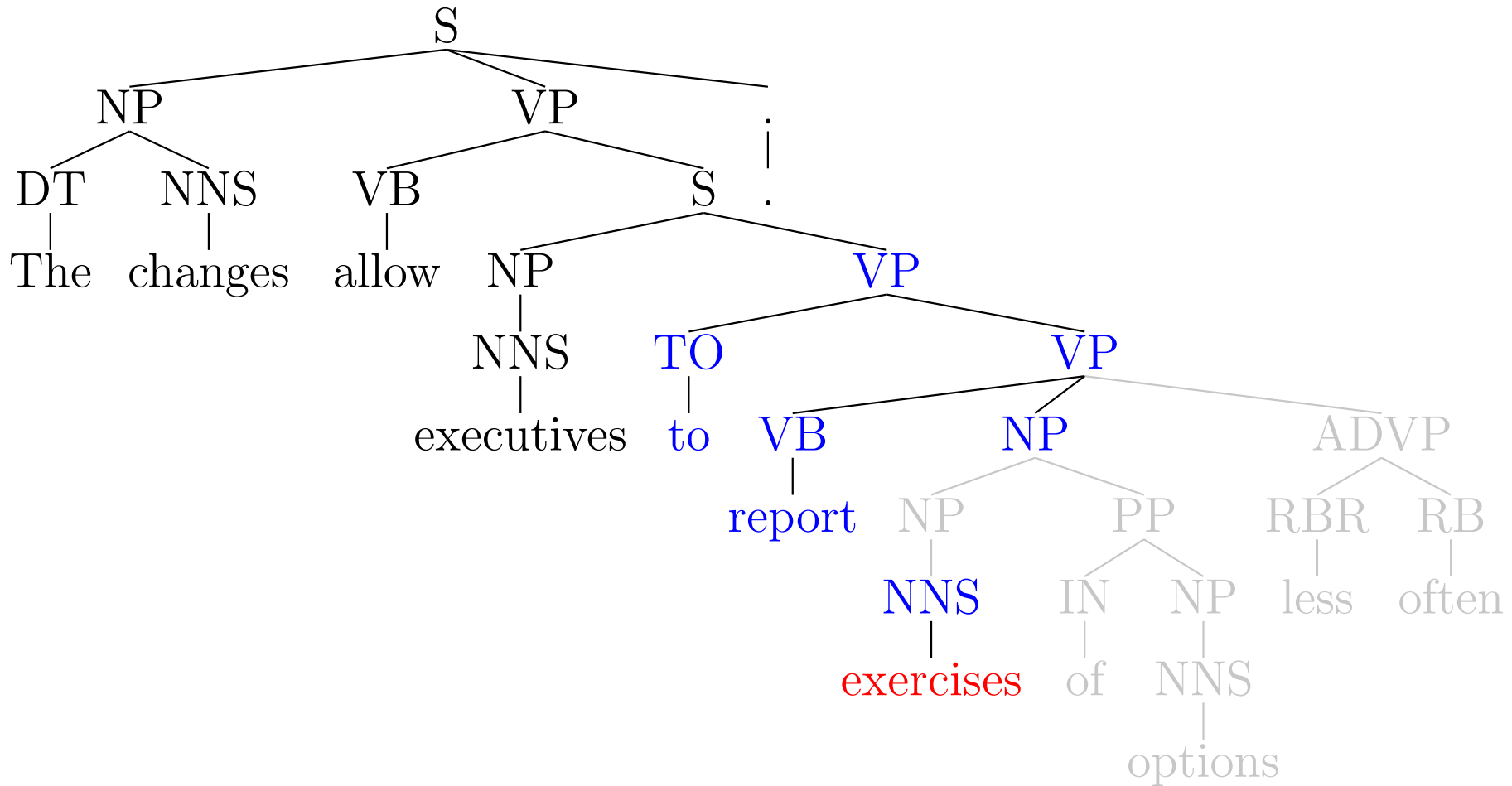
- Predicted node is shown in red
- Conditioning nodes are shown in blue

