COMP 3225

Natural Language Processing

Syntactic Parsing

Stuart E. Middleton

sem03@soton.ac.uk

University of Southampton

Overview

- Ambiguity
- CKY Parsing

- <break - discussion point>

- Span-based Neural Constituency Parsing
- Evaluation

Ambiguity

- Syntactic parsing is assigning a syntactic structure to a sentence
- Constituency structures >> this lecture
- Dependency structures >> next lecture
- Ambiguity is the biggest challenge for parsing
- Structural ambiguity
  - Multiple parse trees possible in a grammar
  - Attachment ambiguity >> constituent could be attached to multiple places in a parse tree

    We saw **the Eiffel Tower** flying to Paris

    We saw NP_SUBJ[**the Eiffel Tower**] VP[flying to Paris]

    PRO_SUBJ[We] VP[saw] NP_OBJ[**the Eiffel Tower**] ADV [flying to Paris]
  - Coordination ambiguity >> phrases that can be conjoined in multiple ways

    old men and women

    [old [**men and women**]]

    [**old men**] and [**women**]
- Syntactic disambiguation is choosing the correct parse

CKY Parsing

- Dynamic Programming is useful to address ambiguity when using context free rules
- Cocke-Kasami Younger (CKY) algorithm is a classic dynamic programming approach to parsing
- Dynamic Programming = chart parsing
- CKY requires grammars in Chomsky Normal Form (CNF)
  - Right side must be (a) two non-terminal nodes or (b) single terminal node
  - Any Context Free Grammar (CFG) can be converted to CNF
  - If a CFG has a single non-terminal node (unit production) then add a dummy terminal node that itself leads to a non-terminal node

  CFG: INF-VP → to VP

  CNF: INF-VP → TO VP; TO → to
  - Use dummy terminal nodes to break up rules with three or more nodes

  CFG: S → AUX NP VP

  CNF: S → X1 VP; X1 → AUX NP

CKY Parsing

- Example

$\mathscr{L}_1$ Grammar	$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$	$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$	$S \rightarrow X1\ VP$
	$X1 \rightarrow Aux\ NP$
$S \rightarrow VP$	$S \rightarrow book \mid include \mid prefer$
	$S \rightarrow Verb\ NP$
	$S \rightarrow X2\ PP$
	$S \rightarrow Verb\ PP$
	$S \rightarrow VP\ PP$
$NP \rightarrow Pronoun$	$NP \rightarrow I \mid she \mid me$
$NP \rightarrow Proper\text{-}Noun$	$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$	$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$	$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$	$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$	$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$	$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$	$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$	$VP \rightarrow X2\ PP$
	$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$	$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$	$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$	$PP \rightarrow Preposition\ NP$

CKY Parsing

- We will encode the parse tree in a 2D matrix called a parse table
- Indices before and after tokens are called fenceposts

  Book that flight

  $_0$ Book $_1$ that $_2$ flight $_3$

- Each cell represents the parse tree entry for (i,j)
  - i = start fencepost index for span
  - j = end fencepost index for span
  - n = length of sentence
  - span(i,j) = constituent phrase with j - i tokens
  - span(0,n) = sentence

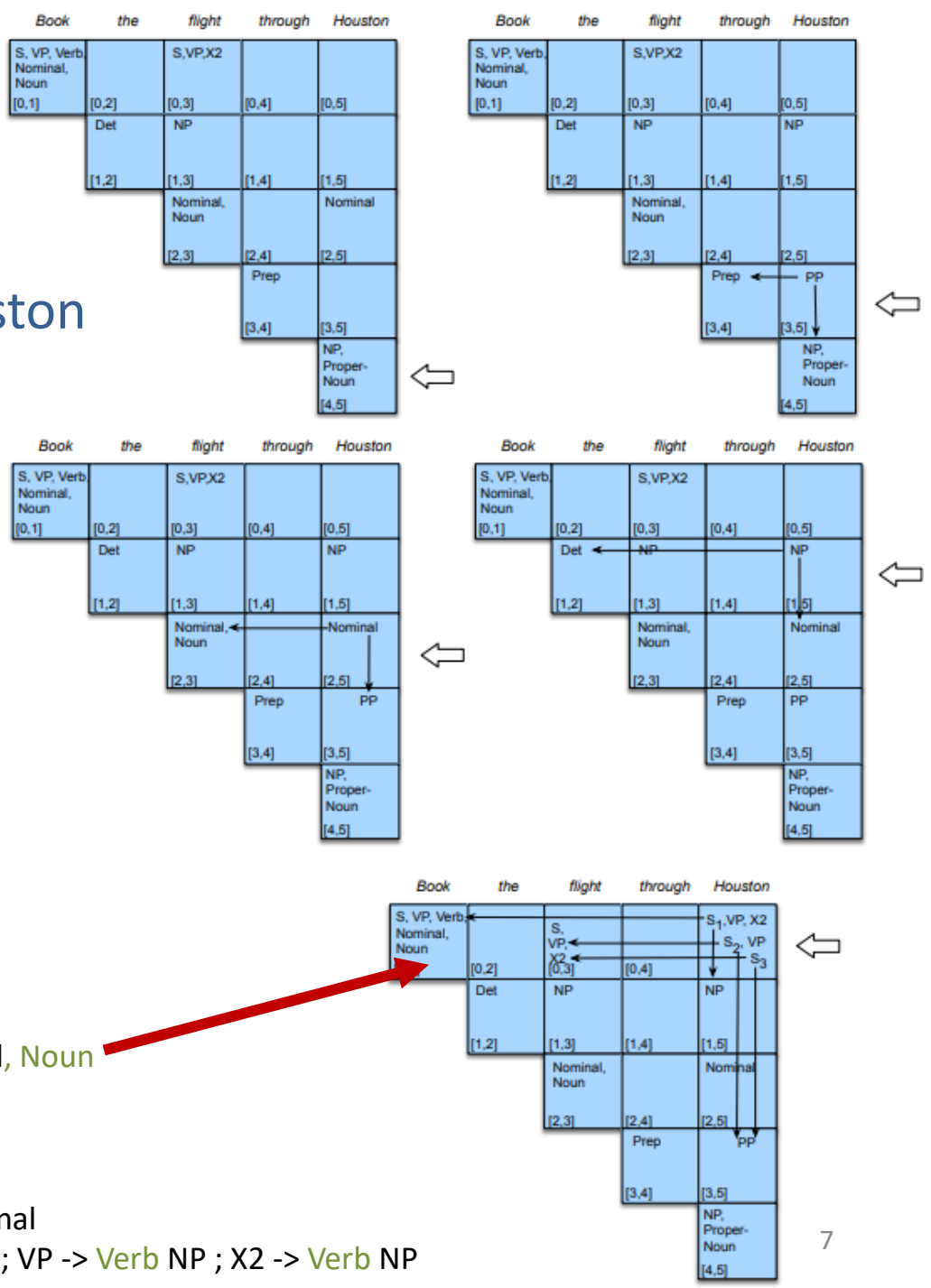
  that flight = span(1,3)

