CS626: Speech, NLP and the Web

Softmax FFNN-BP and Neural Dependency Parsing

Pushpak Bhattacharyya

Computer Science and Engineering Department

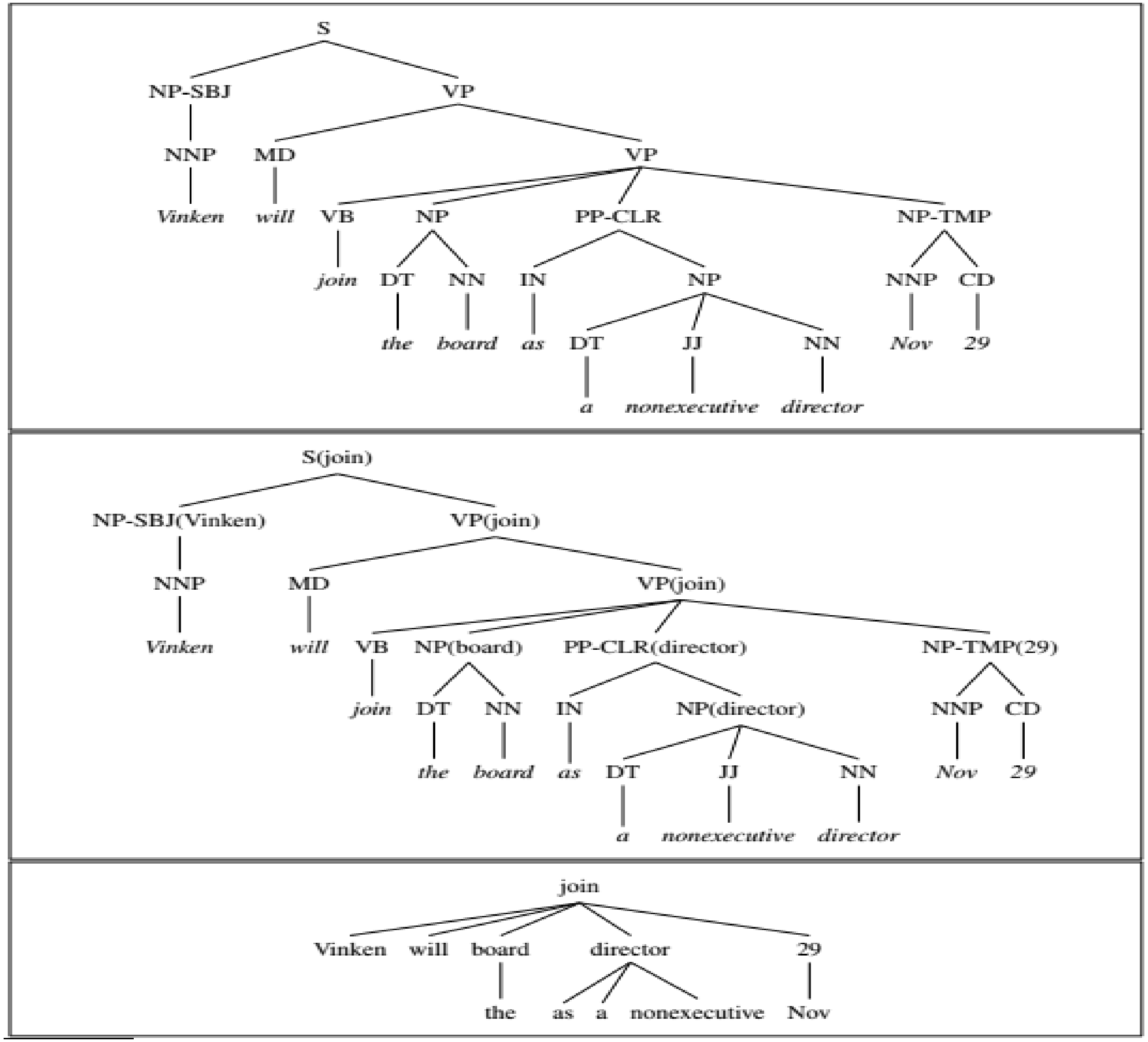
IIT Bombay

Week of 26th October, 2020

Dependency parsing algorithms

Makin approaches

Rule based approach: CP to DP



Speech and Language Processing, Jurafksy & Martin, Ch-15, 2019

Head directionality

- Head final languages
  - Head final structure
  - Head of a phrase follows its complements
  - Example: Hindi is strong head final language
- Head initial languages
  - Head initial structure
  - Head of a phrase precedes its complements
  - Example: English is strong head-initial language
- Head
  - Component of a phrase which determines the category of the phrase
  - Example: Head of a verb phrase is verb

Rule based approach: CP to DP

- Two main acts
  - Identify all the head-dependent relations
  - Identifying the correct dependency relations for above relations
- Task involves
  - Marking the head
  - Make the head of each non-head depend on the head

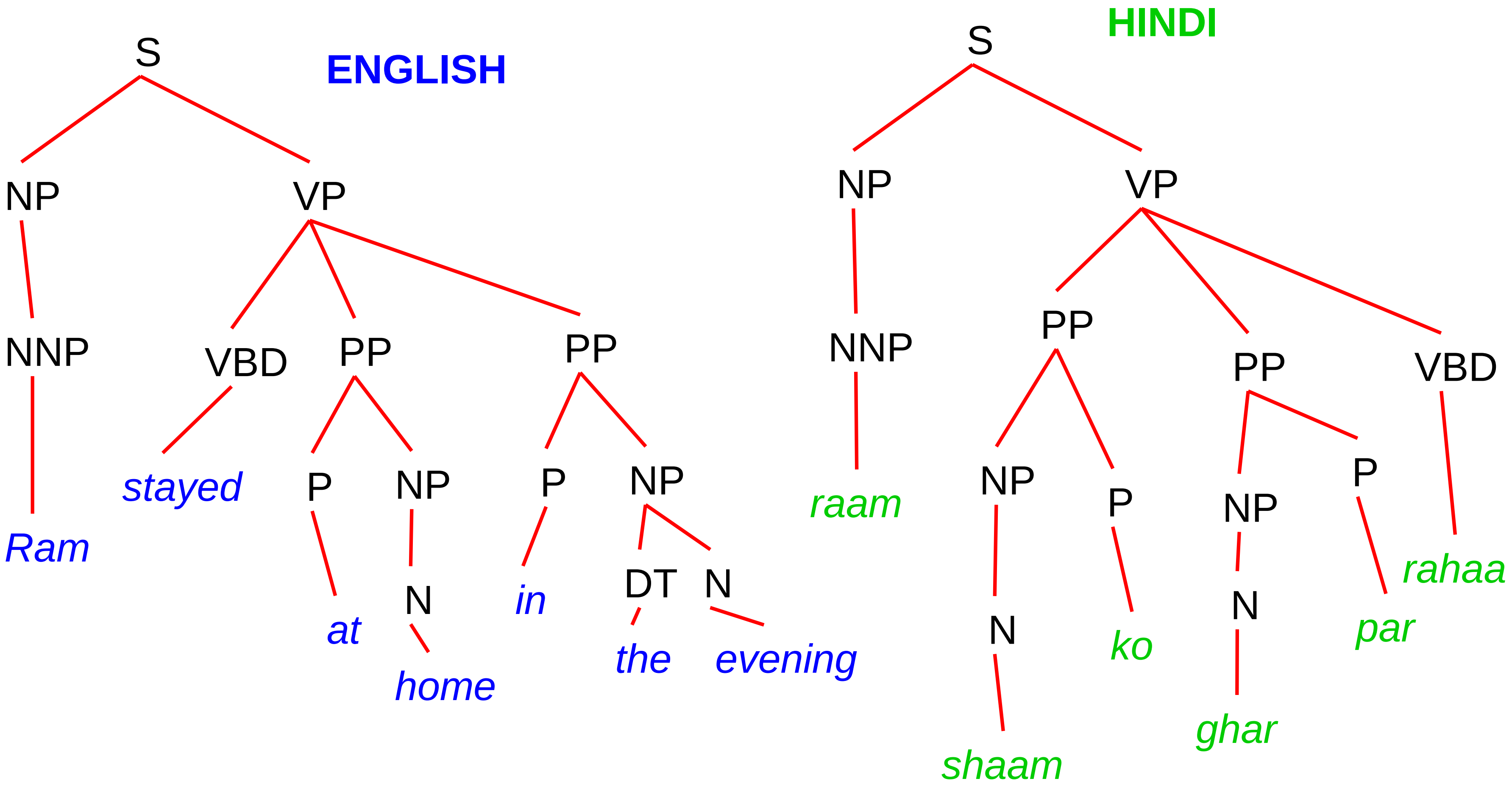
Jurafsky, Dan. *Speech & language processing*. Pearson Education India, 2000.

