Rational Parsing

John Hale
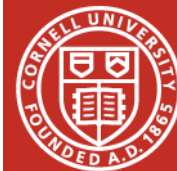
Cornell University
Linguistics Department

Phrase structure
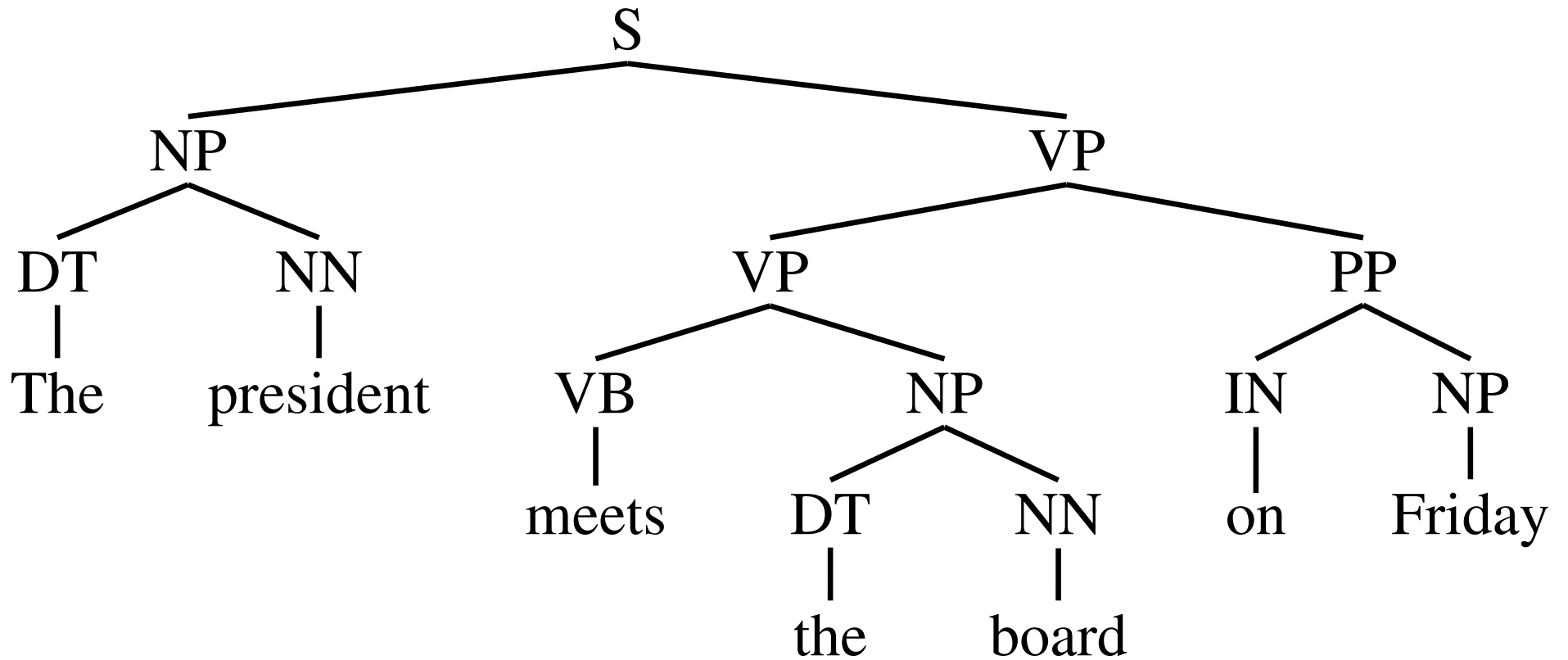
```
                              S
                ┌─────────────┴──────────────┐
               NP                            VP
          ┌─────┴─────┐              ┌────────┴────────┐
         DT          NN             VP                PP
          │           │         ┌────┴────┐       ┌────┴────┐
        The       president    VB        NP      IN        NP
                                │     ┌────┴───┐   │         │
                              meets  DT       NN   on     Friday
                                      │        │
                                     the     board
```

4 operators

$$\frac{\text{shift}(W)}{\text{shiftcomplete}} \quad \Big| \quad \frac{\text{glc}(LHS \rightarrow Trigger\,Rest)}{\text{glccomplete}(LHS \rightarrow Trigger\,Rest)}$$