CKY Parsing

- Example parse table for sent
  Book the flight through Houston

$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$
$S \rightarrow X1\ VP$
$X1 \rightarrow Aux\ NP$
$S \rightarrow book \mid include \mid prefer$
$S \rightarrow Verb\ NP$
$S \rightarrow X2\ PP$
$S \rightarrow Verb\ PP$
$S \rightarrow VP\ PP$
$NP \rightarrow I \mid she \mid me$
$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$
$VP \rightarrow X2\ PP$
$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

and POS tags



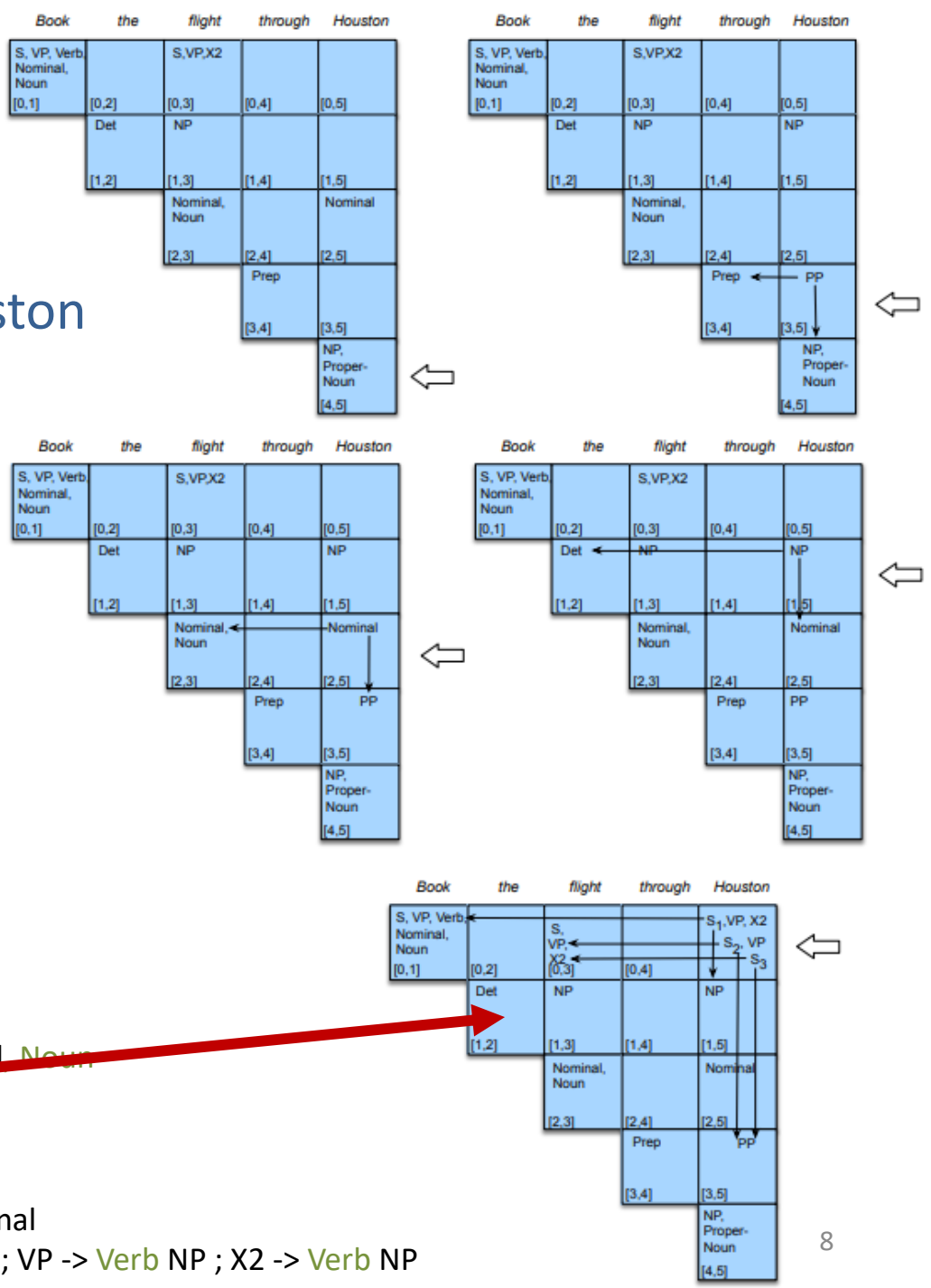
Book = span(0,1) >> productions >> S, VP, Verb, Nominal, Noun
the = span(1,2) >> productions >> Det
Book the = span(0,2) >> productions >> <none>
flight = span(2,3) >> productions >> Nominal, Noun
the flight = span(1,3) >> productions >> NP -> Det Nominal
Book the flight = span(0,3) >> productions >> S -> VP NP; VP -> Verb NP ; X2 -> Verb NP

CKY Parsing

- Example parse table for sent
  Book the flight through Houston

$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$
$S \rightarrow X1\ VP$
$X1 \rightarrow Aux\ NP$
$S \rightarrow book \mid include \mid prefer$
$S \rightarrow Verb\ NP$
$S \rightarrow X2\ PP$
$S \rightarrow Verb\ PP$
$S \rightarrow VP\ PP$
$NP \rightarrow I \mid she \mid me$
$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$
$VP \rightarrow X2\ PP$
$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

and POS tags



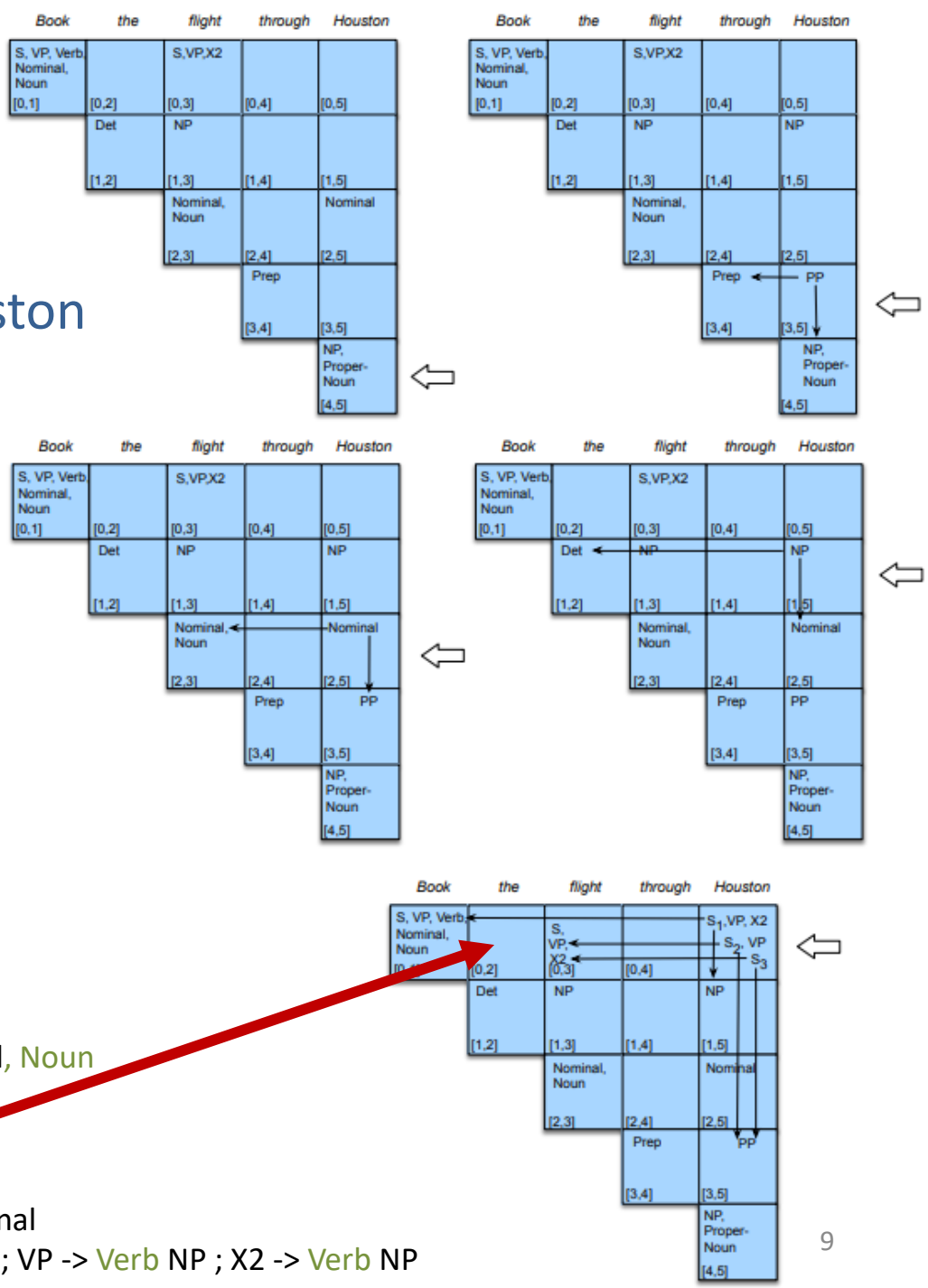
Book = span(0,1) >> productions >> S, VP, Verb, Nominal, Noun
the = span(1,2) >> productions >> Det
Book the = span(0,2) >> productions >> <none>
flight = span(2,3) >> productions >> Nominal, Noun
the flight = span(1,3) >> productions >> NP -> Det Nominal
Book the flight = span(0,3) >> productions >> S -> VP NP; VP -> Verb NP ; X2 -> Verb NP