Example

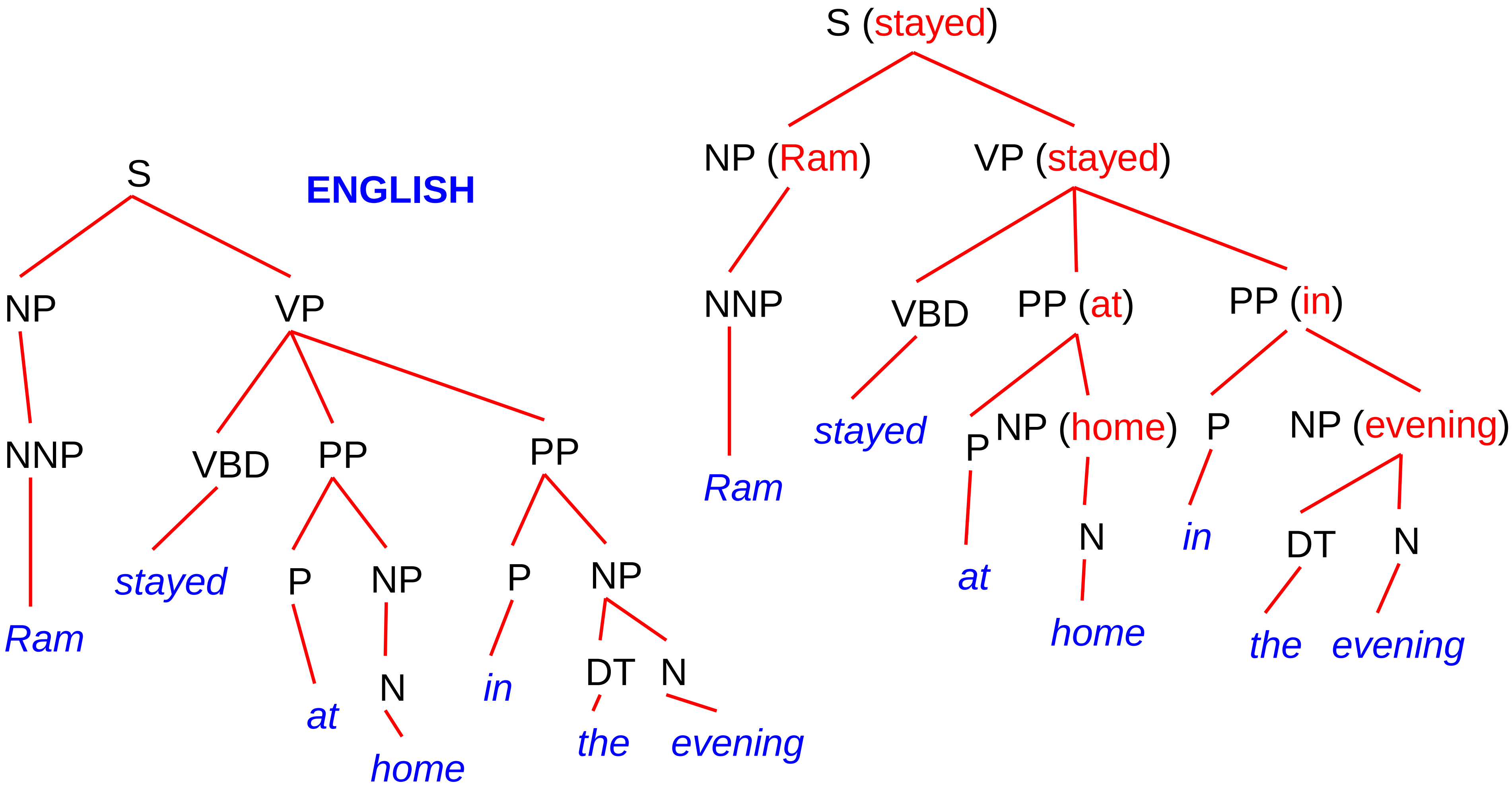
- *Ram stayed at home in the evening*

- *raam shaam ko ghar par rahaa*

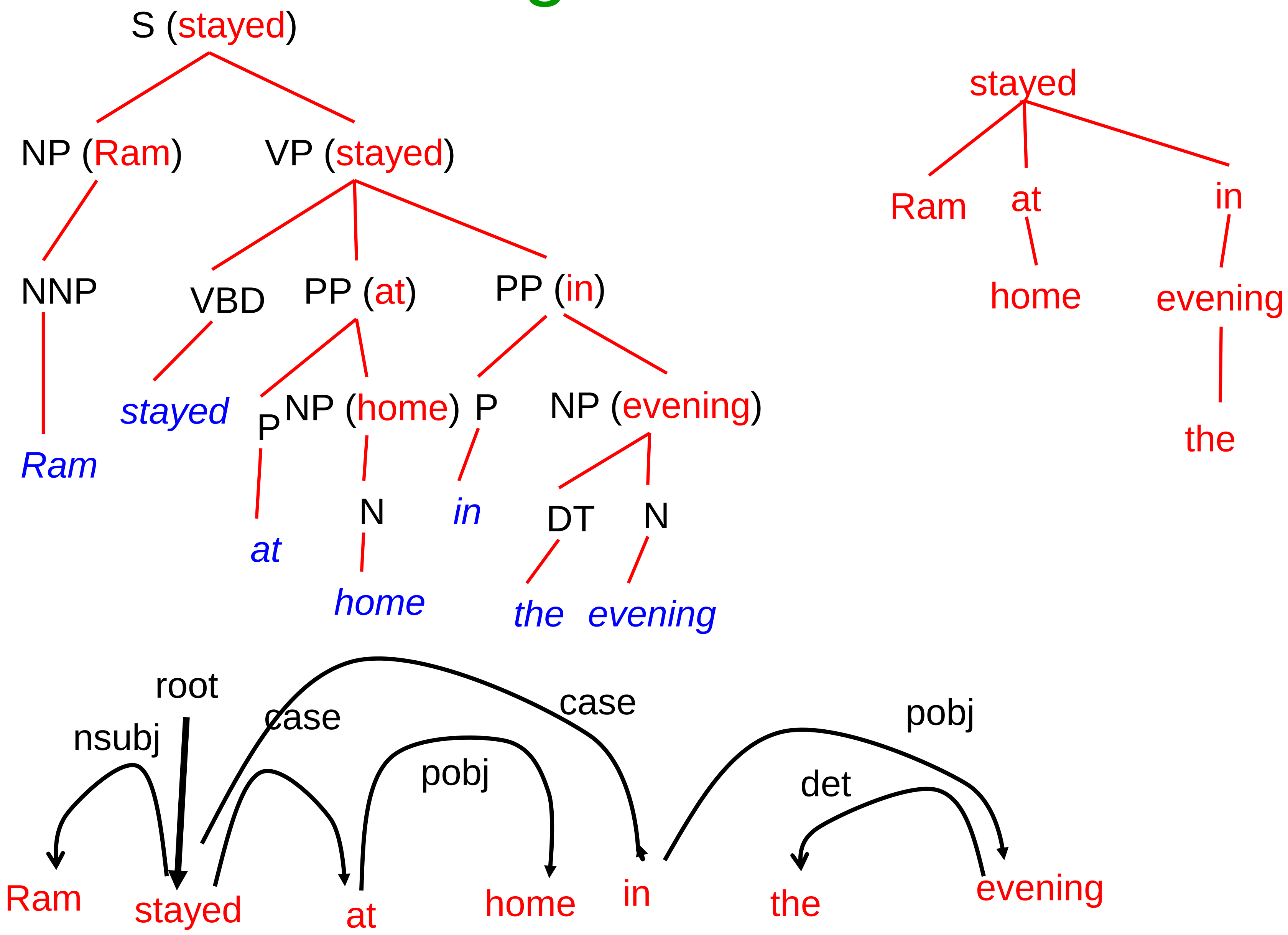
English-Hindi: Head Initial-Head Final



English CP → DP



English CP → DP cntd.



S (stayed)

NP (Ram) VP (stayed)

NNP VBD PP (at) PP (in)

stayed P NP (home) P NP (evening)