Stack machine

1: **procedure** GLC(Grammar,Goals,InputString)
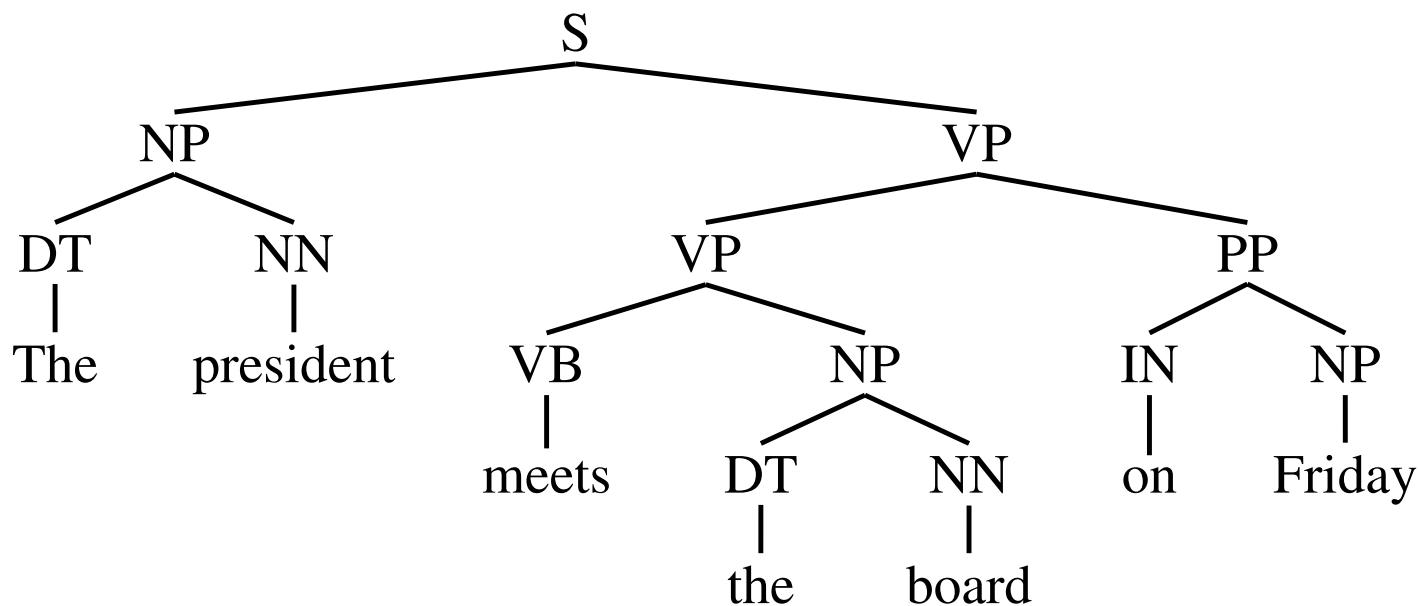2:   choose one of the viable GLCOptions
3:     **if** glc $(LHS \rightarrow Trigger\,Rest)$ **then** $\triangleright$ PREDICT without attaching
4:         GLC with $Trigger$ replaced by $Rest$ on Goals
5:     **else if** glcomplete $(LHS \rightarrow Trigger\,Rest)$ **then** $\triangleright$ PREDICT and attach
6:         GLC with $Trigger$ and matching LHS replaced by $Rest$ on Goals
7:     **else if** shift (W) **then**
8:         GLC with W pushed on Goals, popped from InputString
9:     **else if** shiftcomplete **then**
10:         GLC with matching W popped on Goals and popped from InputString
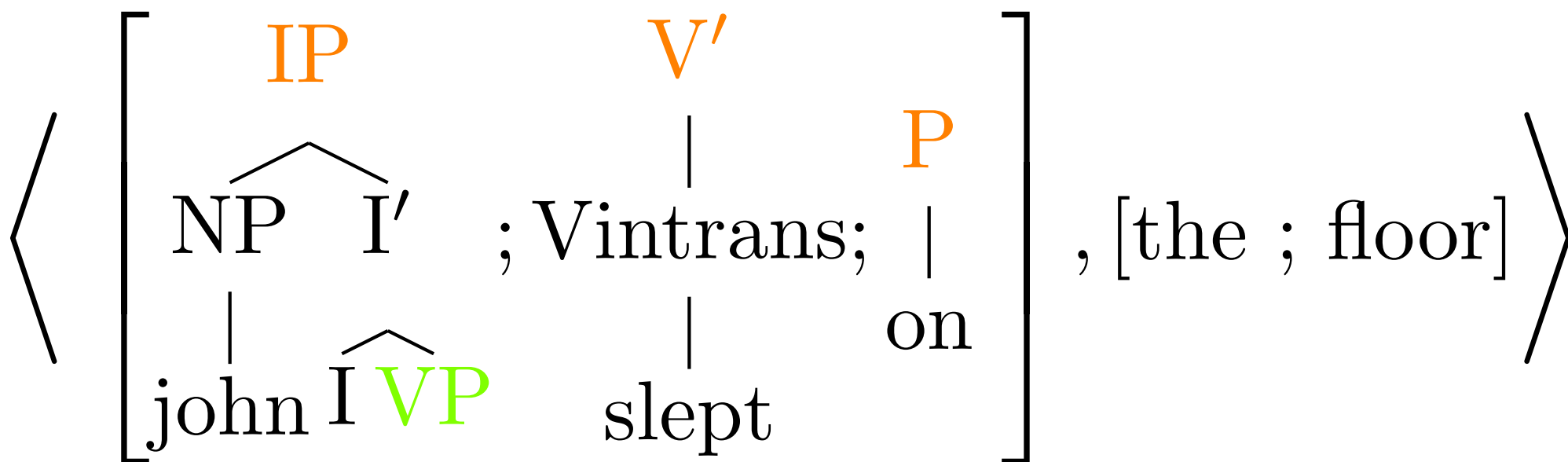11:   **end if**
12: **end procedure**

