Parsing: PCFGs

Ling 571
Deep Processing Techniques for NLP
January 21, 2015

Roadmap

- Motivation: Ambiguity

- Approach:
  - Probabilistic Context-free Grammars (PCFGs)
    - Definition
    - Disambiguation
    - Parsing
    - Evaluation
    - Enhancements

Probabilistic Parsing

- Provides strategy for solving disambiguation problem
  - Compute the probability of all analyses
  - Select the most probable

Probabilistic Parsing

- Provides strategy for solving disambiguation problem
  - Compute the probability of all analyses
  - Select the most probable
- Employed in language modeling for speech recognition
  - N-gram grammars predict words, constrain search
  - Also, constrain generation, translation

PCFGs

- Probabilistic Context-free Grammars
  - Augmentation of CFGs

N	a set of **non-terminal symbols** (or **variables**)
Σ	a set of **terminal symbols** (disjoint from N)
R	a set of **rules** or productions, each of the form $A \rightarrow \beta\ [p]$, where A is a non-terminal, β is a string of symbols from the infinite set of strings $(\Sigma \cup N)*$, and p is a number between 0 and 1 expressing $P(\beta\|A)$
S	a designated **start symbol**

PCFGs

- Probabilistic Context-free Grammars
  - Augmentation of CFGs

N	a set of **non-terminal symbols** (or **variables**)
Σ	a set of **terminal symbols** (disjoint from N)
R	a set of **rules** or productions, each of the form $A \rightarrow \beta$ $\boxed{[p]}$, where A is a non-terminal, β is a string of symbols from the infinite set of strings $(\Sigma \cup N)*$, and p is a number between 0 and 1 expressing $P(\beta\|A)$
S	a designated **start symbol**

PCFGs

- Augment each production with probability that LHS will be expanded as RHS
  - P(A→B) or P(A→B|A), P(RHS|LHS)

PCFGs

- Augment each production with probability that LHS will be expanded as RHS
  - P(A→B) or P(A→B|A), p(RHS|LHS)
  - Sum over all possible expansions is 1

$$\sum_{\beta} P(A \to \beta) = 1$$

PCFGs

- Augment each production with probability that LHS will be expanded as RHS
  - P(A→B) or P(A→B|A), p(RHS|LHS)
  - Sum over all possible expansions is 1

$$\sum_{\beta} P(A \to \beta) = 1$$

- A PCFG is consistent if sum of probabilities of all sentences in language is 1.
  - Recursive rules often yield inconsistent grammars

Example PCFG

Grammar		Lexicon
$S \rightarrow NP\ VP$	[.80]	$Det \rightarrow that\ [.10] \mid a\ [.30] \mid the\ [.60]$
$S \rightarrow Aux\ NP\ VP$	[.15]	$Noun \rightarrow book\ [.10] \mid flight\ [.30]$
$S \rightarrow VP$	[.05]	$\mid meal\ [.15] \mid money\ [.05]$
$NP \rightarrow Pronoun$	[.35]	$\mid flights\ [.40] \mid dinner\ [.10]$
$NP \rightarrow Proper\text{-}Noun$	[.30]	$Verb \rightarrow book\ [.30] \mid include\ [.30]$
$NP \rightarrow Det\ Nominal$	[.20]	$\mid prefer;\ [.40]$
$NP \rightarrow Nominal$	[.15]	$Pronoun \rightarrow I\ [.40] \mid she\ [.05]$
$Nominal \rightarrow Noun$	[.75]	$\mid me\ [.15] \mid you\ [.40]$
$Nominal \rightarrow Nominal\ Noun$	[.20]	$Proper\text{-}Noun \rightarrow Houston\ [.60]$
$Nominal \rightarrow Nominal\ PP$	[.05]	$\mid NWA\ [.40]$
$VP \rightarrow Verb$	[.35]	$Aux \rightarrow does\ [.60] \mid can\ [40]$
$VP \rightarrow Verb\ NP$	[.20]	$Preposition \rightarrow from\ [.30] \mid to\ [.30]$
$VP \rightarrow Verb\ NP\ PP$	[.10]	$\mid on\ [.20] \mid near\ [.15]$
$VP \rightarrow Verb\ PP$	[.15]	$\mid through\ [.05]$
$VP \rightarrow Verb\ NP\ NP$	[.05]	
$VP \rightarrow VP\ PP$	[.15]	
$PP \rightarrow Preposition\ NP$	[1.0]	

Disambiguation

- A PCFG assigns probability to each parse tree T for input S.
  - Probability of T: product of all rules to derive T

Disambiguation

- A PCFG assigns probability to each parse tree T for input S.
  - Probability of T: product of all rules to derive T

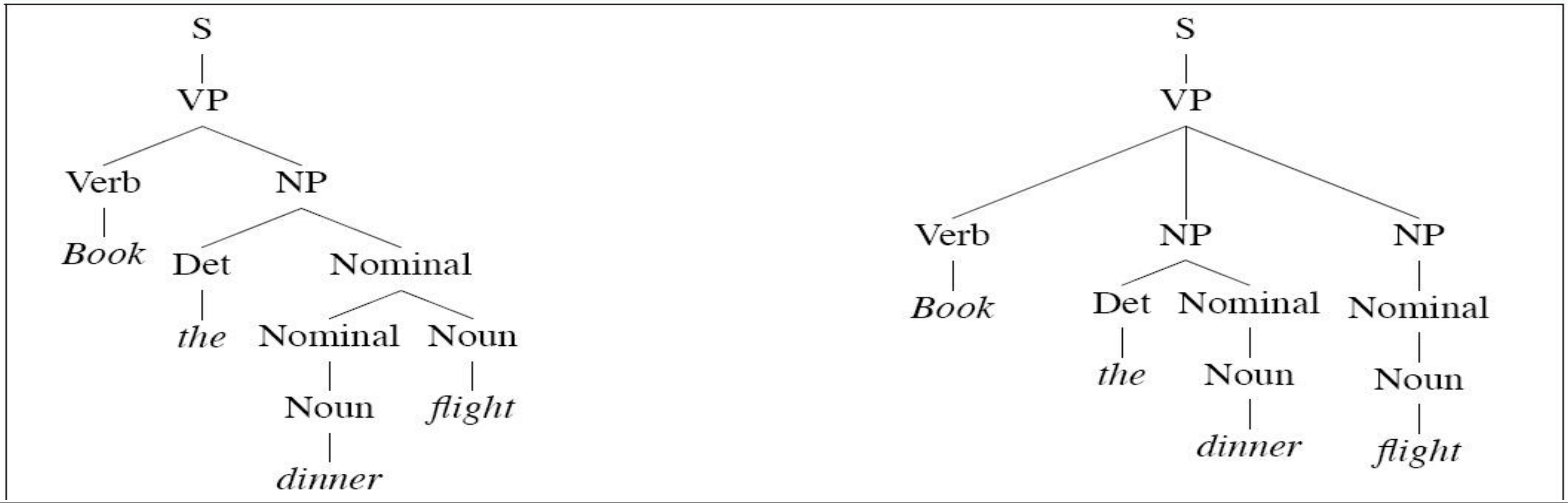
$$P(T,S) = \prod_{i=1}^{n} P(RHS_i \mid LHS_i)$$