Ram N *in* DT N

at *home* *the* *evening*

stayed

Ram at in

home evening

the

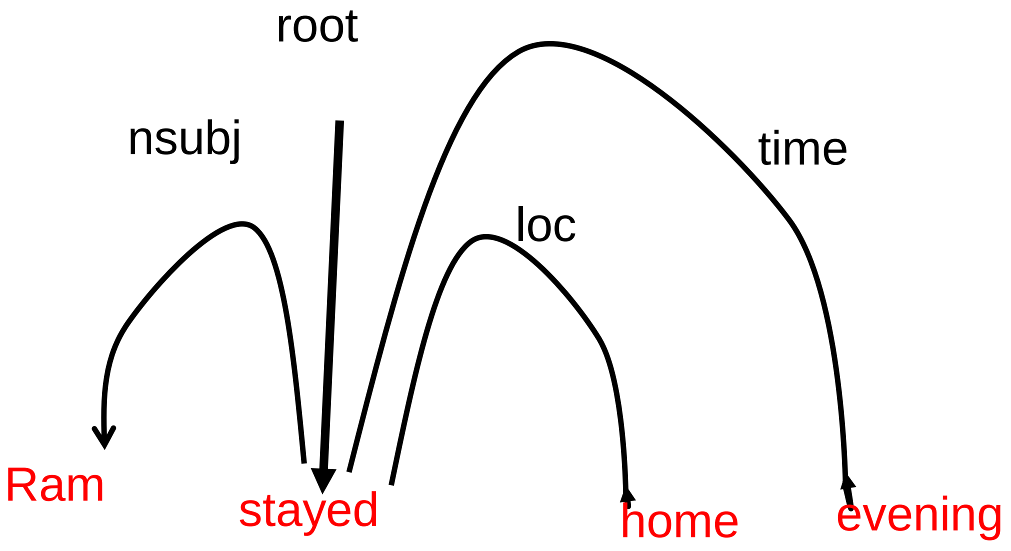
root

nsubj case case pobj

pobj det

Ram stayed at home in the evening

English Compact DP with labels



root

nsubj

time

loc

Ram stayed home evening

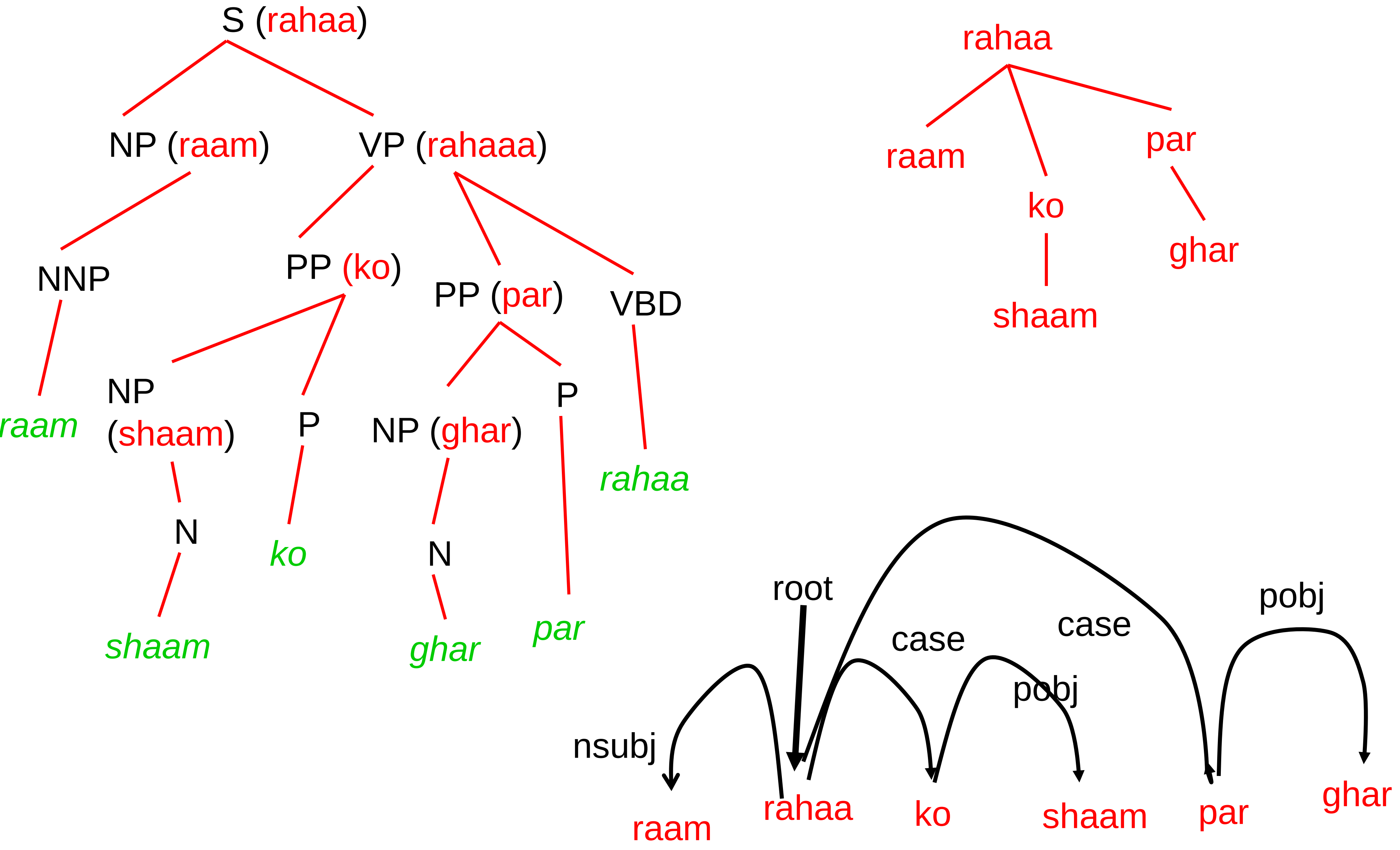
Relevant discussion: different types of subjects (nsubj, dsubj, gsubj)

- *nsubj*: Nominative subject
- Languages do have non-nominal subject
  - Dative subject is very common in Indian languages
  - Example
    - Hindi: मुझे आम चाहिए
    - English: I want Mango.
    - मुझे: Dative case (सम्प्रदान कारक) (dsubj: dative subject)
    - I: Nominative case (nsubj)

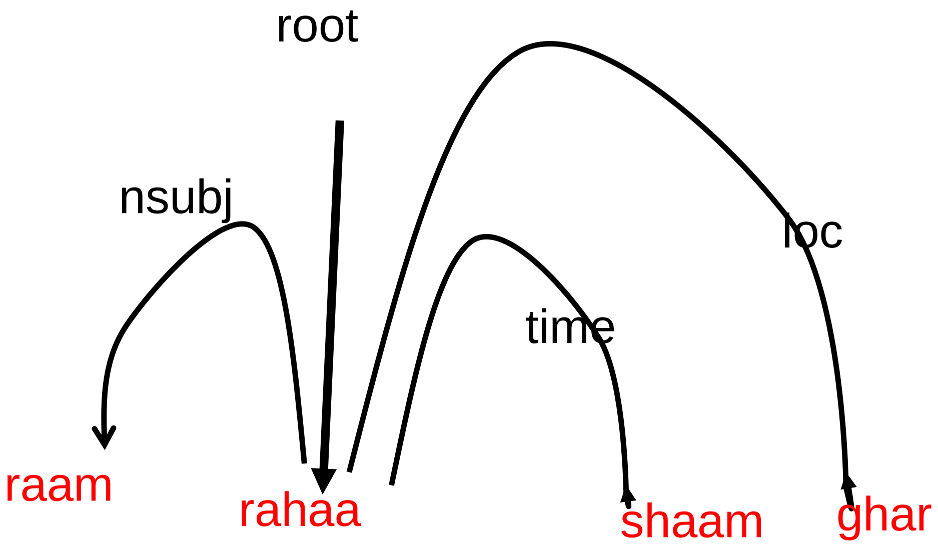
Deeper representation

- Deeper representation towards meaning
  - *nsubj* is shallower than *agent*
  - *nsubj* will lead to *agent*
  - Semantic role labeled graph
    - Compact DP with labels
    - The relations captures the exact semantics of the situation
    - Case relationships: l*oc*, *time*, *agent*
      - Indicate semantic relationships of the noun with the main verb

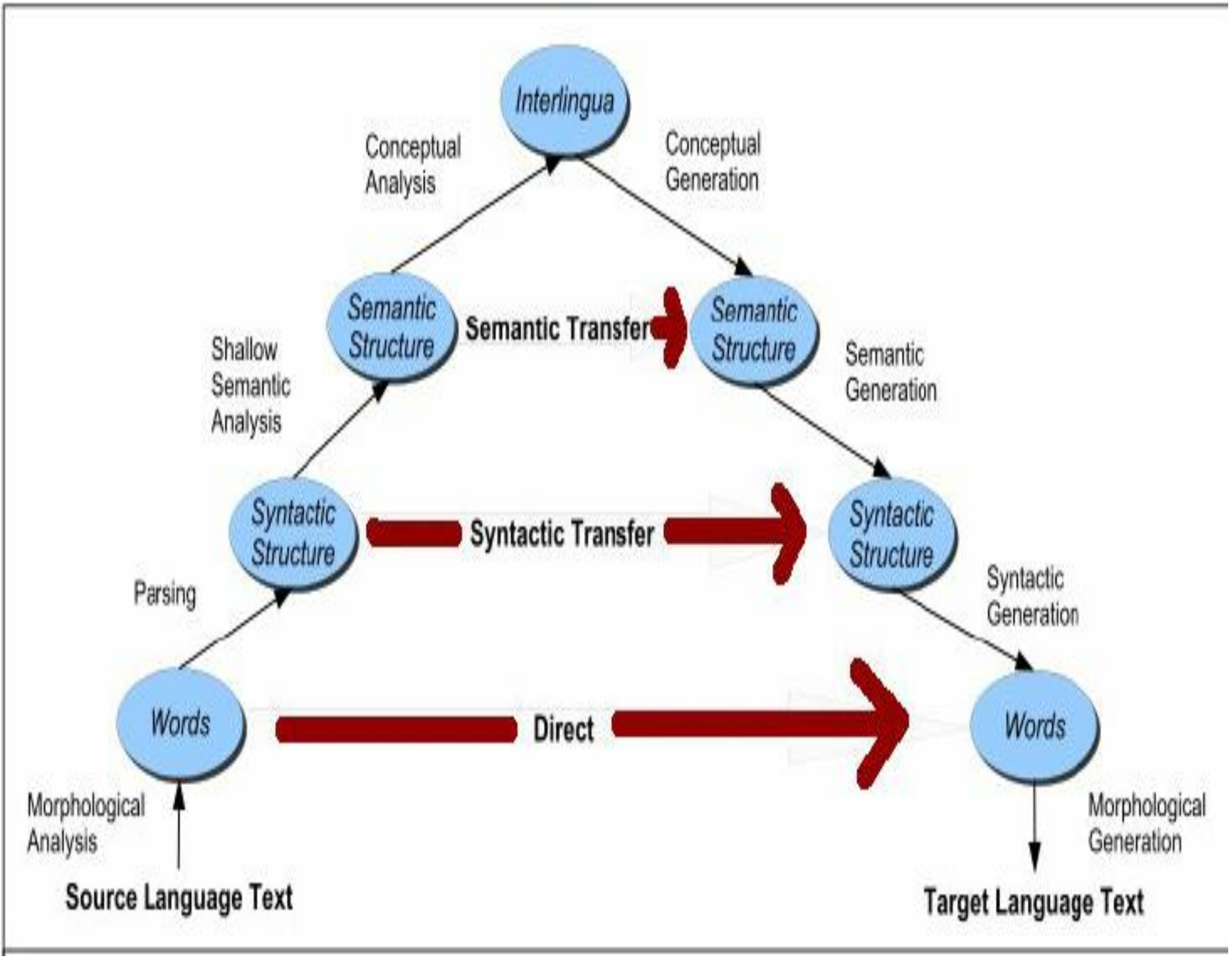
Hindi CP→DP


S (rahaa)
NP (raam)
VP (rahaaa)
NNP
PP (ko)
PP (par)
VBD
raam
NP (shaam)
P
NP (ghar)
P
rahaa
N
ko
N
shaam
ghar
par
rahaa
raam
ko
par
shaam
ghar
root
case
case
pobj
pobj
nsubj
raam
rahaa
ko
shaam
par
ghar