Left Corner

Shift The
| [S] The
Glc DT -> The><
| [S] DT
Glc NP -> DT><NN
| [S] NP [NN]
Shift president
| [S] NP [NN] president
GlcComplete NN -> president><
| [S] NP
GlcComplete S -> NP><VP
| [VP]
Shift meets
| [VP] meets
Glc VB -> meets><
| [VP] VB
Glc VP -> VB><NP
| [VP] VP [NP]
Shift the
| [VP] VP [NP] the
Glc DT -> the><
| [VP] VP [NP] DT
GlcComplete NP -> DT><NN
| [VP] VP [NN]
Shift board
| [VP] VP [NN] board
GlcComplete NN -> board><
| [VP] VP
GlcComplete VP -> VP><PP
| [PP]
Shift on
| [PP] on
Glc IN -> on><
| [PP] IN
GlcComplete PP -> IN><NP
| [NP]
Shift Friday
| [NP] Friday
GlcComplete NP -> Friday><
|

Left Corner

Shift The
| [S] The
Glc DT -> The><
| [S] DT
Glc NP -> DT><NN
| [S] NP [NN]
Shift president
| [S] NP [NN] president
GlcComplete NN -> president><
| [S] NP
GlcComplete S -> NP><VP
| [VP]
Shift meets
| [VP] meets
Glc VB -> meets><
| [VP] VB
Glc VP -> VB><NP
| [VP] VP [NP]
Shift the
| [VP] VP [NP] the
Glc DT -> the><
| [VP] VP [NP] DT
GlcComplete NP -> DT><NN
| [VP] VP [NN]
Shift board
| [VP] VP [NN] board
GlcComplete NN -> board><
| [VP] VP
GlcComplete VP -> VP><PP
| [PP]
Shift on
| [PP] on
Glc IN -> on><
| [PP] IN
GlcComplete PP -> IN><NP
| [NP]
Shift Friday
| [NP] Friday
GlcComplete NP -> Friday><
|

Expectations

$$\left\langle \begin{bmatrix} \text{IP} & & \text{V}' & & \\ \quad \overset{\frown}{\quad} & & \mid & & \text{P} \\ \text{NP} \quad \text{I}' & ; & \text{Vintrans}; & \mid & \\ \mid & & \mid & \text{on} \\ \text{john} \quad \overset{\frown}{\text{I} \,\, \text{VP}} & & \text{slept} & \end{bmatrix}, [\text{the ; floor}] \right\rangle$$