Disambiguation

- A PCFG assigns probability to each parse tree T for input S.
  - Probability of T: product of all rules to derive T

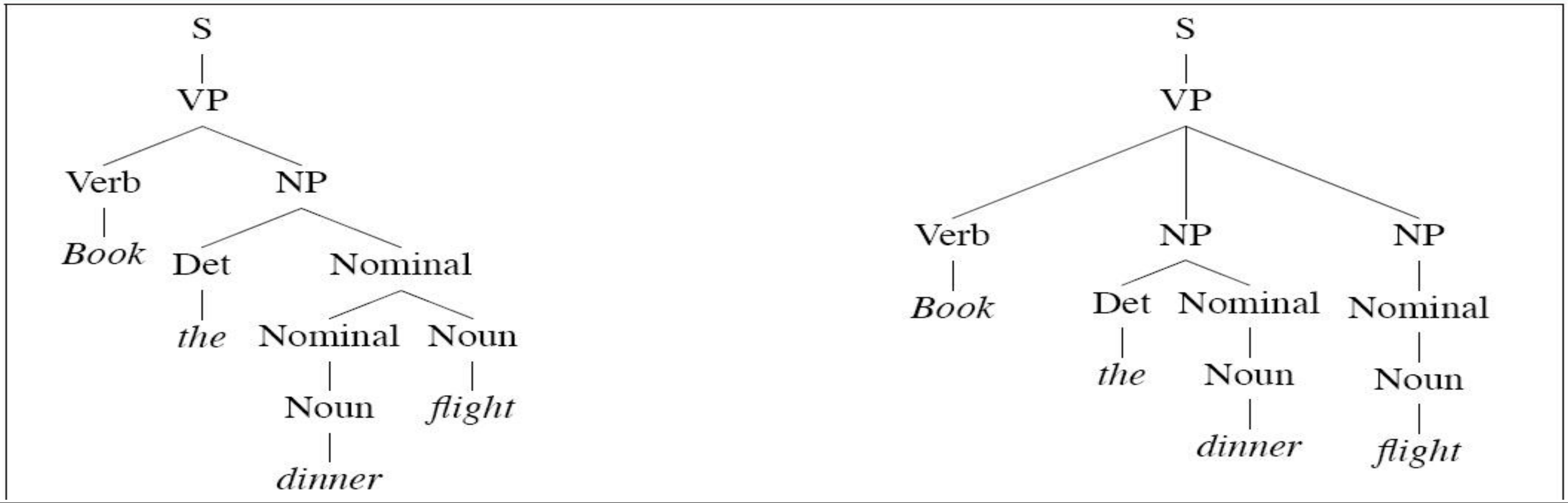
$$P(T,S) = \prod_{i=1}^{n} P(RHS_i \mid LHS_i)$$

$$P(T,S) = P(T)P(S \mid T) = P(T)$$



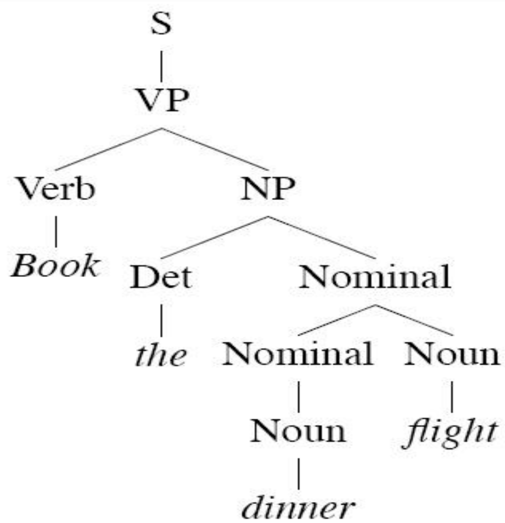
Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