Hindi Compact DP with labels



A Digression: Vauquois Triangle in Machine Translation



The deeper the NLP analysis, the closer the representations and the easier the analysis

Grammatical head and Semantic Head

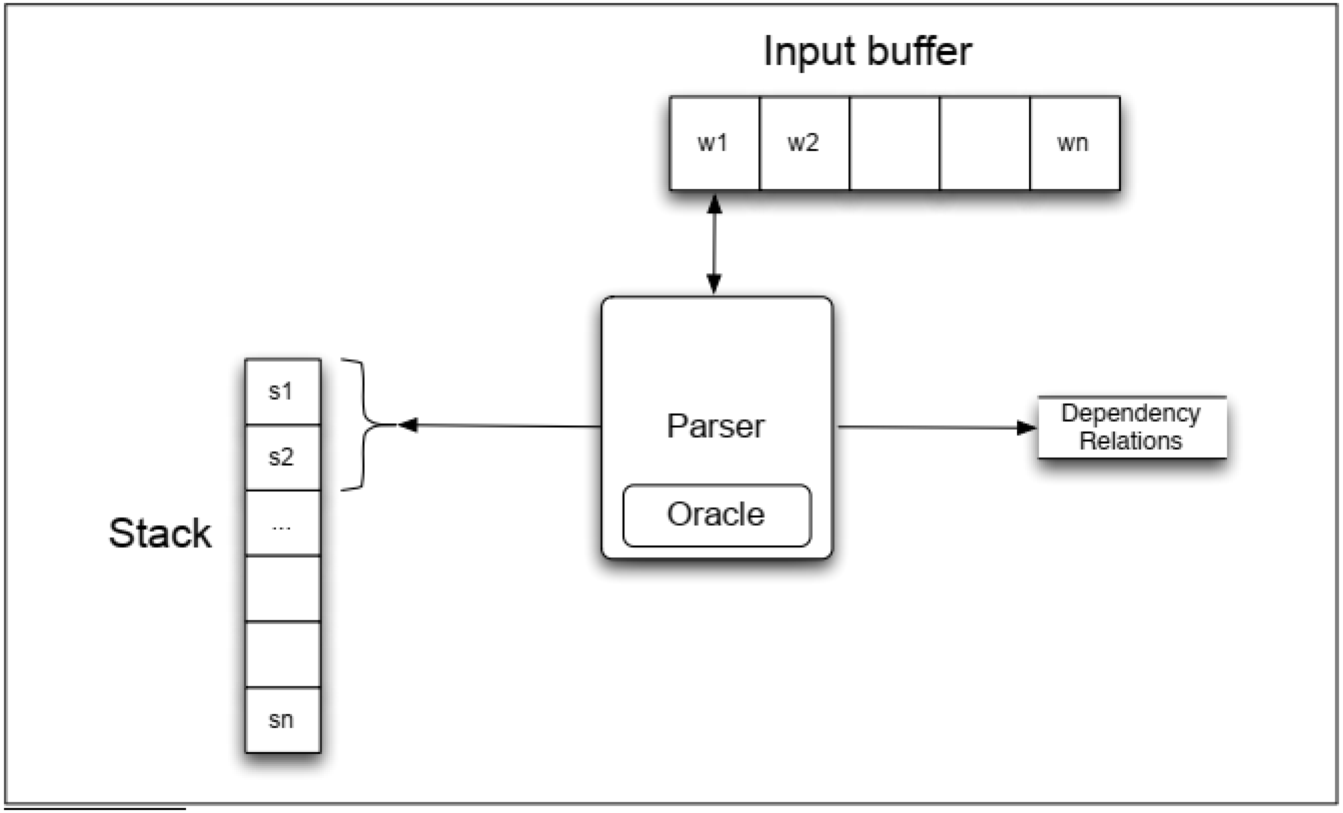
- Grammatical Head: Word whose POS gives the Phrase its name

- Semantic Head: Word that carries the “semantic load”

Data Driven: two approaches

- Transition-based
  - State machine for mapping a sentence to its dependency graph
  - Inducing a model for predicting the next transition, given the current state and the transition history so far.
- Graph-based
  - Induce a model for assigning scores to the candidate dependency graphs for a sentence
  - Find the maximum-scoring dependency Tree
  - Maximum spanning tree (MST) parsing

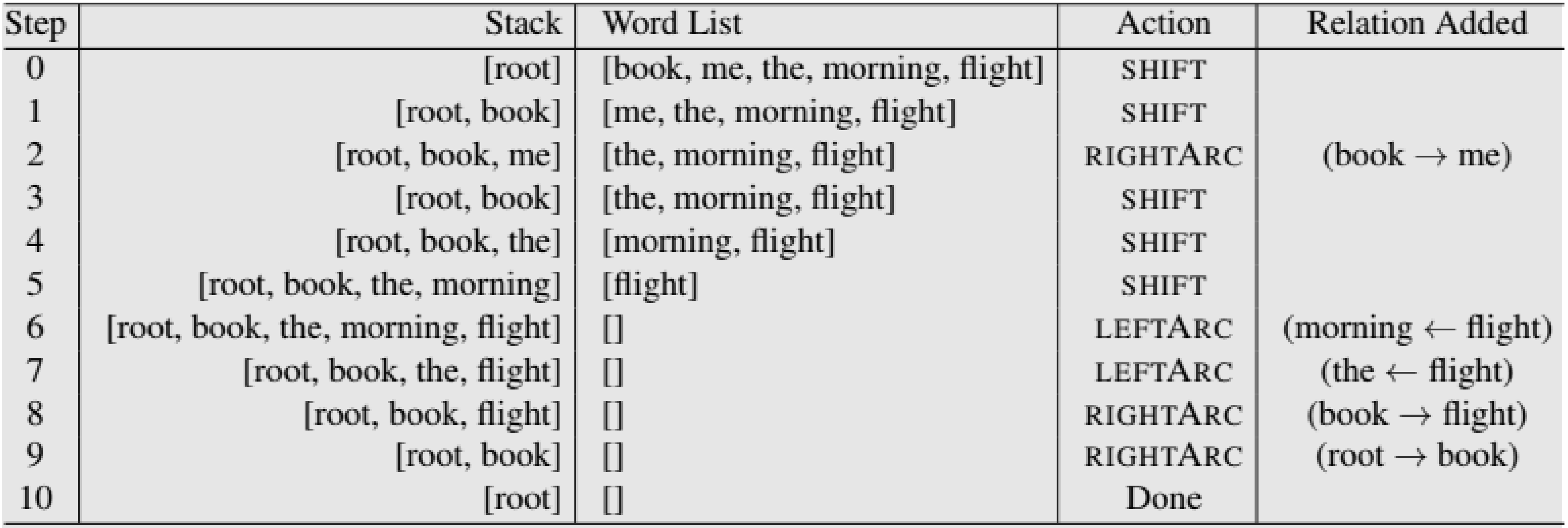
Basic Transition Based DP



Examines top two elements of the stack and selects an action based on consulting an oracle that examines the current configuration.