CKY Parsing

- Example parse table for sent
  Book the flight through Houston

$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$
$S \rightarrow X1\ VP$
$X1 \rightarrow Aux\ NP$
$S \rightarrow book \mid include \mid prefer$
$S \rightarrow Verb\ NP$
$S \rightarrow X2\ PP$
$S \rightarrow Verb\ PP$
$S \rightarrow VP\ PP$
$NP \rightarrow I \mid she \mid me$
$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$
$VP \rightarrow X2\ PP$
$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

and POS tags



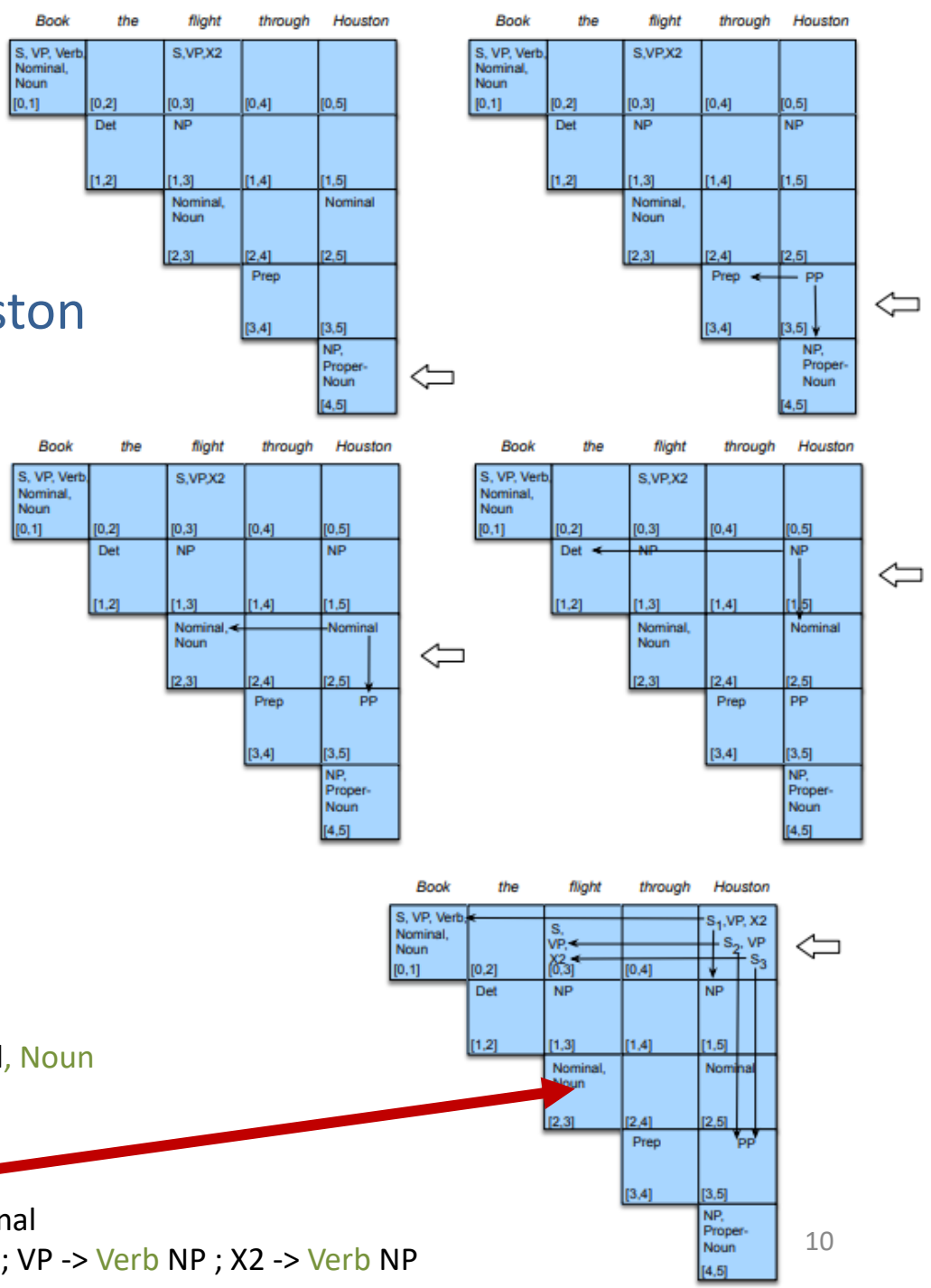
Book = span(0,1) >> productions >> S, VP, Verb, Nominal, Noun
the = span(1,2) >> productions >> Det
Book the = span(0,2) >> productions >> <none>
flight = span(2,3) >> productions >> Nominal, Noun
the flight = span(1,3) >> productions >> NP -> Det Nominal
Book the flight = span(0,3) >> productions >> S -> VP NP; VP -> Verb NP ; X2 -> Verb NP