NP/Z garden path

(1)   a. while Mary was mending   a sock fell on the floor

            b. while Mary was mending , a sock fell on the floor
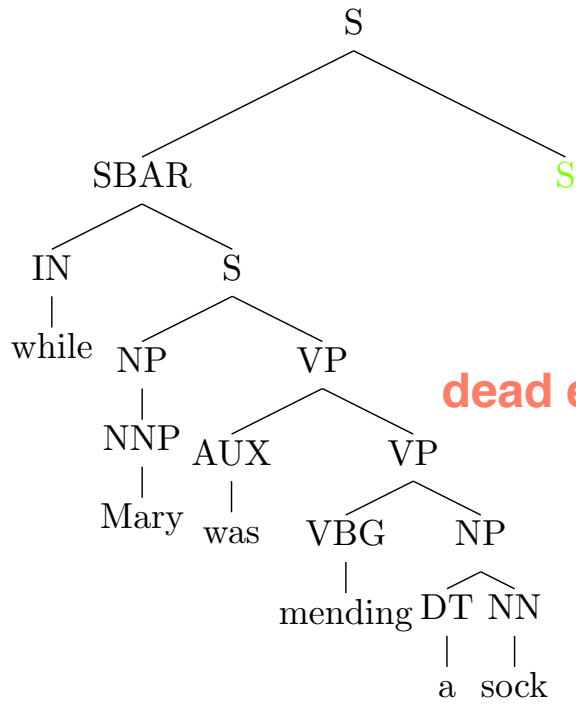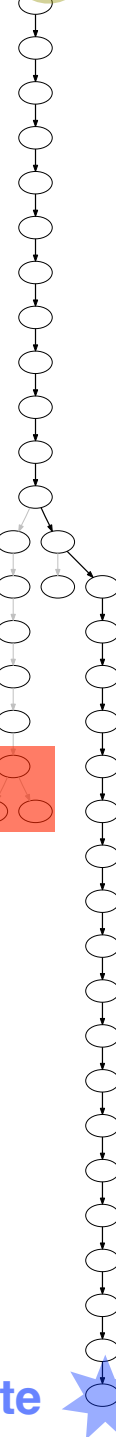
Phrase-structure grammar

S	→	NP	VP	
S	→	SBAR	S	
S	→	SBAR	,	S
SBAR	→	IN	S	
VP	→	VBG		
VP	→	VBG	NP	
VP	→	VBD	PP	
VP	→	AUX	VP	
VBG	→	mending		
VBD	→	fell		
VBD	→	was		
PP	→	IN	NP	
NP	→	NNP		
NP	→	DT	NN	
NNP	→	Mary		
NN	→	sock		
NN	→	floor		
IN	→	while		
IN	→	on		
DT	→	the		
DT	→	a		

Visited states

initial state

S
├── SBAR
│   ├── IN
│   │   └── while
│   └── S
│   ├── NP
│   │   └── NNP
│   │   └── Mary
│   └── VP
│   ├── AUX
│   │   └── was
│   └── VP
│   ├── VBG
│   │   └── mending
│   └── NP
│   ├── DT
│   │   └── a
│   └── NN
│   └── sock
└── S

dead end

goal state

S
├── SBAR
│   ├── IN
│   │   └── while
│   └── S
│   ├── NP
│   │   └── NNP
│   │   └── Mary
│   └── VP
│   ├── AUX
│   │   └── was
│   └── VP
│   └── VBG
│   └── mending
└── S
    ├── NP
    │   ├── DT
    │   │   └── a
    │   └── NN
    │   └── sock
    └── VP
        ├── VBD
        │   └── fell
        └── PP
            ├── IN
            │   └── on
            └── NP
                ├── DT
                │   └── the
                └── NN
                    └── floor

Visitation sequence

S
├── SBAR
│   ├── IN
│   │   └── while
│   └── S
│   ├── NP
│   │   └── NNP
│   │   └── Mary
│   └── VP
│   ├── AUX
│   │   └── was
│   └── VP
│   └── VBG
│   └── mending
└── S
    ├── ,
    └── NP
        ├── DT
        │   └── a
        └── NN,   sock