Speech and Language Processing, Jurafksy & Martin, Ch-15, 2019

Example: transition based



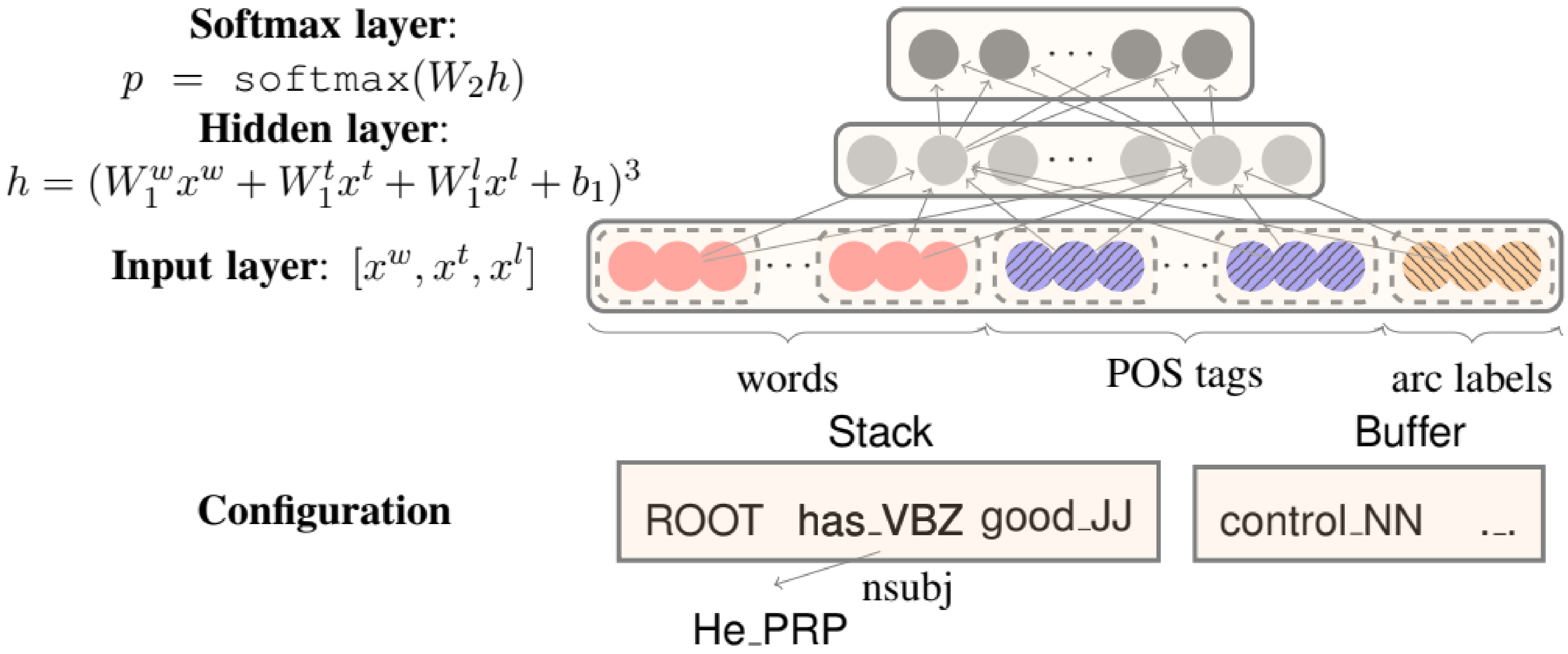
Step	Stack	Word List	Action	Relation Added
0	[root]	[book, me, the, morning, flight]	SHIFT	
1	[root, book]	[me, the, morning, flight]	SHIFT	
2	[root, book, me]	[the, morning, flight]	RIGHTARC	(book → me)
3	[root, book]	[the, morning, flight]	SHIFT	
4	[root, book, the]	[morning, flight]	SHIFT	
5	[root, book, the, morning]	[flight]	SHIFT	
6	[root, book, the, morning, flight]	[]	LEFTARC	(morning ← flight)
7	[root, book, the, flight]	[]	LEFTARC	(the ← flight)
8	[root, book, flight]	[]	RIGHTARC	(book → flight)
9	[root, book]	[]	RIGHTARC	(root → book)
10	[root]	[]	Done	

Trace of a transition-based parse

Speech and Language Processing, Jurafksy & Martin, Ch-15, 2019

A neural transition based parser

(chen and Manning 2014)



Softmax layer:

$$p = \mathtt{softmax}(W_2 h)$$

Hidden layer:

$$h = (W_1^w x^w + W_1^t x^t + W_1^l x^l + b_1)^3$$

Input layer: $[x^w, x^t, x^l]$

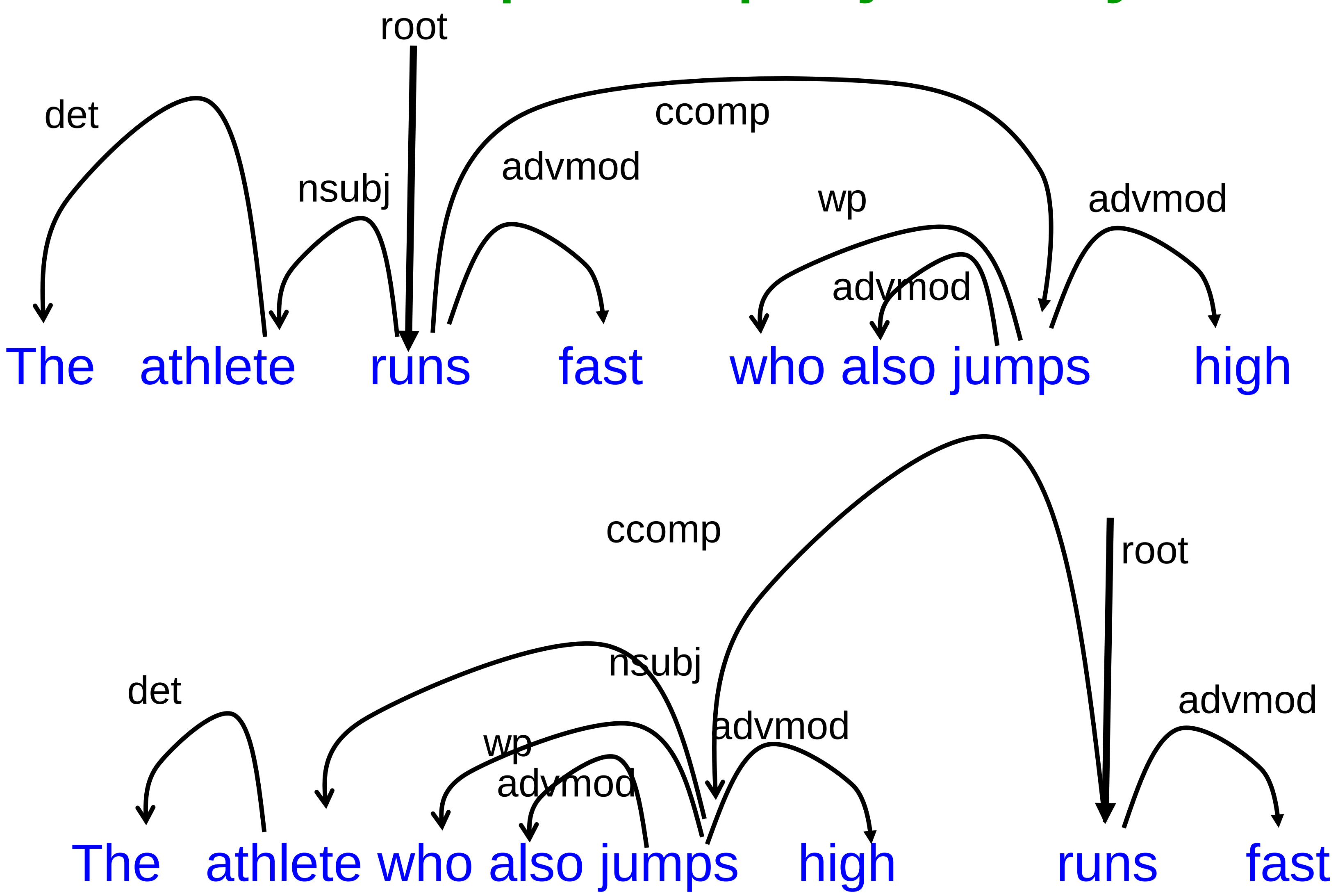
Configuration

Notion of Projectivity

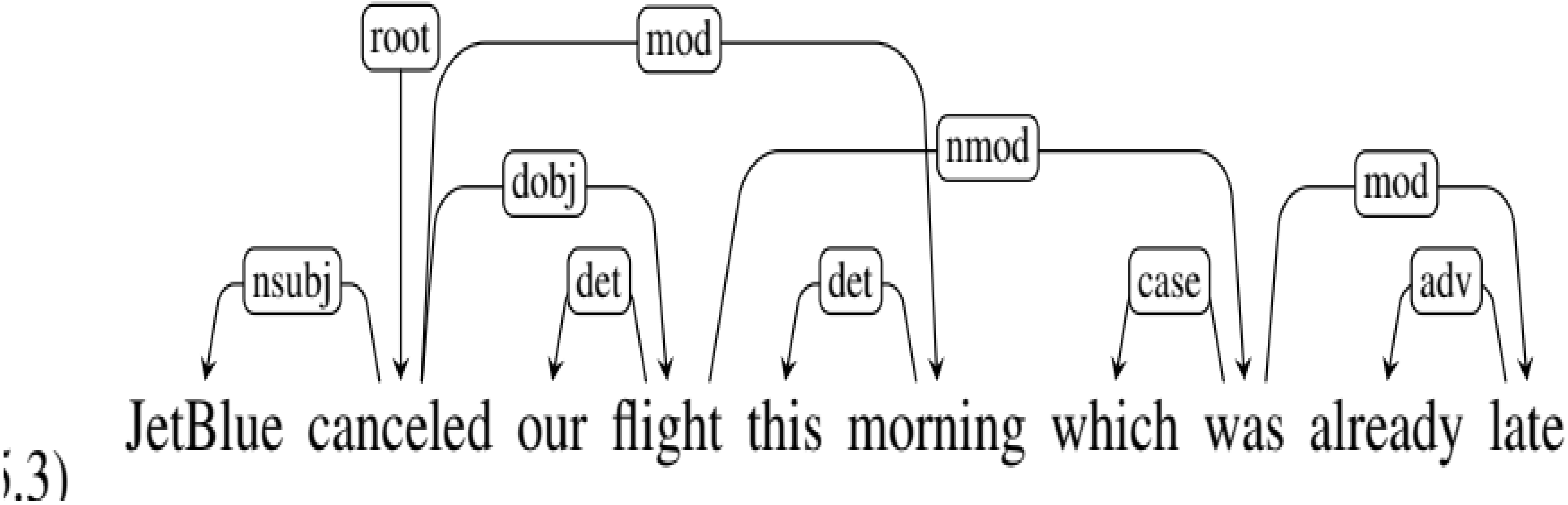
Definition

- An arc from a head to a dependent is said to be projective if there is a path from the head to every word that lies between the head and the dependent in the sentence
- A dependency tree is then said to be projective if all the arcs that make it up are projective
- ***Intuition***- the dependency graph can be drawn on a plane w/o crossing of arcs