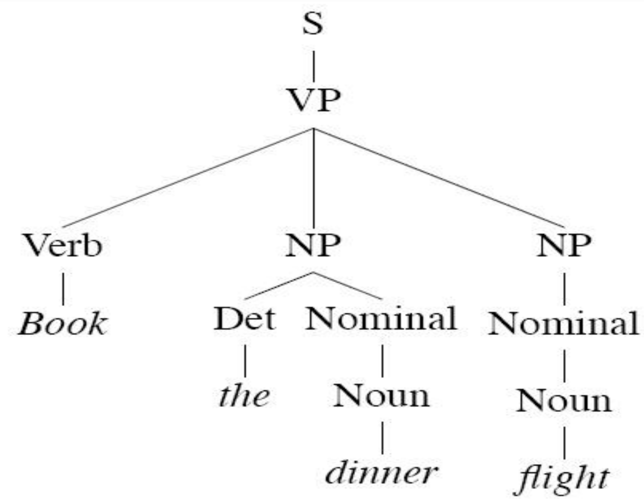
P(T,S)=0.05



Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

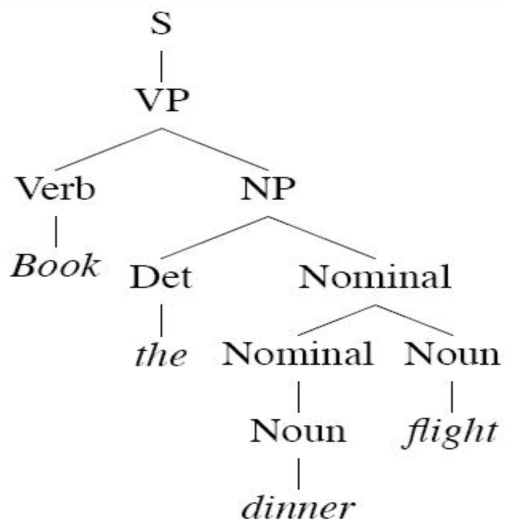
P(T,S)=0.05*0.2

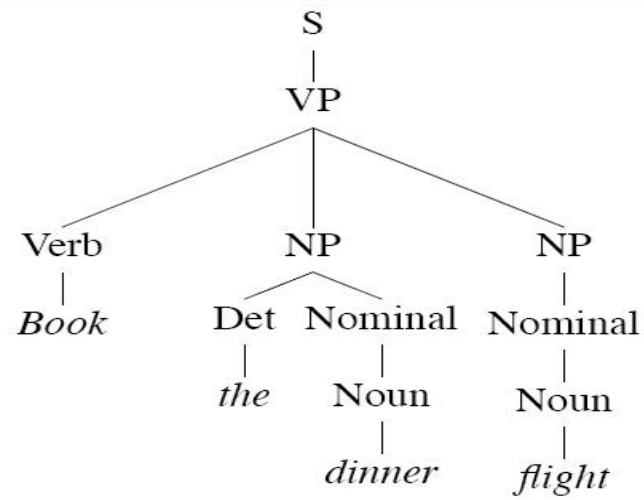




Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

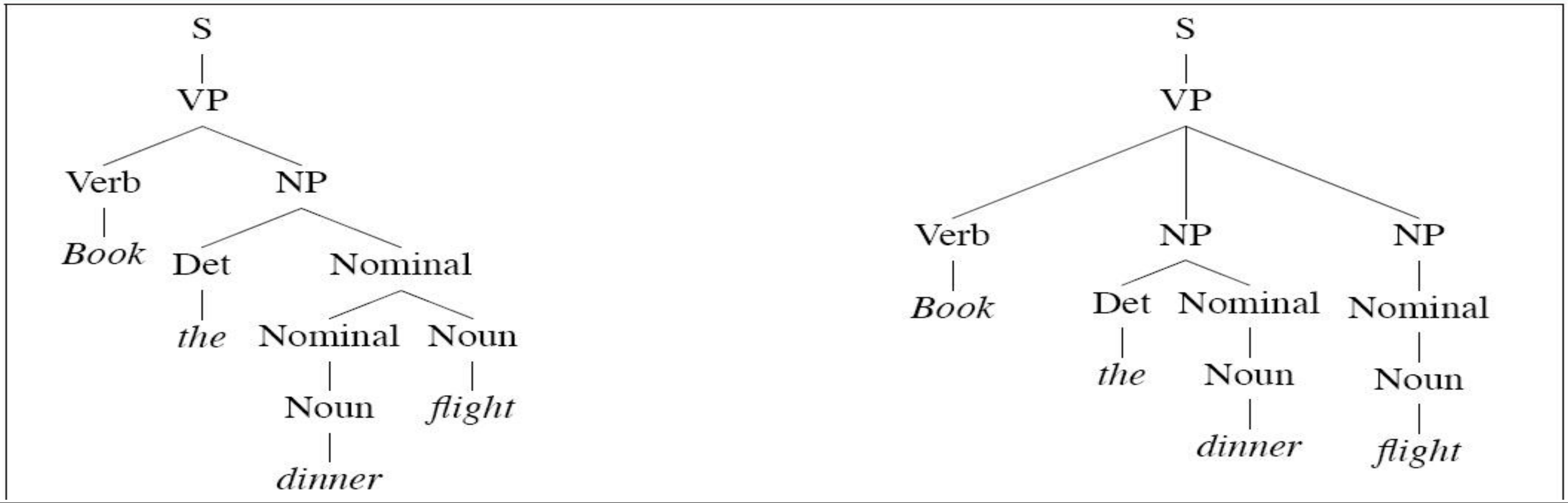
P(T,S)=0.05*0.2*0.2





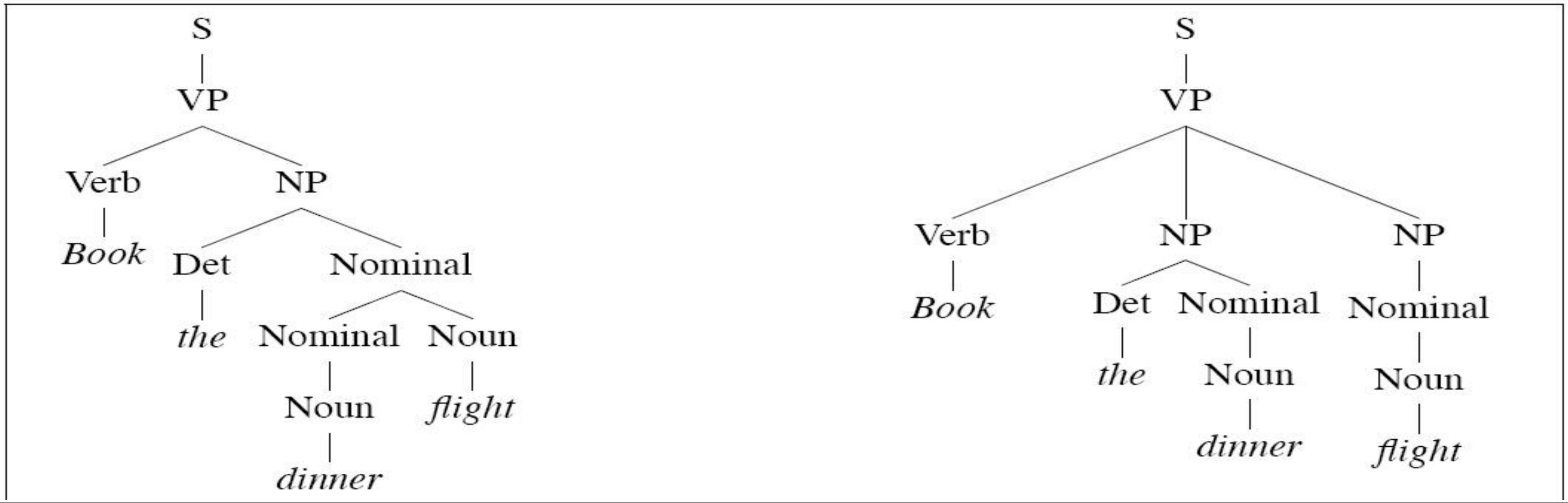
Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