$g(n)$	$\hat{h}(n)$	$f(n)$
18	71	89

- - - - garden path
———— globally-correct

dead end
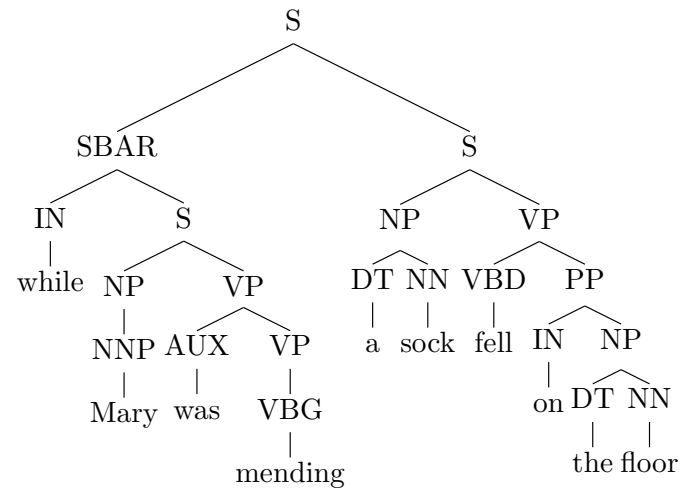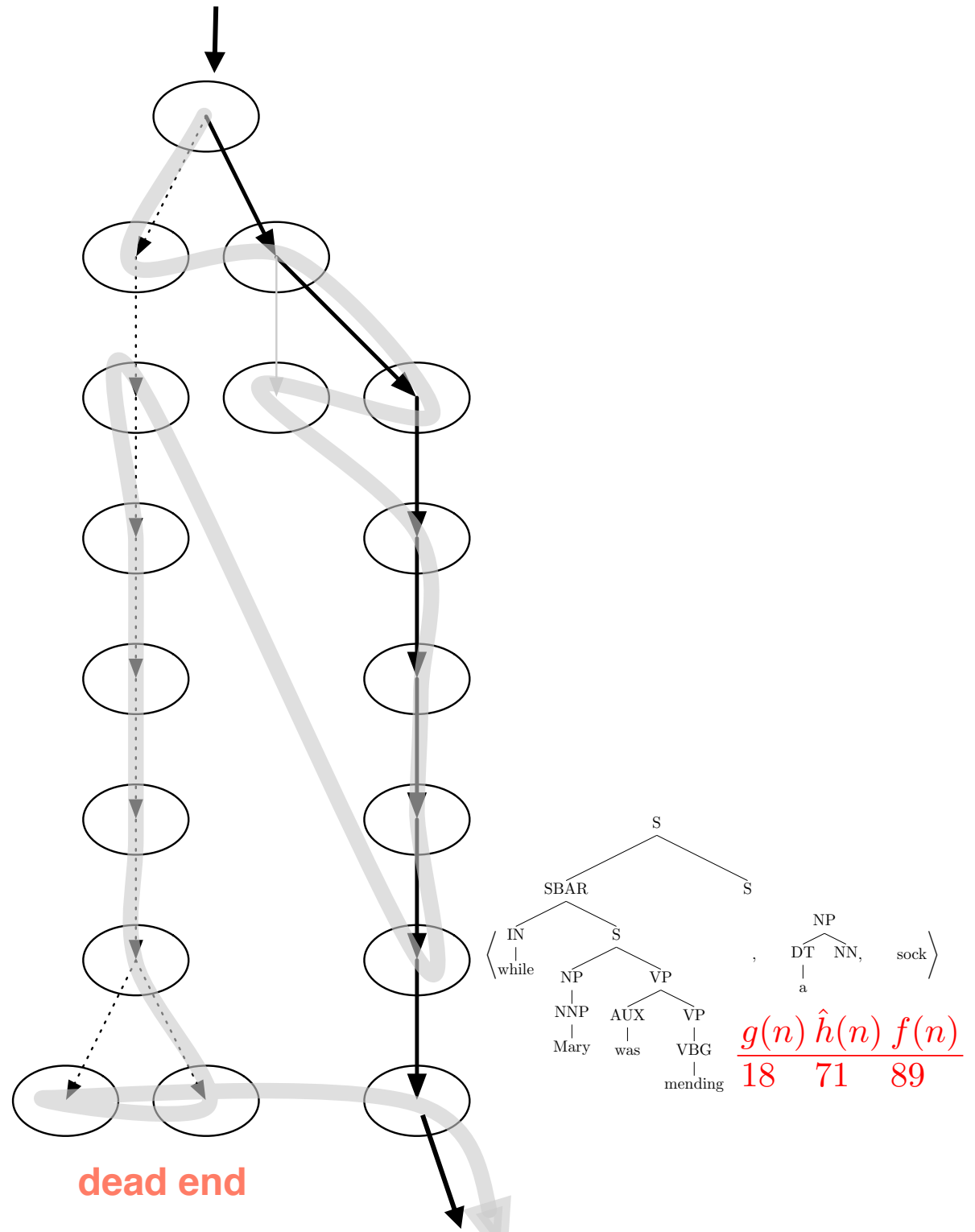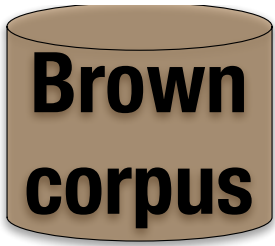
Informed search: A*