CKY Parsing

- Example parse table for sent
  Book the flight through Houston

$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$
$S \rightarrow X1\ VP$
$X1 \rightarrow Aux\ NP$
$S \rightarrow book \mid include \mid prefer$
$S \rightarrow Verb\ NP$
$S \rightarrow X2\ PP$
$S \rightarrow Verb\ PP$
$S \rightarrow VP\ PP$
$NP \rightarrow I \mid she \mid me$
$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$
$VP \rightarrow X2\ PP$
$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

and POS tags



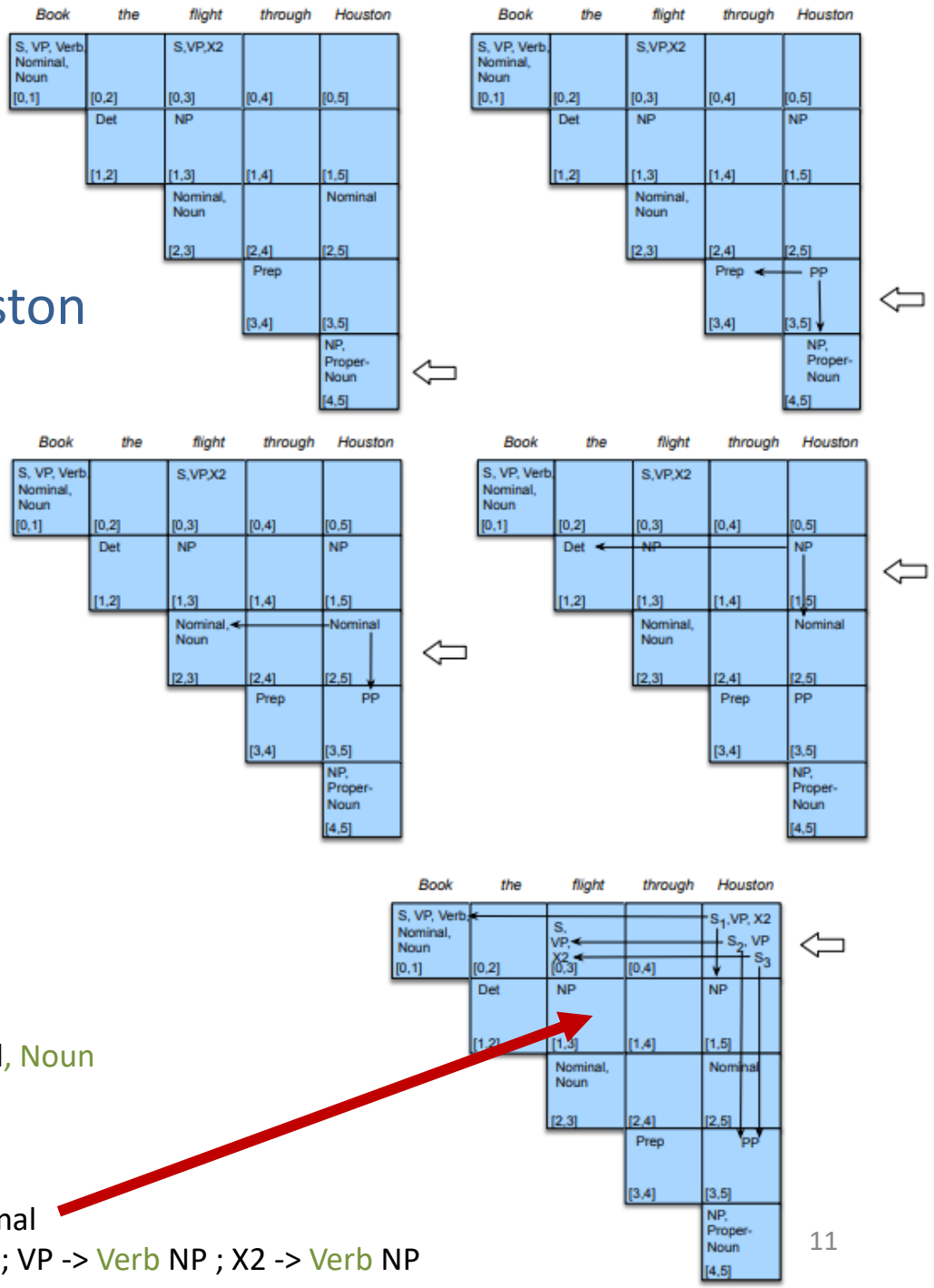
Book = span(0,1) >> productions >> S, VP, Verb, Nominal, Noun
the = span(1,2) >> productions >> Det
Book the = span(0,2) >> productions >> <none>
flight = span(2,3) >> productions >> Nominal, Noun
the flight = span(1,3) >> productions >> NP -> Det Nominal
Book the flight = span(0,3) >> productions >> S -> VP NP; VP -> Verb NP ; X2 -> Verb NP

CKY Parsing

- Example parse table for sent
  Book the flight through Houston

$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$
$S \rightarrow X1\ VP$
$X1 \rightarrow Aux\ NP$
$S \rightarrow book \mid include \mid prefer$
$S \rightarrow Verb\ NP$
$S \rightarrow X2\ PP$
$S \rightarrow Verb\ PP$
$S \rightarrow VP\ PP$
$NP \rightarrow I \mid she \mid me$
$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$
$VP \rightarrow X2\ PP$
$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

and POS tags



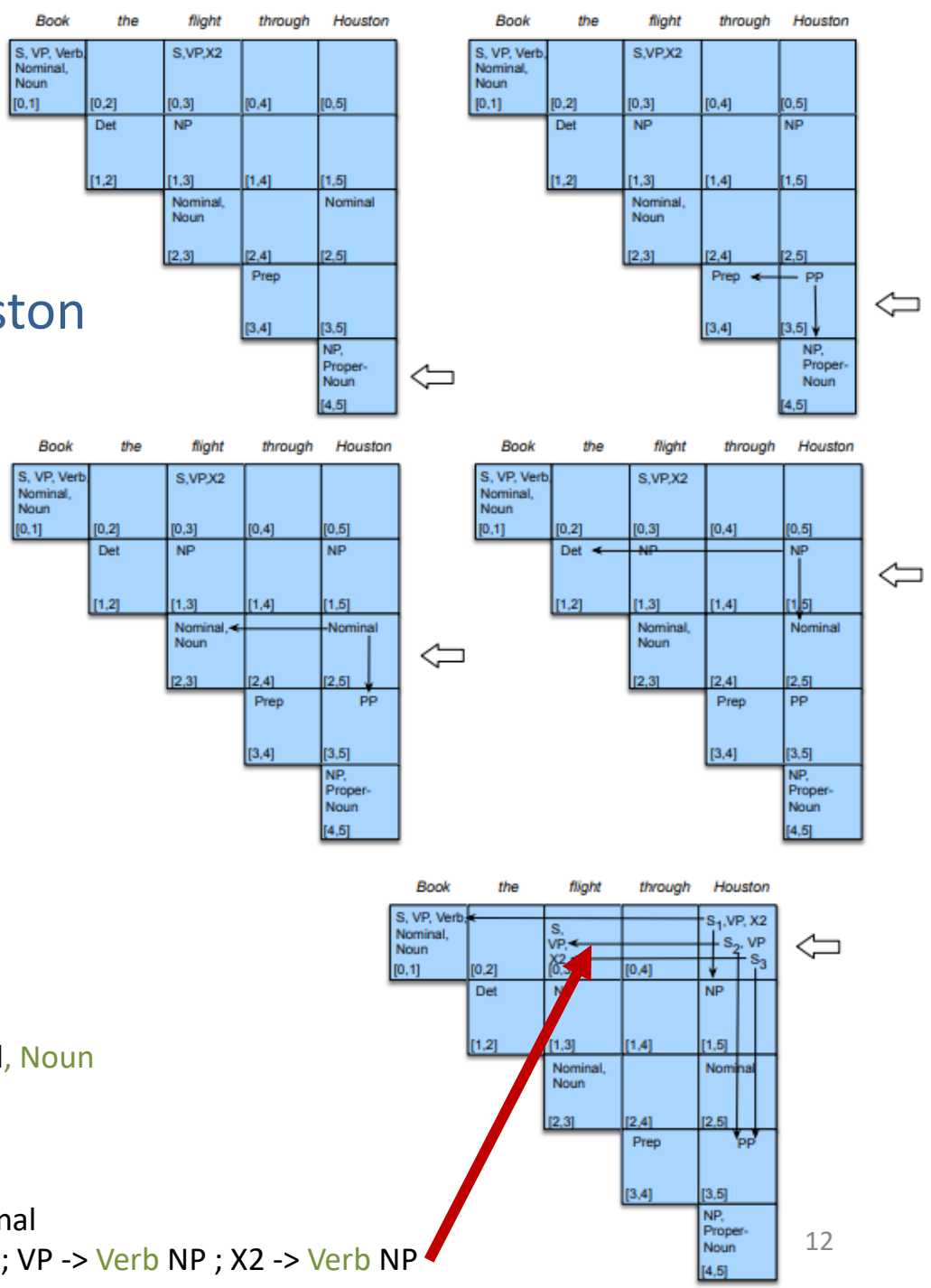
Book = span(0,1) >> productions >> S, VP, Verb, Nominal, Noun
the = span(1,2) >> productions >> Det
Book the = span(0,2) >> productions >> <none>
flight = span(2,3) >> productions >> Nominal, Noun
the flight = span(1,3) >> productions >> NP -> Det Nominal
Book the flight = span(0,3) >> productions >> S -> VP NP; VP -> Verb NP ; X2 -> Verb NP