Example of projectivity


root
det
nsubj
advmod
ccomp
wp
advmod
advmod
The athlete runs fast who also jumps high
ccomp
root
nsubj
det
advmod
wp
advmod
advmod
The athlete who also jumps high runs fast

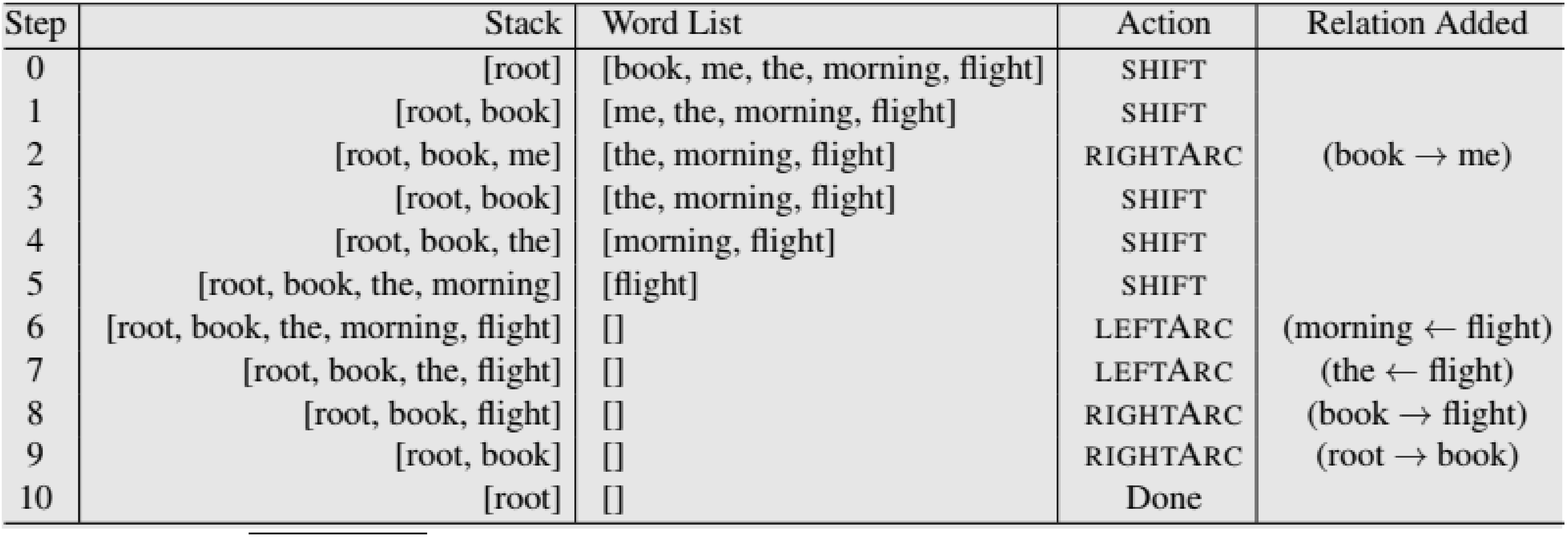
Another example


root
mod
nmod
dobj
mod
nsubj
det
det
case
adv
JetBlue canceled our flight this morning which was already late
;.3)

Learning of transitions

Speech and NLP, J & M, Ch 15, 2019.

Recall: transition based DP

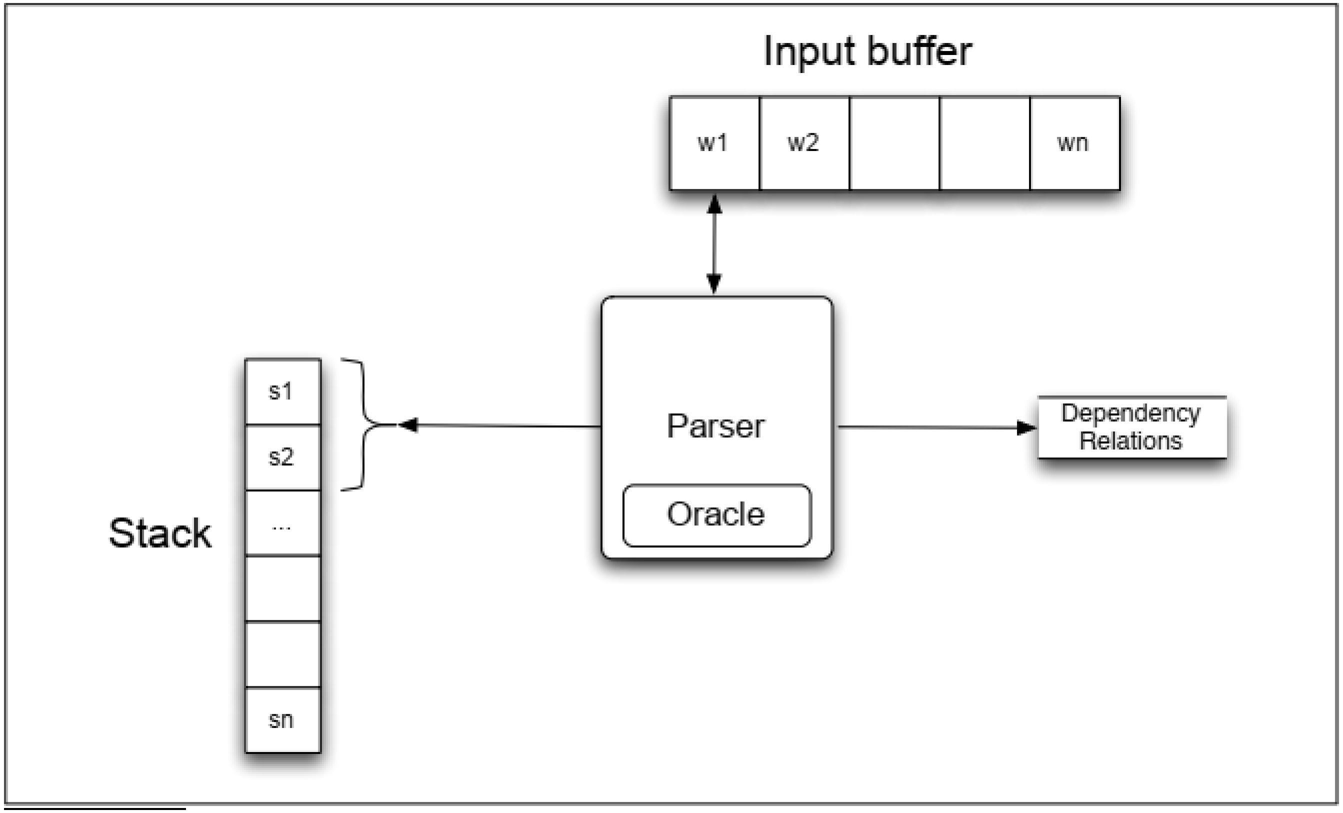


Step	Stack	Word List	Action	Relation Added
0	[root]	[book, me, the, morning, flight]	SHIFT	
1	[root, book]	[me, the, morning, flight]	SHIFT	
2	[root, book, me]	[the, morning, flight]	RIGHTARC	(book → me)
3	[root, book]	[the, morning, flight]	SHIFT	
4	[root, book, the]	[morning, flight]	SHIFT	
5	[root, book, the, morning]	[flight]	SHIFT	
6	[root, book, the, morning, flight]	[]	LEFTARC	(morning ← flight)
7	[root, book, the, flight]	[]	LEFTARC	(the ← flight)
8	[root, book, flight]	[]	RIGHTARC	(book → flight)
9	[root, book]	[]	RIGHTARC	(root → book)
10	[root]	[]	Done	