$$f(n) = g(n) + \hat{h}(n)$$

KNOWN COST TO GET HERE FROM START

A GUESS ABOUT HOW MUCH IT'LL COST TO FINISH FROM HERE

ESTIMATED VALUE OF A SEARCH NODE

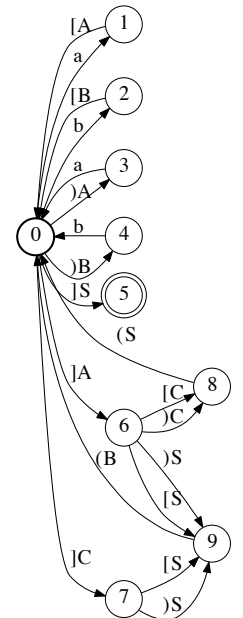| stack contents	attestations	average # steps to goal	standard error
VP S	55790	44.936	0.1572
S	53991	10.542	0.0986
NP S	43635	33.092	0.1633
NP	38844	55.791	0.2126
NP S S	34415	47.132	0.2122
S S	33578	52.800	0.2195
PP S	30693	34.454	0.1915
IN PP S	27272	32.379	0.2031
DT NP S	22375	34.478	0.2306
AUX VP S	16447	46.536	0.2863
VBD VP S	16224	43.057	0.2826
VB VP S	13548	40.404	0.3074
the NP S	12507	34.120	0.3046
NP NP S	12092	43.821	0.3269
DT	10440	66.452	0.3907

Brown corpus

$\hat{h}(n)$

Soar opportunities

parsing strategies
are pushdown transducers
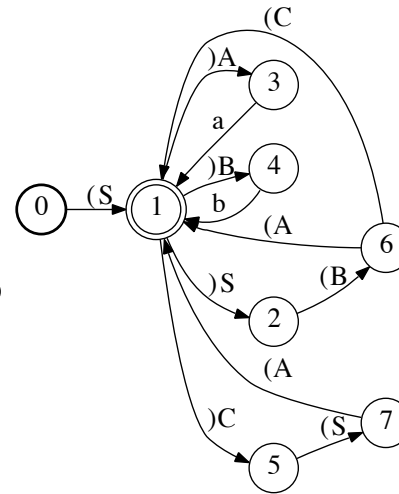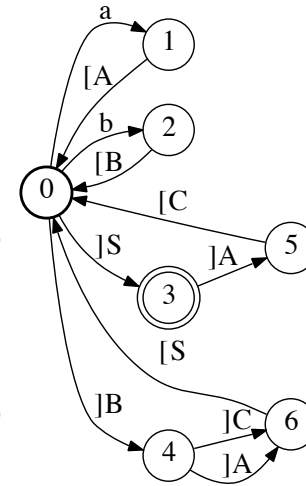
(a) left parser   (b) right parser   (c) left corner parser

idea: a `^waypoint` has an `^input` queue and `^stack`
PDT transitions as operators
tune numeric prefs with RL
chunk operator sequences

Read the paper

COGNITIVE SCIENCE
A Multidisciplinary Journal

What a Rational Parser Would Do

John T. Hale

Department of Linguistics, Cornell University

Talk to me!

jthale@cornell.edu