CKY Parsing

- Example parse table for sent
  Book the flight through Houston

$\mathscr{L}_1$ in CNF
$S \rightarrow NP\ VP$
$S \rightarrow X1\ VP$
$X1 \rightarrow Aux\ NP$
$S \rightarrow book \mid include \mid prefer$
$S \rightarrow Verb\ NP$
$S \rightarrow X2\ PP$
$S \rightarrow Verb\ PP$
$S \rightarrow VP\ PP$
$NP \rightarrow I \mid she \mid me$
$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$
$VP \rightarrow X2\ PP$
$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$

and POS tags



Book = span(0,1) >> productions >> S, VP, Verb, Nominal, Noun
the = span(1,2) >> productions >> Det
Book the = span(0,2) >> productions >> <none>
flight = span(2,3) >> productions >> Nominal, Noun
the flight = span(1,3) >> productions >> NP -> Det Nominal
Book the flight = span(0,3) >> productions >> S -> VP NP; VP -> Verb NP ; X2 -> Verb NP

CKY Parsing

- CKY Algorithm
  - Below algorithm will populate a parse table
  - To make it a parser allow multiple non-terminal options per cell

function CKY-PARSE(*words*, *grammar*) **returns** *table*

for $j \leftarrow$ **from** 1 **to** LENGTH(*words*) **do**
  for all $\{A \mid A \rightarrow words[j] \in grammar\}$
    $table[j-1, j] \leftarrow table[j-1, j] \cup A$
  for $i \leftarrow$ **from** $j-2$ **downto** 0 **do**
    for $k \leftarrow i+1$ **to** $j-1$ **do**
      for all $\{A \mid A \rightarrow BC \in grammar$ **and** $B \in table[i,k]$ **and** $C \in table[k,j]\}$
        $table[i,j] \leftarrow table[i,j] \cup A$

Loop on columns 1..N

Populate end column cell with its a POS production

Loop on all possible cells in column (bottom up)

Populate cell if there is a production that will generate its value
e.g. Book the flight, span(0,3)

i = 0, j = 3, k = 1, A = S
B = table[i,k] = VP = Book
C = table[k,j] = NP = the flight

Break

- Panopto Quiz - discussion point

- Does the CKY parsing algorithm help disambiguate possible parse trees?

Yes
Sort of
No

Break

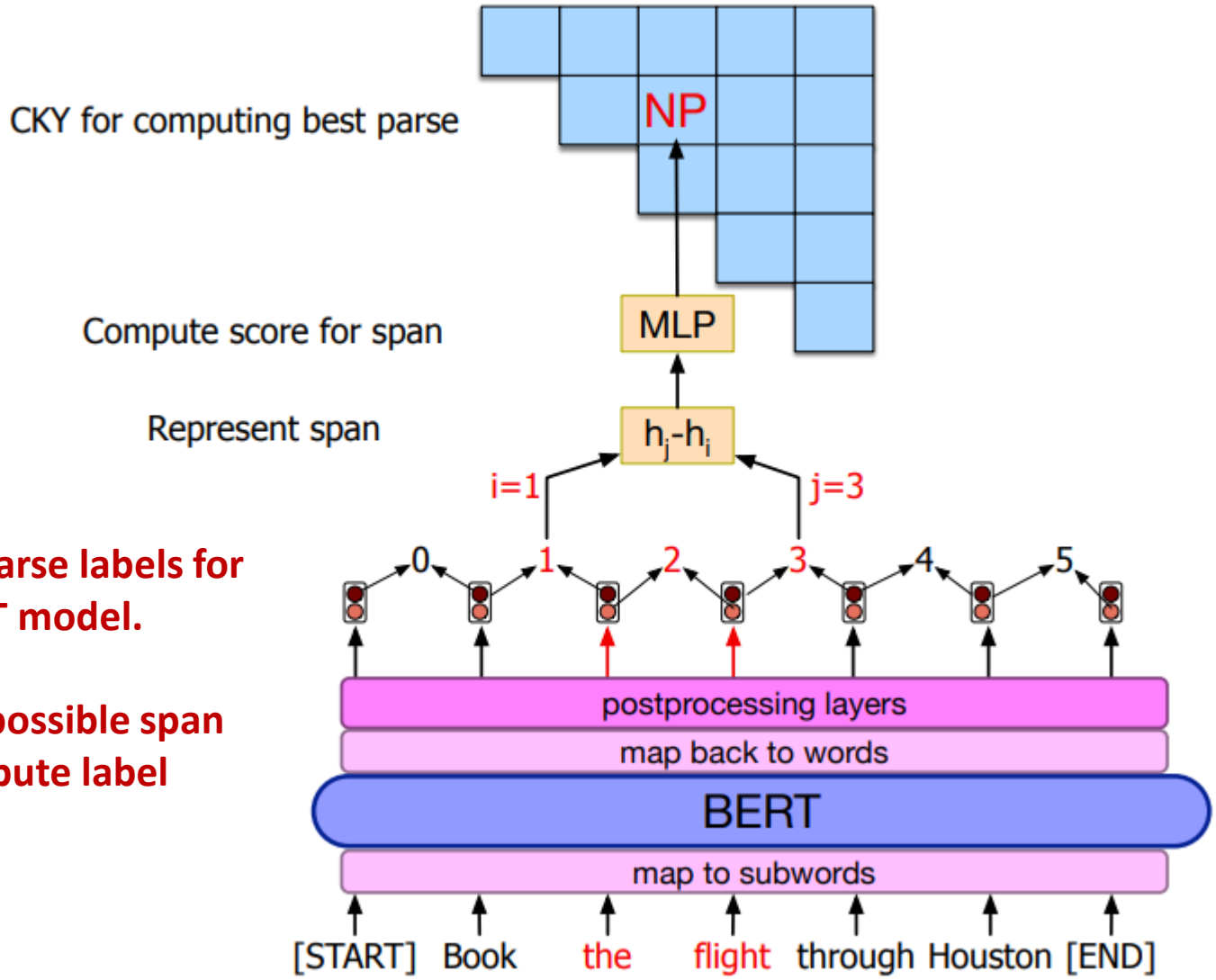
- Panopto Quiz - discussion point

- Does the CKY parsing algorithm help disambiguate possible parse trees?