$P(T,S)=0.05*0.2*0.2*0.2$



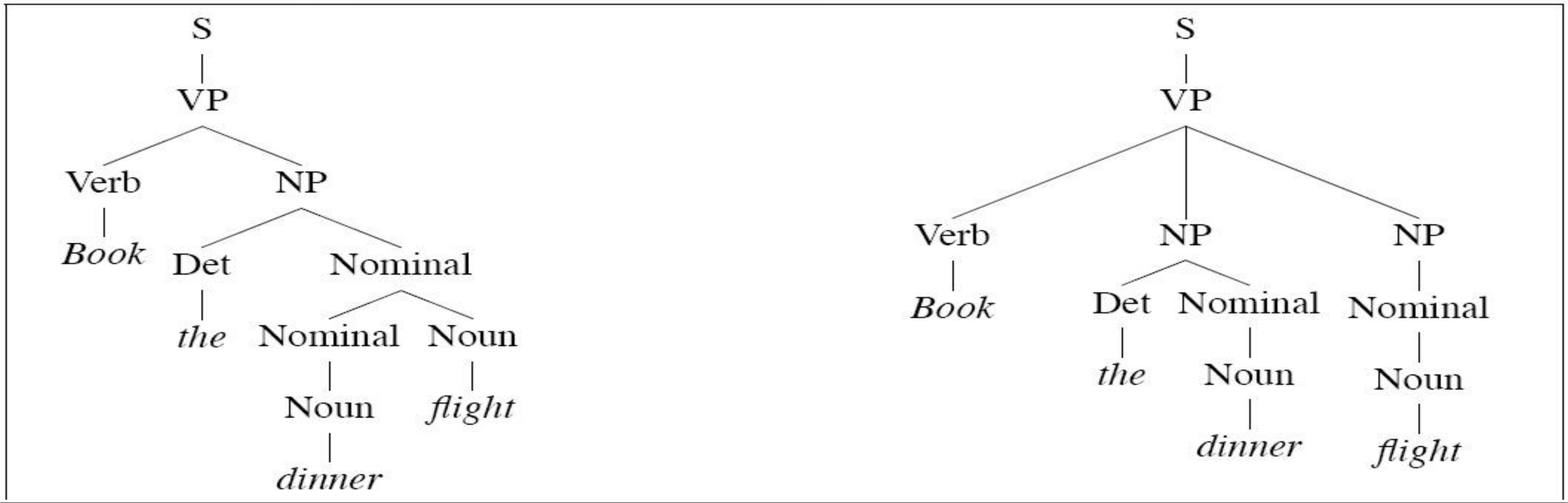
Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75



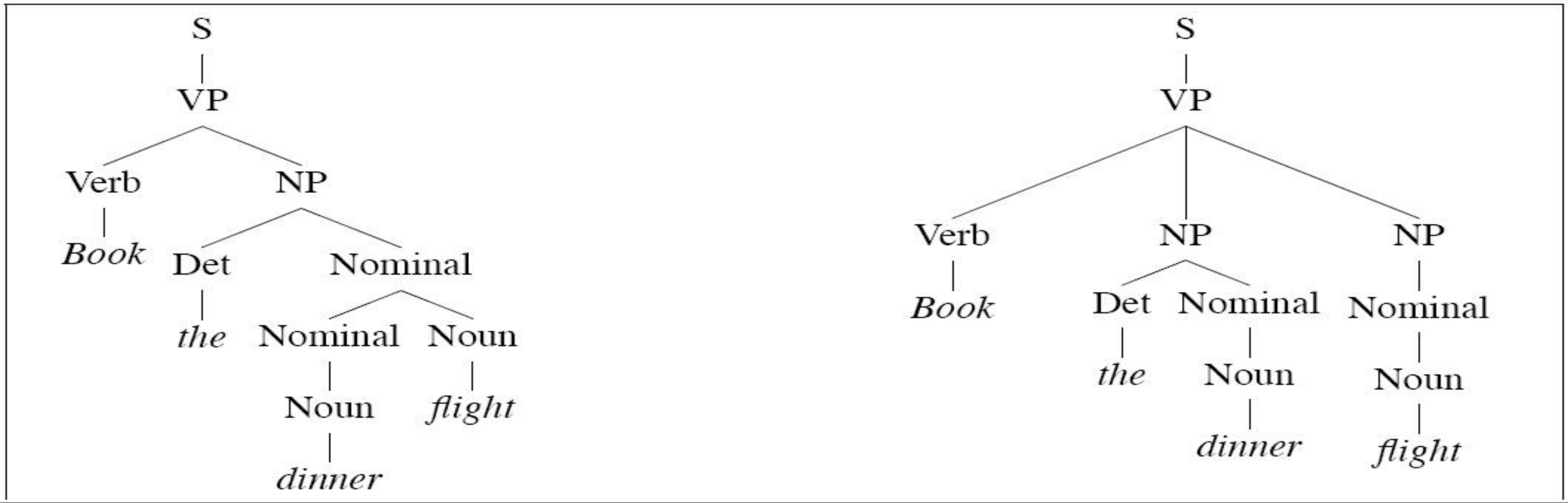
Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

$P(T,S)=0.05*0.2*0.2*0.2*0.75*$
0.3



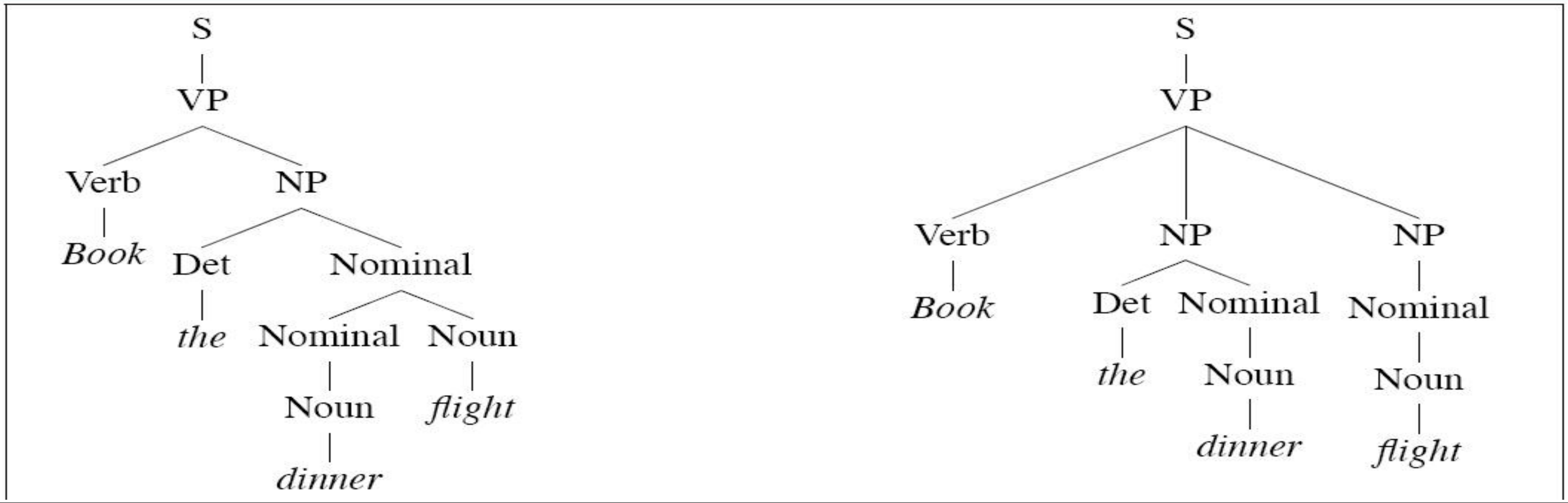
Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6