Trace of a transition-based parse

Speech and Language Processing, Jurafksy & Martin, Ch-15, 2019

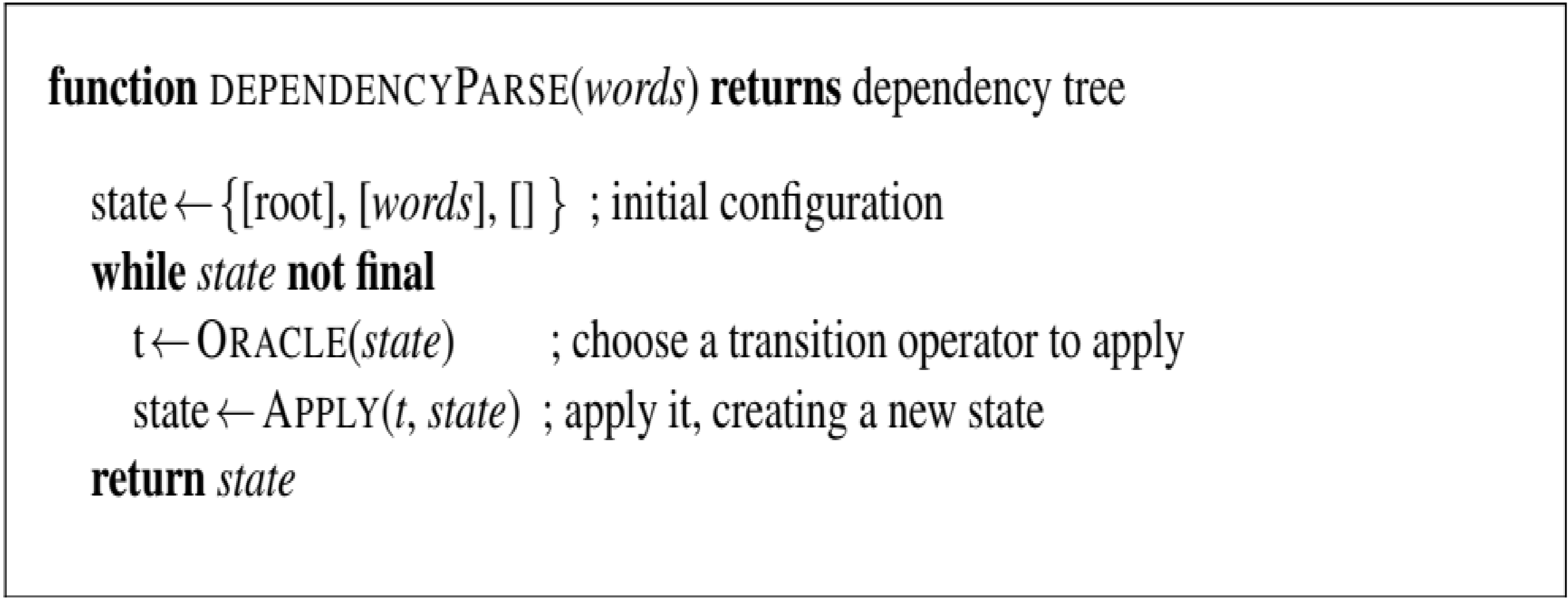
Basic Transition Based DP



Examines top two elements of the stack and selects an action based on consulting an oracle that examines the current configuration.

Speech and Language Processing, Jurafksy & Martin, Ch-15, 2019

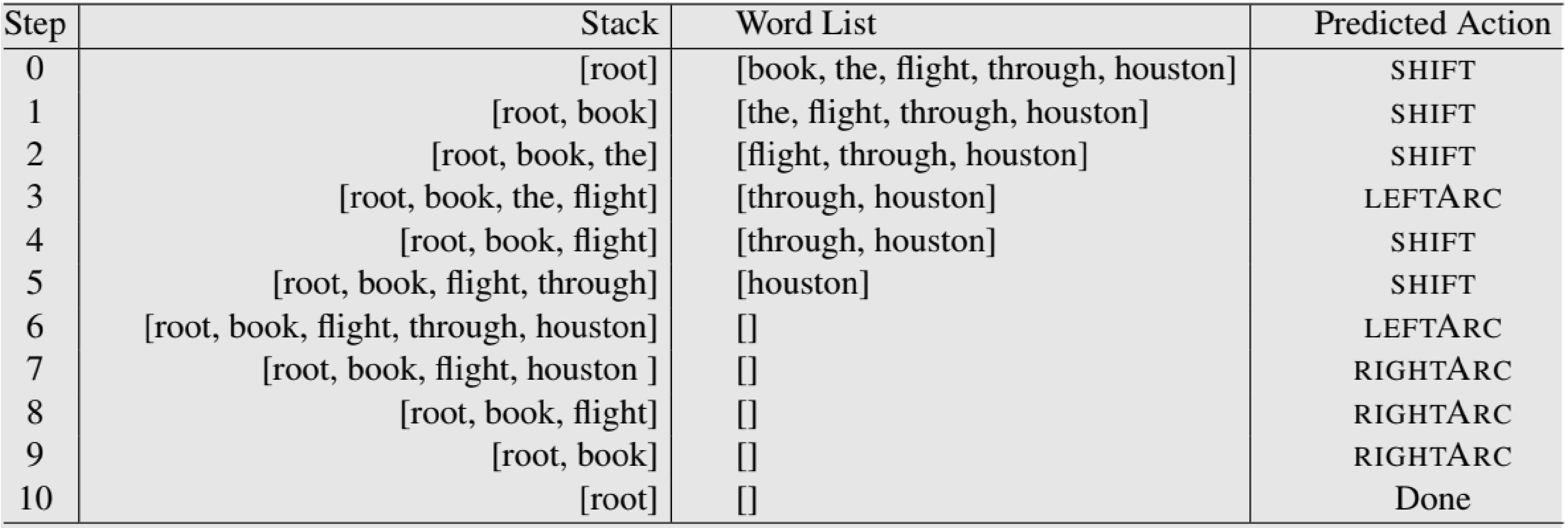
Operators: shift, leftarc, rightarc



```
function DEPENDENCYPARSE(words) returns dependency tree

  state ← {[root], [words], [] }  ; initial configuration
  while state not final
    t ← ORACLE(state)        ; choose a transition operator to apply
    state ← APPLY(t, state)  ; apply it, creating a new state
  return state
```

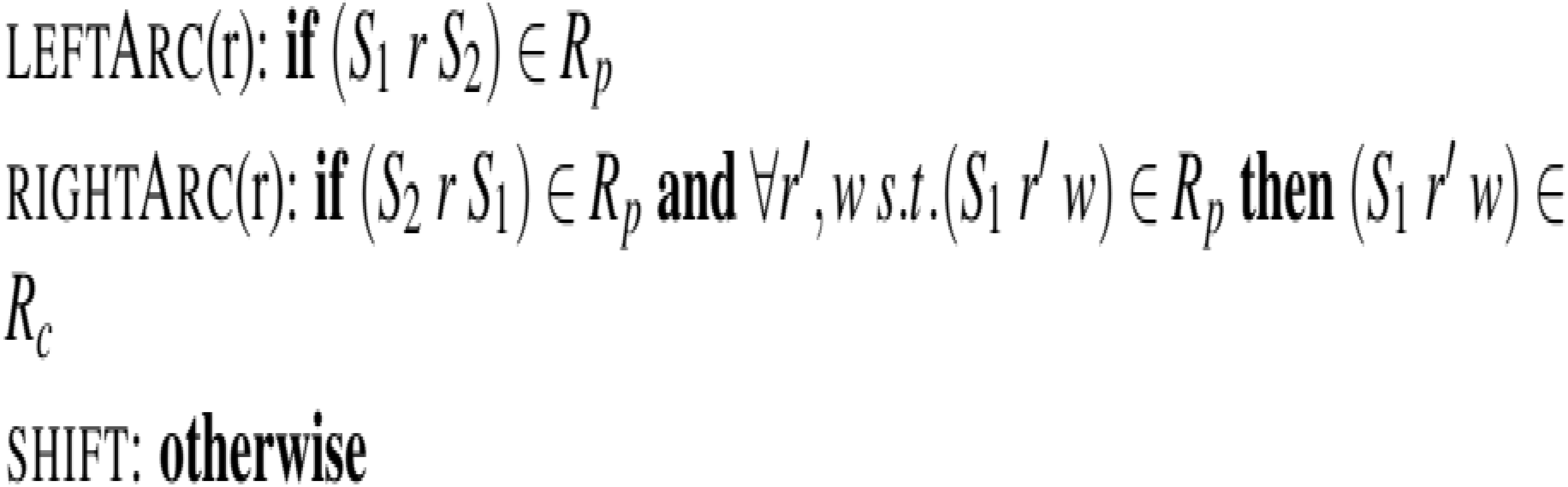
Generation of Training Data



Step	Stack	Word List	Predicted Action
0	[root]	[book, the, flight, through, houston]	SHIFT
1	[root, book]	[the, flight, through, houston]	SHIFT
2	[root, book, the]	[flight, through, houston]	SHIFT
3	[root, book, the, flight]	[through, houston]	LEFTARC
4	[root, book, flight]	[through, houston]	SHIFT
5	[root, book, flight, through]	[houston]	SHIFT
6	[root, book, flight, through, houston]	[]	LEFTARC
7	[root, book, flight, houston]	[]	RIGHTARC
8	[root, book, flight]	[]	RIGHTARC
9	[root, book]	[]	RIGHTARC
10	[root]	[]	Done

Training data

How are operators generated

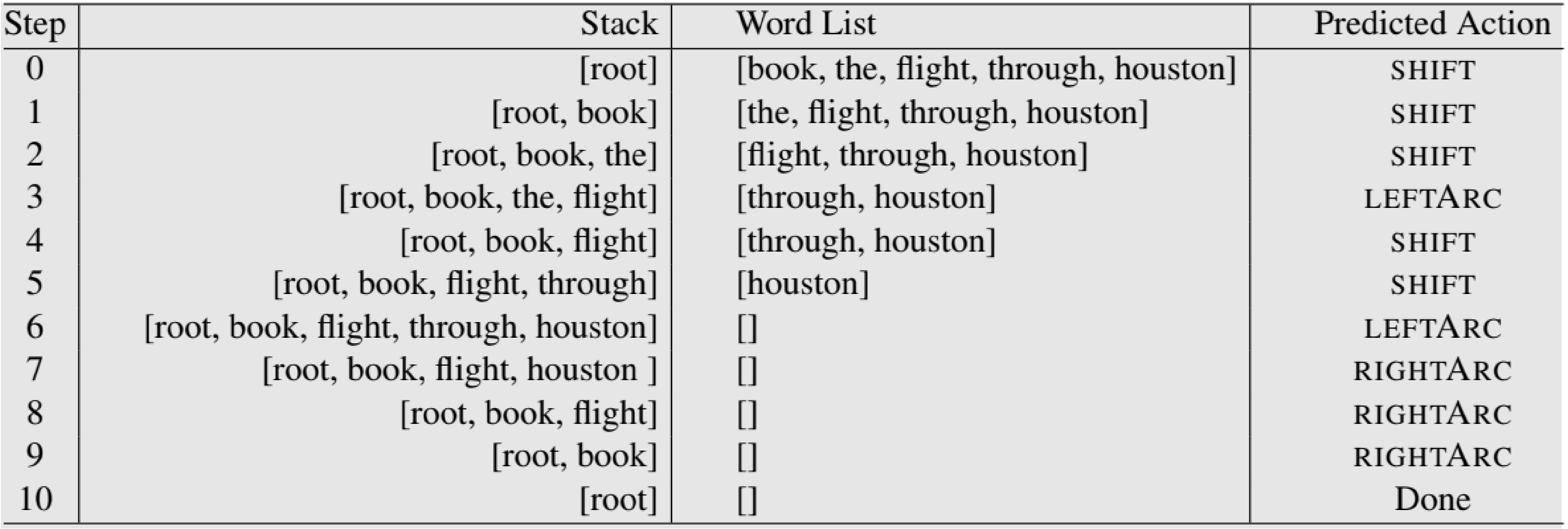


LEFTARC(r): **if** $(S_1\ r\ S_2) \in R_p$

RIGHTARC(r): **if** $(S_2\ r\ S_1) \in R_p$ **and** $\forall r', w\ s.t.(S_1\ r'\ w) \in R_p$ **then** $(S_1\ r'\ w) \in R_c$