Yes
Sort of
No, it populates the parse table with all possible parse trees - it does not choose the best one! For that we need to add a classifier (see next!)

Span-based Neural Constituency Parsing

- Neural CKY - span-based constituency parsing



Classify constituency parse labels for text spans using a BERT model.

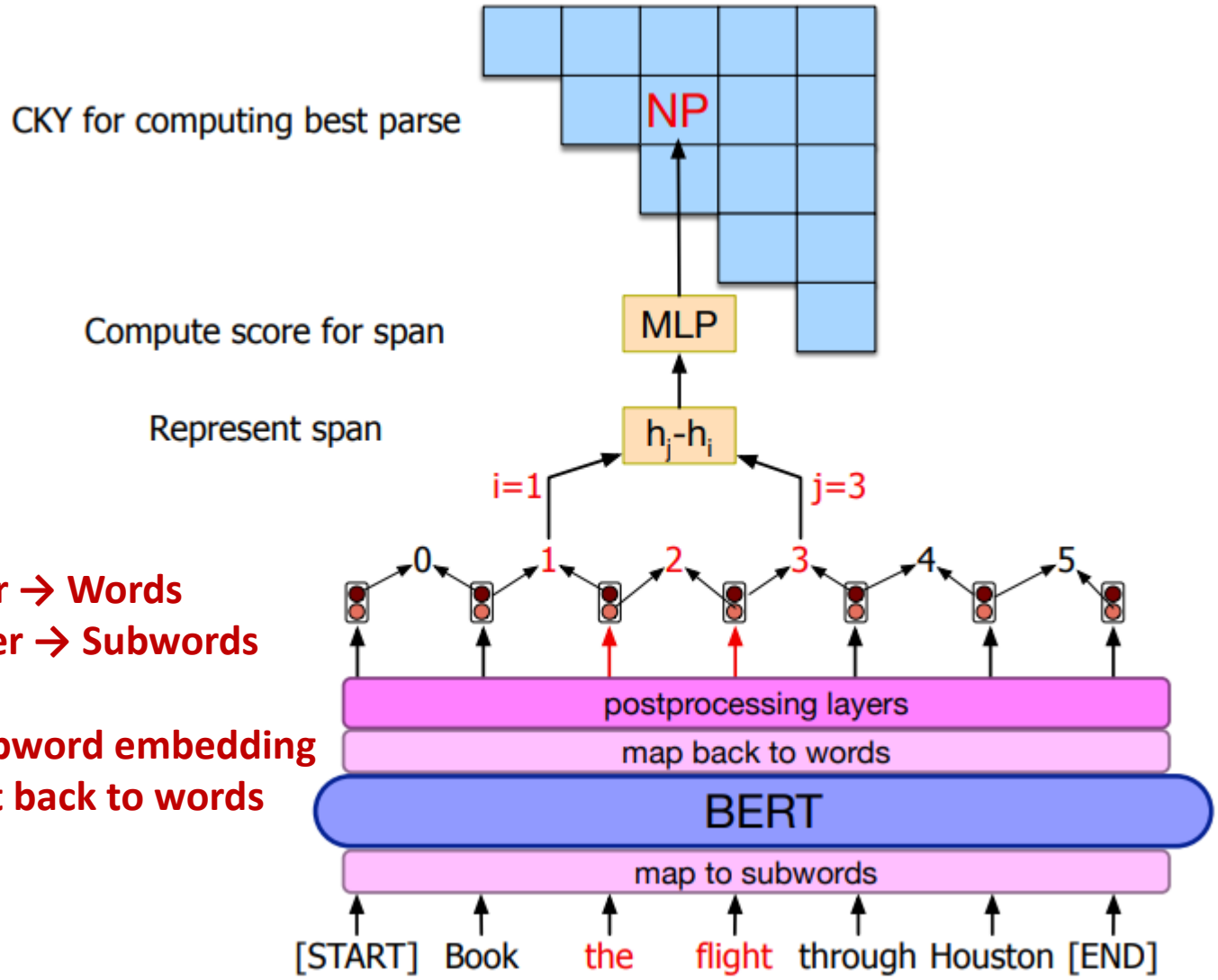
Provide as input each possible span in a sentence and compute label

e.g.
span(1,3) = the flight

Cite: Gaddy, D., Stern, M., and Klein, D. (2018). What's going on in neural constituency parsers? an analysis. NAACL HLT

Span-based Neural Constituency Parsing

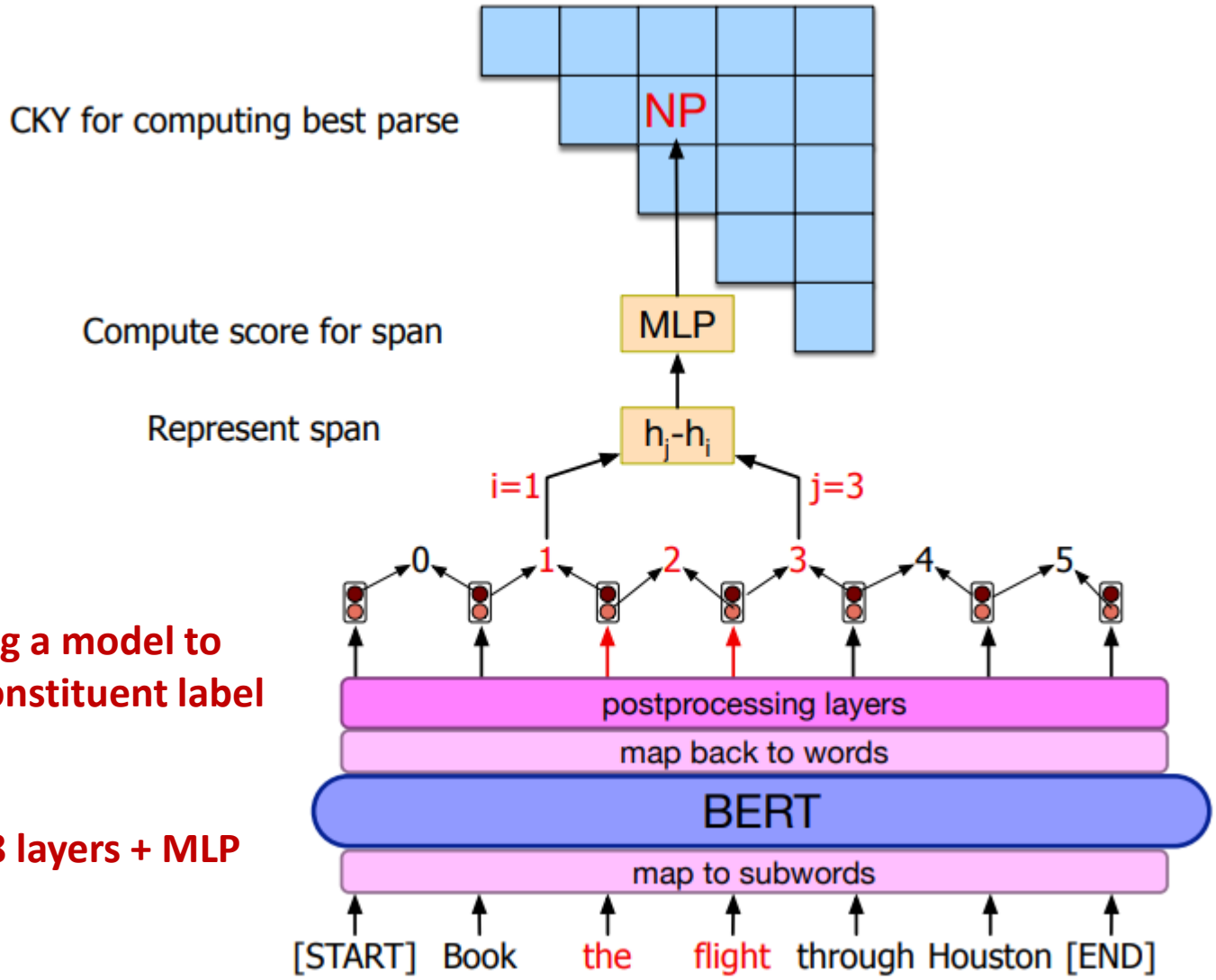
- Neural CKY - span-based constituency parsing