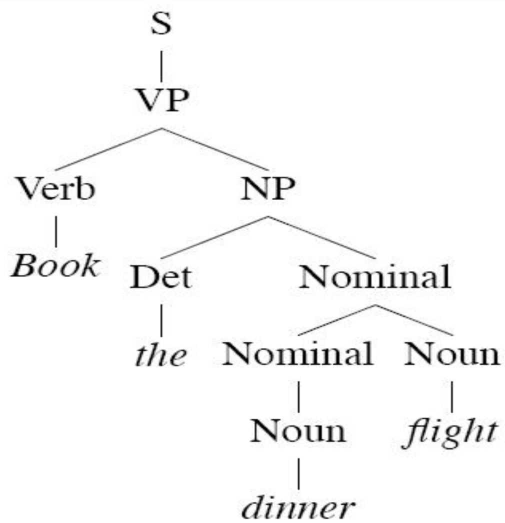
Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

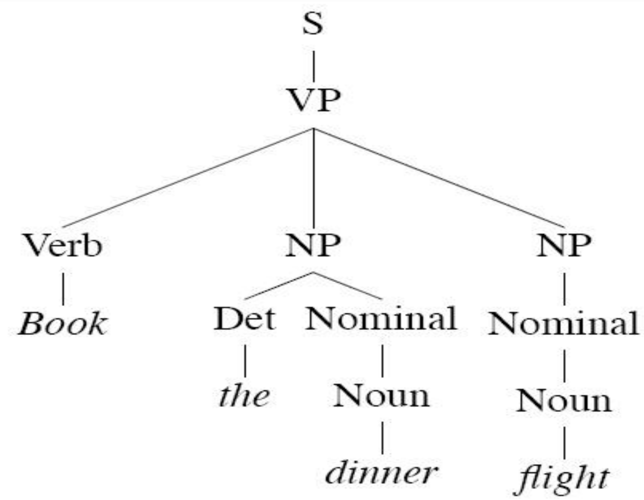
P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1



Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1*0.4=2.2x10^-6

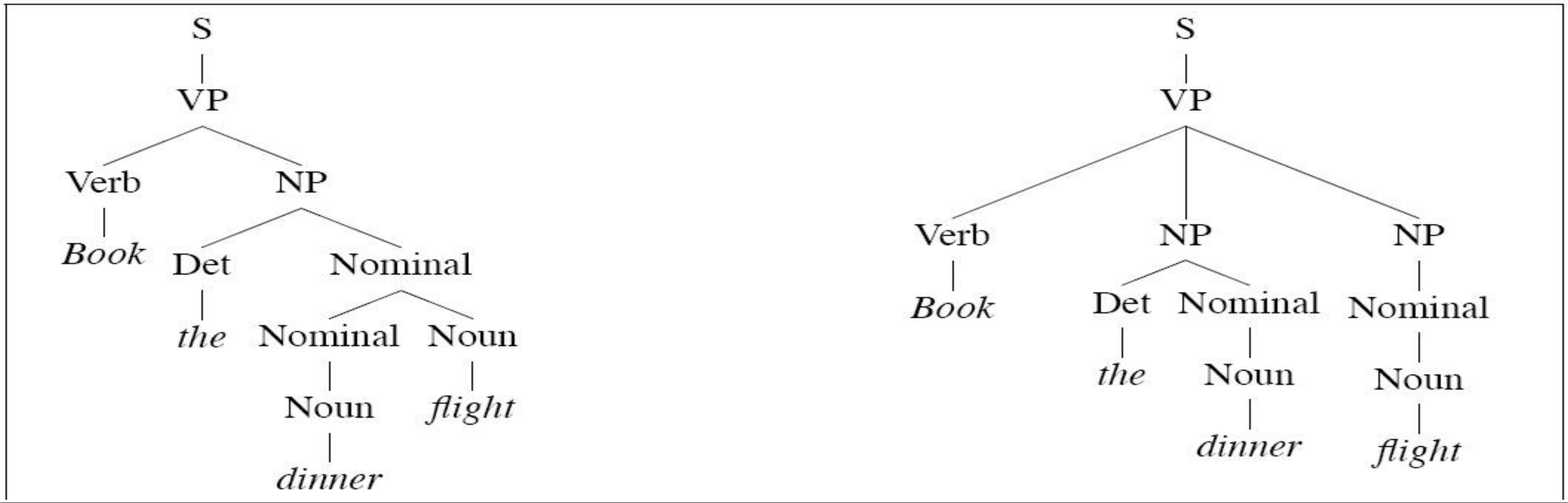




Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1*0.4=2.2x10^-6

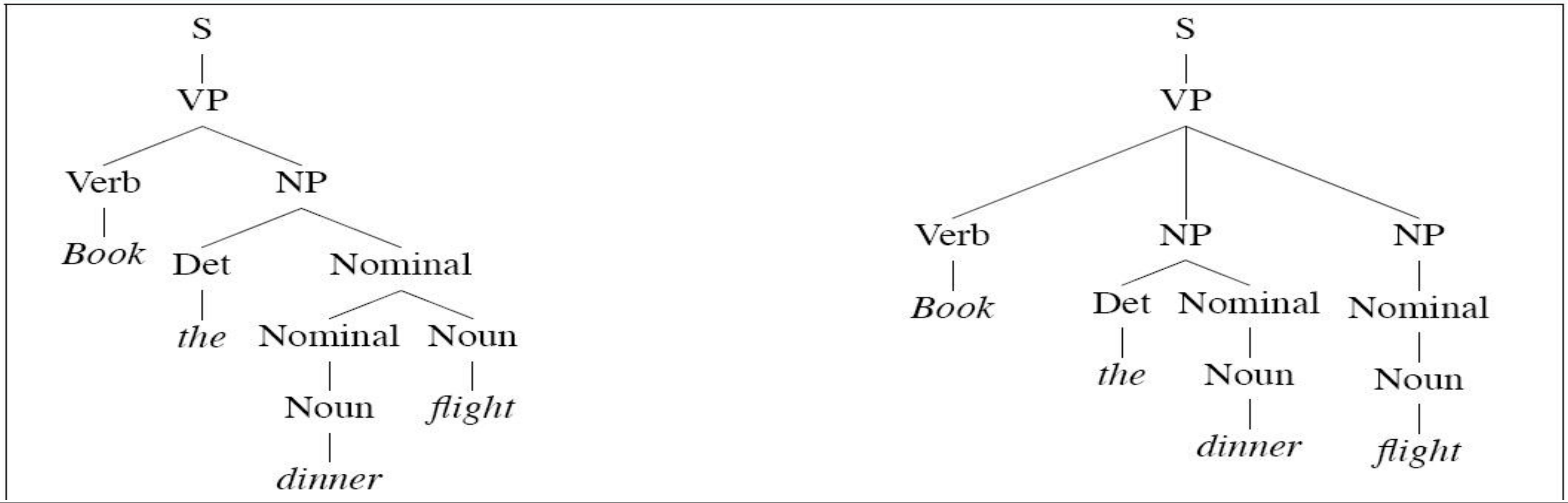
P(T,S)=0.05



Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1*0.4=2.2x10^-6

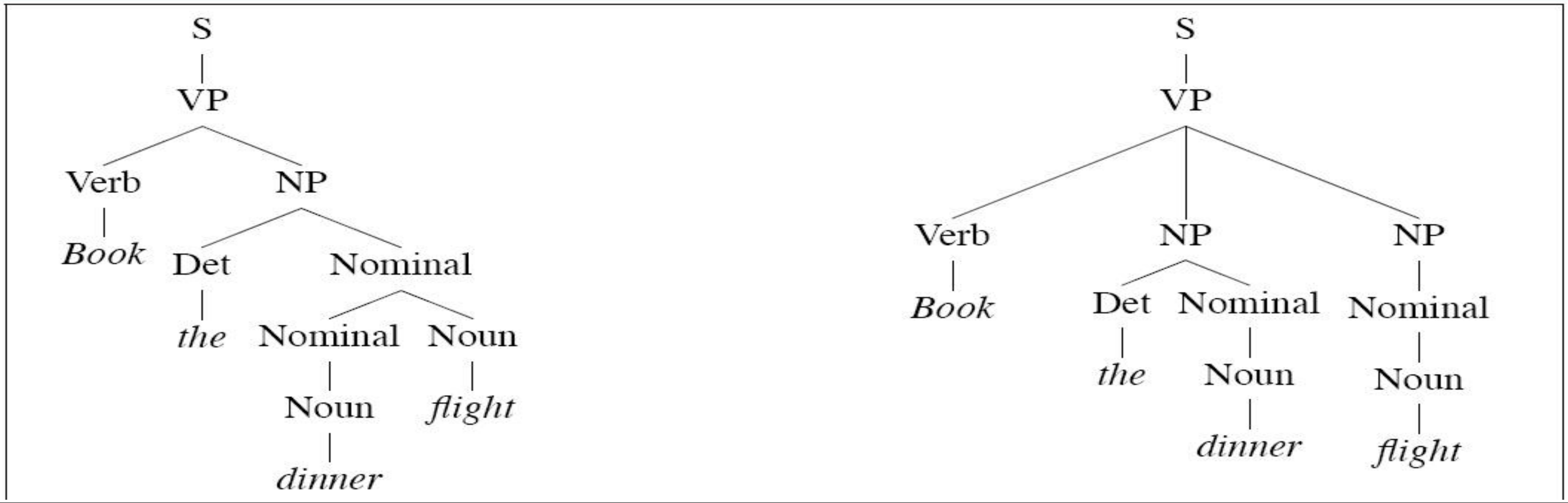
P(T,S)=0.05*0.1



Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1*0.4=2.2x10^-6

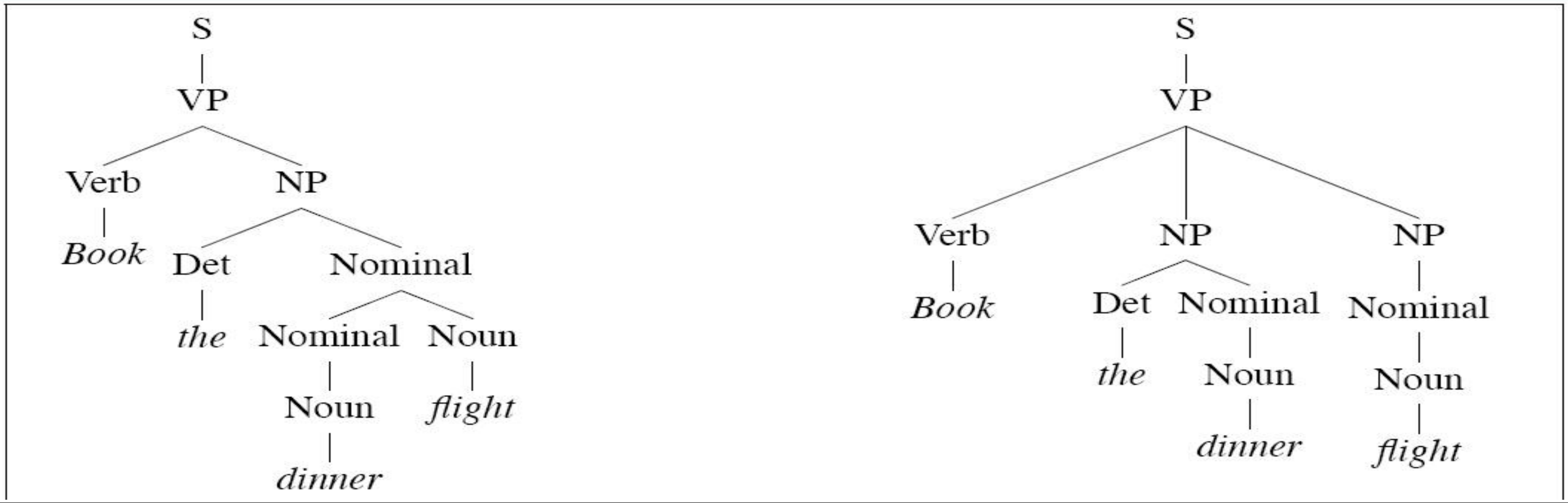
P(T,S)=0.05*0.1*0.15



Rules			P	Rules			P
S	$\longrightarrow$	VP	.05	S	$\longrightarrow$	VP	.05
VP	$\longrightarrow$	Verb NP	.20	VP	$\longrightarrow$	Verb NP NP	.10
NP	$\longrightarrow$	Det Nominal	.20	NP	$\longrightarrow$	Det Nominal	.20
Nominal	$\longrightarrow$	Nominal Noun	.20	NP	$\longrightarrow$	Nominal	.15
Nominal	$\longrightarrow$	Noun	.75	Nominal	$\longrightarrow$	Noun	.75
				Nominal	$\longrightarrow$	Noun	.75
Verb	$\longrightarrow$	book	.30	Verb	$\longrightarrow$	book	.30
Det	$\longrightarrow$	the	.60	Det	$\longrightarrow$	the	.60
Noun	$\longrightarrow$	dinner	.10	Noun	$\longrightarrow$	dinner	.10
Noun	$\longrightarrow$	flights	.40	Noun	$\longrightarrow$	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1*0.4=2.2x10^-6

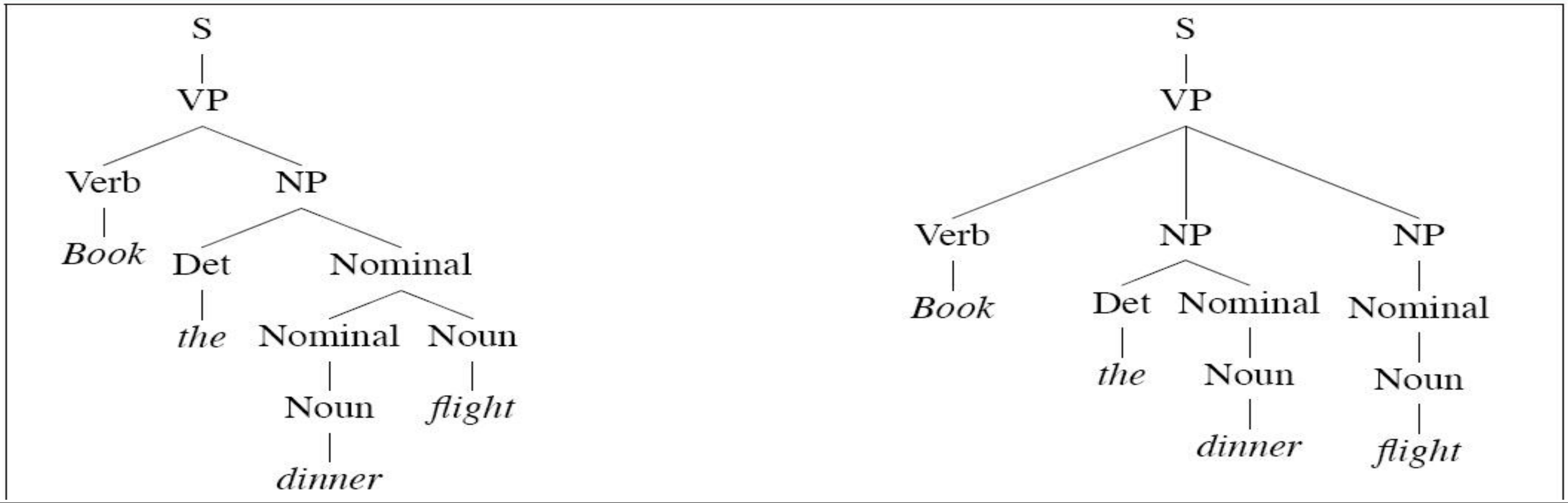
P(T,S)=0.05*0.1*0.15*0.75



Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1*0.4=2.2x10^-6

P(T,S)=0.05*0.1*0.15*0.75*0.75*



Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10	Noun	→	dinner	.10
Noun	→	flights	.40	Noun	→	flights	.40

P(T,S)=0.05*0.2*0.2*0.2*0.75*