Parsing model from Charniak 2001



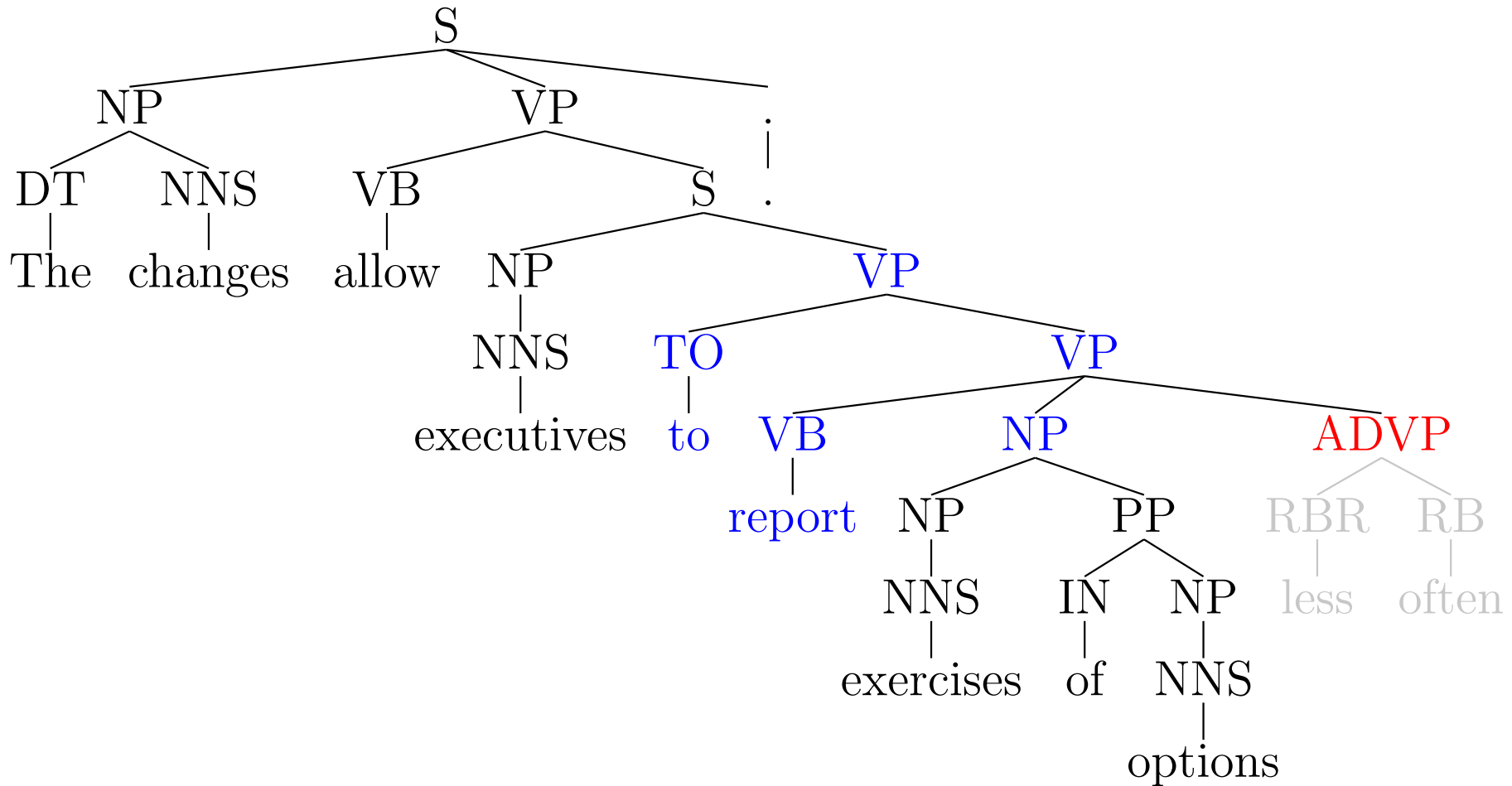
- Predicted node is shown in red
- Conditioning nodes are shown in blue

Parsing model from Charniak 2001



- Predicted node is shown in red
- Conditioning nodes are shown in blue

Parsing model from Charniak 2001



- Predicted node is shown in red
- Conditioning nodes are shown in blue