Sent → Word Tokenizer → Words
→ Word Piece Tokenizer → Subwords
→ BERT embeddings
Use first (or last) subword embedding for each word to get back to words
→ Word embedding

Span-based Neural Constituency Parsing

- Neural CKY - span-based constituency parsing

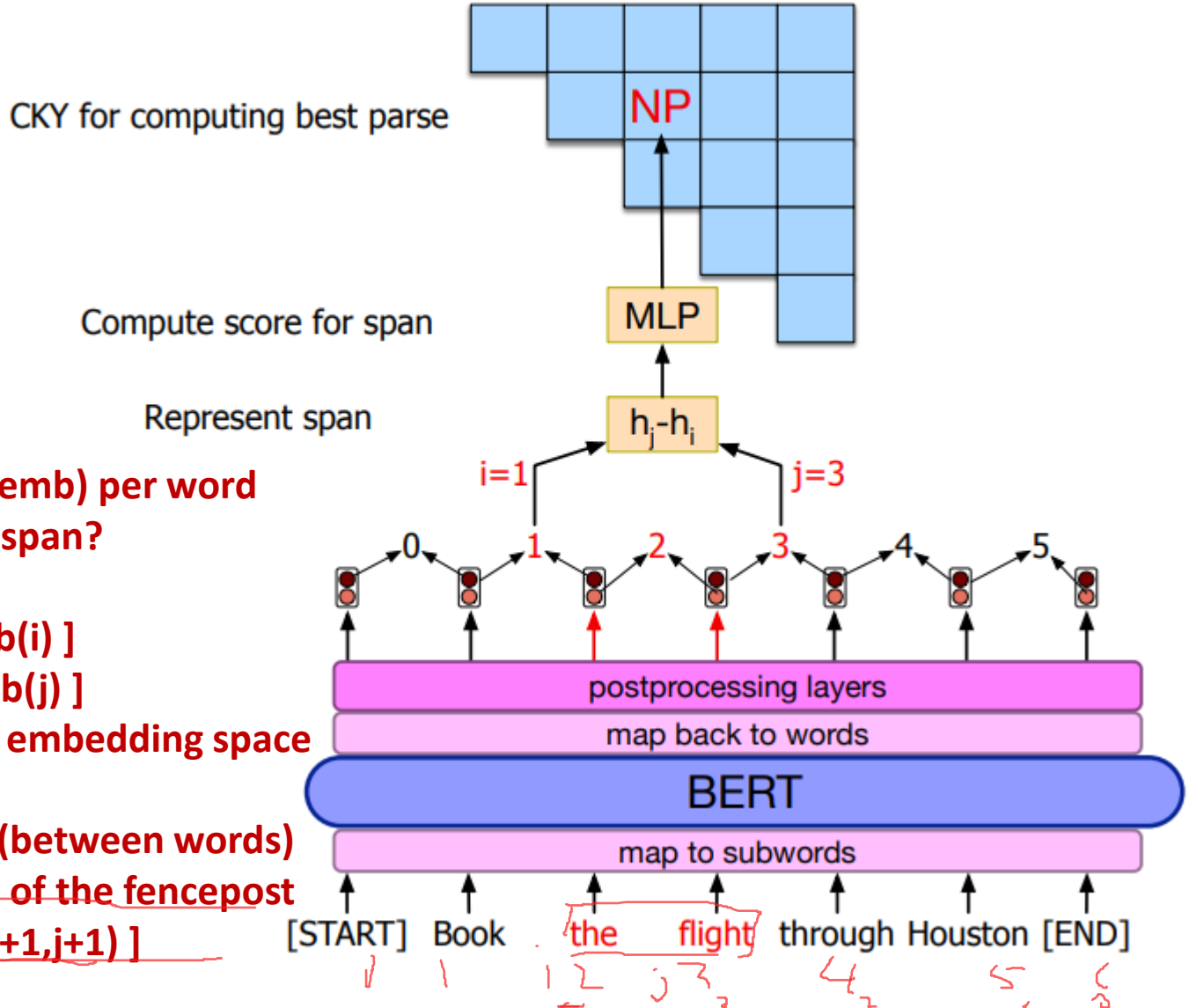


Word embedding
→ post processing using a model to compute possible constituent label scores