0.3*0.6*0.1*0.4=2.2x10^-6

P(T,S)=0.05*0.1*0.15*0.75*0.75*
0.3*0.6*0.1*0.4=6.1x10^-7

Formalizing Disambiguation

- Select T such that:

$$\hat{T}(S) = \underset{T s.t, S=yield(T)}{\operatorname{argmax}} P(T)$$

  - String of words S is *yield* of parse tree over S
  - Select tree that maximizes probability of parse

Parsing Problem for PCFGs

- Select T such that:

$$\hat{T}(S) = \underset{T s.t, S = yield(T)}{\operatorname{argmax}} P(T)$$

  - String of words S is *yield* of parse tree over S
  - Select tree that maximizes probability of parse

Parsing Problem for PCFGs

- Select T such that:

$$\hat{T}(S) = \underset{Ts.t,S=yield(T)}{\operatorname{argmax}} P(T)$$

  - String of words S is *yield* of parse tree over S
  - Select tree that maximizes probability of parse

- Extend existing algorithms: CKY & Earley
  - Most modern PCFG parsers based on CKY
    - Augmented with probabilities

Probabilistic CKY

- Like regular CKY
  - Assume grammar in Chomsky Normal Form (CNF)
    - Productions:
      - A -> B C or A -> w
  - Represent input with indices b/t words
    - E.g., $_0$ Book $_1$ that $_2$ flight $_3$ through $_4$ Houston $_5$

Probabilistic CKY

- Like regular CKY
  - Assume grammar in Chomsky Normal Form (CNF)
    - Productions:
      - A -> B C or A -> w
  - Represent input with indices b/t words
    - E.g., $_0$ Book $_1$ that $_2$ flight $_3$ through $_4$ Houston $_5$
- For input string length *n* and non-terminals *V*
  - Cell[i,j,A] in (n+1)x(n+1)xV matrix contains
    - Probability that constituent A spans [i,j]