SHIFT: **otherwise**

Generation of Training Data



Step	Stack	Word List	Predicted Action
0	[root]	[book, the, flight, through, houston]	SHIFT
1	[root, book]	[the, flight, through, houston]	SHIFT
2	[root, book, the]	[flight, through, houston]	SHIFT
3	[root, book, the, flight]	[through, houston]	LEFTARC
4	[root, book, flight]	[through, houston]	SHIFT
5	[root, book, flight, through]	[houston]	SHIFT
6	[root, book, flight, through, houston]	[]	LEFTARC
7	[root, book, flight, houston]	[]	RIGHTARC
8	[root, book, flight]	[]	RIGHTARC
9	[root, book]	[]	RIGHTARC
10	[root]	[]	Done

Training data

A neural transition based parser
(chen and Manning 2014)

Softmax layer:

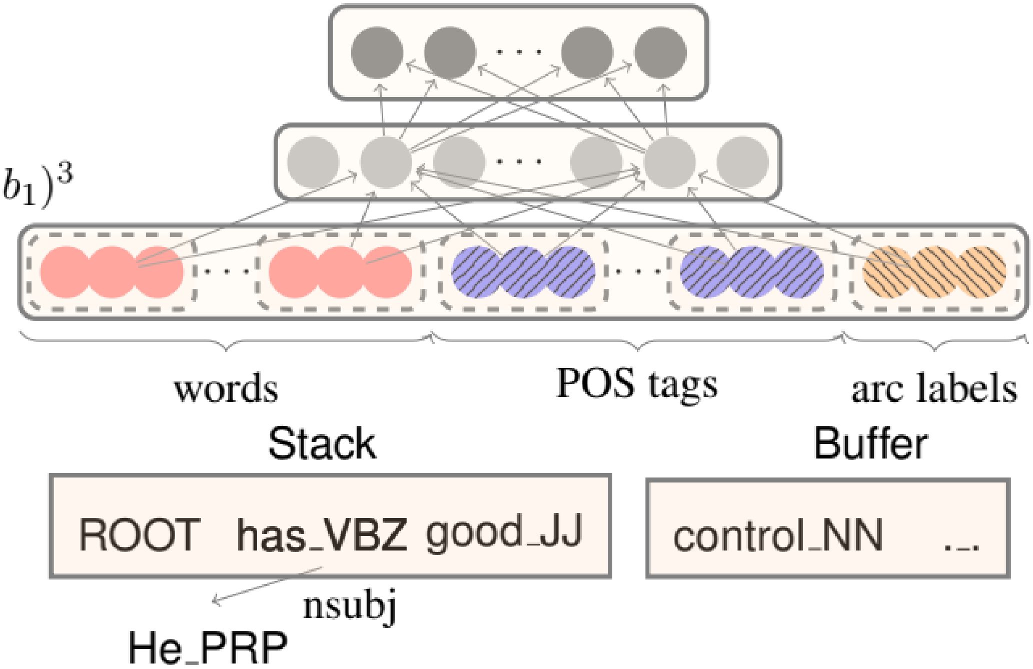
$$p = \texttt{softmax}(W_2 h)$$

Hidden layer:

$$h = (W_1^w x^w + W_1^t x^t + W_1^l x^l + b_1)^3$$

Input layer: $[x^w, x^t, x^l]$

Configuration



Features: example sentence “*cancelled flights to Houston*”

$$\langle s_1.w = \mathit{flights}, op = \mathit{shift} \rangle$$

$$\langle s_2.w = \mathit{canceled}, op = \mathit{shift} \rangle$$

$$\langle s_1.t = \mathit{NNS}, op = \mathit{shift} \rangle$$

$$\langle s_2.t = \mathit{VBD}, op = \mathit{shift} \rangle$$

$$\langle b_1.w = \mathit{to}, op = \mathit{shift} \rangle$$

$$\langle b_1.t = \mathit{TO}, op = \mathit{shift} \rangle$$

$$\langle s_1.wt = \mathit{flightsNNS}, op = \mathit{shift} \rangle$$

DP across languages

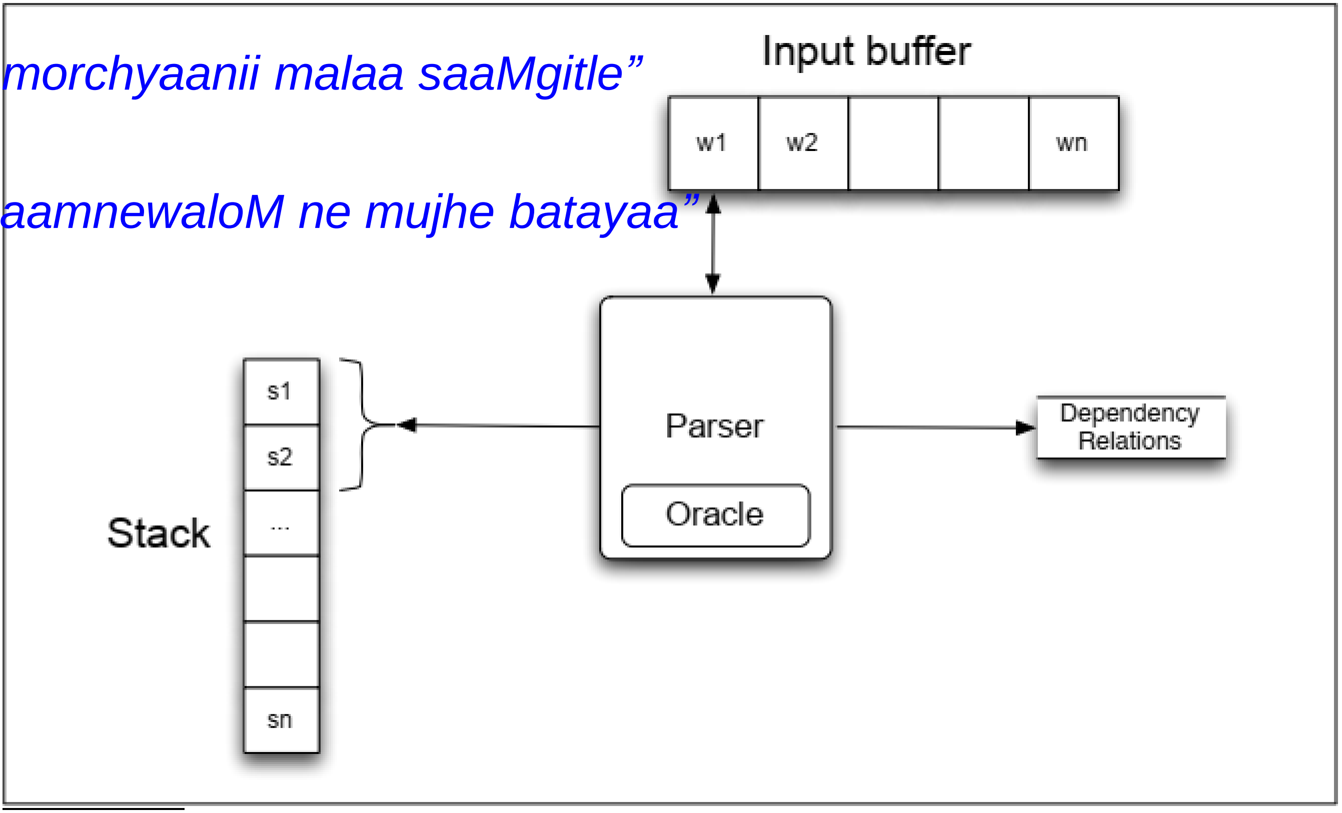
- *“people in front of the house told me”*

- *“gharaasamorchyaanii malaa saaMgitle”*

- *“ghar ke saamnewaloM ne mujhe batayaa”*

Multilingual DP

- *“people in front of the house told me”*
- *“gharaasamorchyaanii malaa saaMgitle”*
- *“ghar ke saamnewaloM ne mujhe batayaa”*



Examines top two elements of the stack and selects an action based on consulting an oracle that examines the current configuration.

Speech and Language Processing, Jurafksy & Martin, Ch-15, 2019

Essence of DP

- Cannot pop a *head* out of the stack if any of its dependents remains on the stack

- The above works if the sentence is projective