e.g. Transformer with 8 layers + MLP

Span-based Neural Constituency Parsing

- Neural CKY - span-based constituency parsing



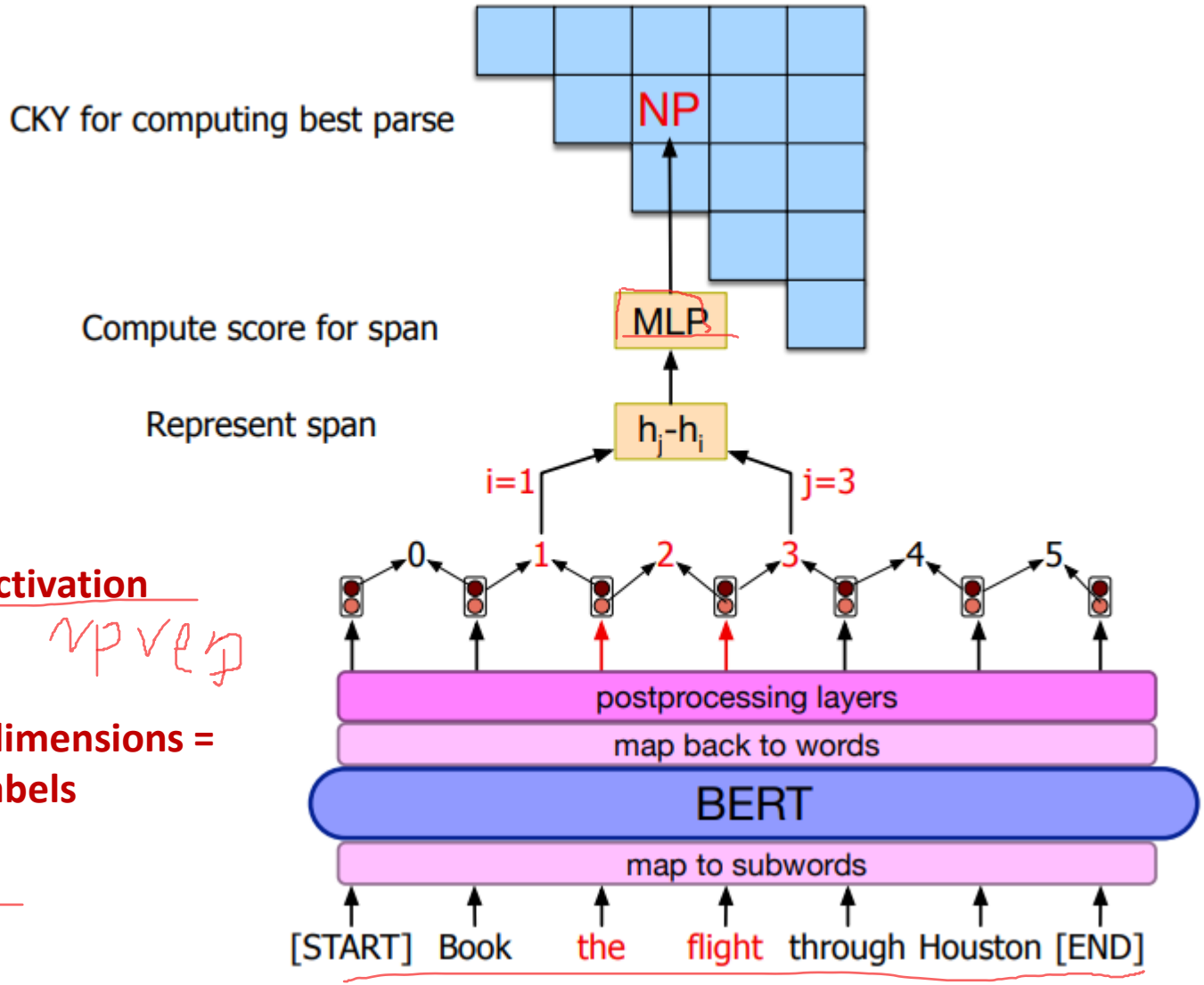
How can embeddings (emb) per word be used to represent a span?

fwd(i,j) = [emb(j) - emb(i)]
bwd(i,j) = [emb(i) - emb(j)]
= directional vectors in embedding space

But .. need fenceposts (between words) so concat left and right of the fencepost
v(i,j) = [fwd(i,j) ; bwd(i+1,j+1)]

Span-based Neural Constituency Parsing

- Neural CKY - span-based constituency parsing



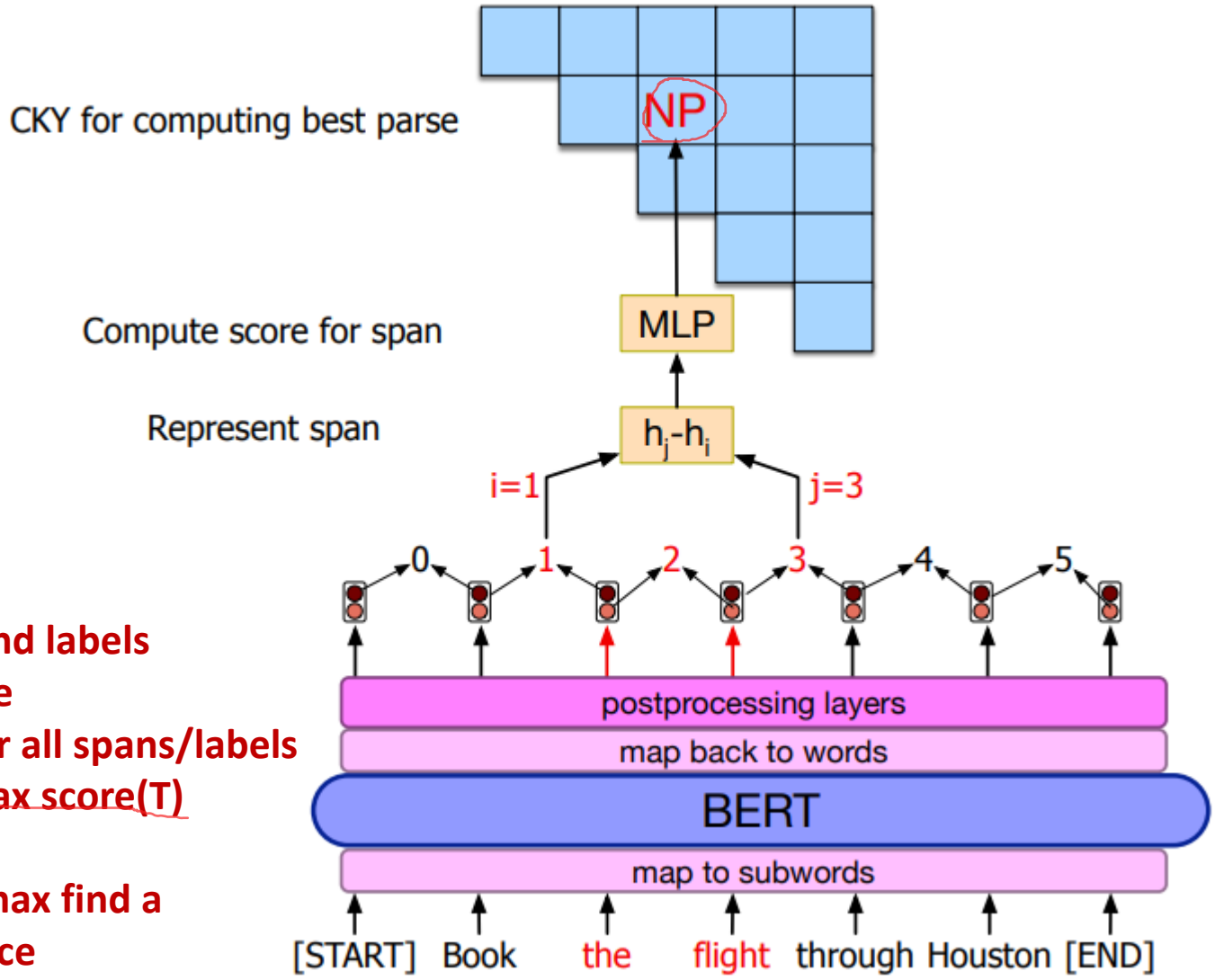
Compute score using activation function g e.g. ReLU

MLP output layer has dimensions = num of non-terminal labels

score(i,j,L) = g(v(i,j))
where L = label set

Span-based Neural Constituency Parsing

- Neural CKY - span-based constituency parsing



T = parse tree
= (i,j,l) for all spans and labels
score(T) = score for tree
= sum(score(i,j,l)) for all spans/labels
best parse tree = argmax score(T)

note: 95% of time argmax find a complete tree in practice

Evaluation

- Parsers are usually evaluated with datasets annotated with a single parse tree per sentence
  - Example dataset - Penn Treebank 3

  - Metric = F1
  - TP = correct constituent within reference (gold) parse tree

  - Community metric = PARSEVAL
  - standard implementation = evalb = https://nlp.cs.nyu.edu/evalb/

Required Reading

- Constituency Grammars
  - Jurafsky and Martin, Speech and Language Processing, 3rd edition (online) >> chapter 13

Questions

- Panopto Quiz - 1 minute brainstorm for interactive questions

  Please write down in Panopto quiz in **1 minute** two or three questions that you would like to have answered at the next interactive session.

  Do it **right now** while its fresh.

  Take a screen shot of your questions and **bring them with you** at the interactive session so you have something to ask.