Probabilistic CKY Algorithm

function PROBABILISTIC-CKY(*words*,*grammar*) **returns** most probable parse and its probability

  for $j \leftarrow$ **from** 1 **to** LENGTH(*words*) **do**
    for all $\{ A \mid A \rightarrow words[j] \in grammar \}$
      $table[j-1, j, A] \leftarrow P(A \rightarrow words[j])$
    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** $table[i,k,B] > 0$ **and** $table[k,j,C] > 0 \}$
          if $(table[i,j,A] < P(A \rightarrow BC) \times table[i,k,B] \times table[k,j,C])$ **then**
            $table[i,j,A] \leftarrow P(A \rightarrow BC) \times table[i,k,B] \times table[k,j,C]$
            $back[i,j,A] \leftarrow \{k, B, C\}$
  return BUILD_TREE(*back*[1, LENGTH(*words*), *S*]), *table*[1, LENGTH(*words*), *S*]

PCKY Grammar Segment

S	$\rightarrow$	$NP\ VP$	.80	Det	$\rightarrow$	the	.40
NP	$\rightarrow$	$Det\ N$	.30	Det	$\rightarrow$	a	.40
VP	$\rightarrow$	$V\ NP$	.20	N	$\rightarrow$	$meal$	.01
V	$\rightarrow$	$includes$	.05	N	$\rightarrow$	$flight$	.02

PCKY Matrix:
The flight includes a meal

Det: 0.4 **[0,1]**				

PCKY Matrix:
The flight includes a meal

Det: 0.4 **[0,1]**				
	N: 0.02 [1,2]			

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024 [0,2]**			
	N: 0.02 [1,2]			

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**			
	N: 0.02 [1,2]			
		V: 0.05 [2,3]		

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024 [0,2]**			
	N: 0.02 [1,2]	[1,3]		
		V: 0.05 [2,3]		

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**		
	N: 0.02 [1,2]	[1,3]		
		V: 0.05 [2,3]		

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**		
	N: 0.02 [1,2]	[1,3]		
		V: 0.05 [2,3]		
			Det: 0.4 [3,4]	

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**		
	N: 0.02 [1,2]	[1,3]		
		V: 0.05 [2,3]	[2,4]	
			Det: 0.4 [3,4]	

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**		
	N: 0.02 [1,2]	[1,3]	[1,4]	
		V: 0.05 [2,3]	[2,4]	
			Det: 0.4 [3,4]	

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**	**[0,4]**	
	N: 0.02 [1,2]	[1,3]	[1,4]	
		V: 0.05 [2,3]	[2,4]	
			Det: 0.4 [3,4]	

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**	**[0,4]**	
	N: 0.02 [1,2]	[1,3]	[1,4]	
		V: 0.05 [2,3]	[2,4]	
			Det: 0.4 [3,4]	
				N: 0.01 [4,5]

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**	**[0,4]**	
	N: 0.02 [1,2]	[1,3]	[1,4]	
		V: 0.05 [2,3]	[2,4]	
			Det: 0.4 [3,4]	NP: 0.3*0.4*0.01 =0.0012 [3,5]
				N: 0.01 [4,5]

PCKY Matrix: The flight includes a meal

Det: 0.4 **[0,1]**	**NP: 0.3*0.4*0.02 =.0024** **[0,2]**	**[0,3]**	**[0,4]**	
	N: 0.02 [1,2]	[1,3]	[1,4]	
		V: 0.05 [2,3]	[2,4]	VP: 0.2*0.05* 0.0012=0.0 00012 [2,5]
			Det: 0.4 [3,4]	NP: 0.3*0.4*0.01 =0.0012 [3,5]
				N: 0.01 [4,5]

PCKY Matrix:
The flight includes a meal

Det: 0.4 **[0,1]**	**NP:** **0.3*0.4*0.02** **=.0024** **[0,2]**	**[0,3]**	**[0,4]**	**S: 0.8*** **0.000012*** **0.0024** **[0,5]**
	N: 0.02 [1,2]	[1,3]	[1,4]	[1,5]
		V: 0.05 [2,3]	[2,4]	VP: 0.2*0.05* 0.0012=0.0 00012 [2,5]
			Det: 0.4 [3,4]	NP: 0.3*0.4*0.01 =0.0012 [3,5]
				N: 0.01 [4,5]

Learning Probabilities

- Simplest way:
  - Treebank of parsed sentences

Learning Probabilities

- Simplest way:
  - Treebank of parsed sentences
  - To compute probability of a rule, count:
    - Number of times non-terminal is expanded

Learning Probabilities

- Simplest way:
  - Treebank of parsed sentences
  - To compute probability of a rule, count:
    - Number of times non-terminal is expanded
    - Number of times non-terminal is expanded by given rule

$$P(\alpha \rightarrow \beta \mid \alpha) = \frac{Count(\alpha \rightarrow \beta)}{\sum_{\gamma} Count(\alpha \rightarrow \gamma)}$$

Learning Probabilities

- Simplest way:
  - Treebank of parsed sentences
  - To compute probability of a rule, count:
    - Number of times non-terminal is expanded
    - Number of times non-terminal is expanded by given rule

$$P(\alpha \rightarrow \beta \mid \alpha) = \frac{Count(\alpha \rightarrow \beta)}{\sum_{\gamma} Count(\alpha \rightarrow \gamma)} = \frac{Count(\alpha \rightarrow \beta)}{Count(\alpha)}$$

  - Alternative: Learn probabilities by re-estimating
    - (Later)

Probabilistic Parser Development Paradigm

- Training:
  - (Large) Set of sentences with associated parses (Treebank)
    - E.g., Wall Street Journal section of Penn Treebank, sec 2-21
      - 39,830 sentences
    - Used to estimate rule probabilities

Probabilistic Parser Development Paradigm

- Training:
  - (Large) Set of sentences with associated parses (Treebank)
    - E.g., Wall Street Journal section of Penn Treebank, sec 2-21
      - 39,830 sentences
    - Used to estimate rule probabilities
- Development (dev):
  - (Small) Set of sentences with associated parses (WSJ, 22)
    - Used to tune/verify parser; check for overfitting, etc.

Probabilistic Parser Development Paradigm

- Training:
  - (Large) Set of sentences with associated parses (Treebank)
    - E.g., Wall Street Journal section of Penn Treebank, sec 2-21
      - 39,830 sentences
    - Used to estimate rule probabilities
- Development (dev):
  - (Small) Set of sentences with associated parses (WSJ, 22)
    - Used to tune/verify parser; check for overfitting, etc.
- Test:
  - (Small-med) Set of sentences w/parses (WSJ, 23)
    - 2416 sentences
  - Held out, used for final evaluation