

Processamento Estatístico da Linguagem Natural

Aula 19

Professora Bianca
(Sala 302 – Bloco E)
bianca@ic.uff.br

<http://www.ic.uff.br/~bianca/peln/>

Aula 19 - 18/11/2008

1

Aula de Hoje

- Cap. 13 – Jurafsky & Martin – Parsing Sintático
 - Seções 13.2, 13.3 e 13.4

Aula 19 - 18/11/2008

2

Comparação Top-Down vs. Bottom-Up

- O top-down não perde tempo explorando árvores que não podem gerar um S.
- O bottom-up não perde tempo explorando árvores que não são consistentes com a frase em questão.

Aula 19 - 18/11/2008

3

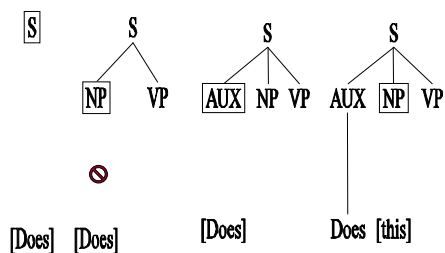
Combinando Top-Down e Bottom-Up

- Há diversas maneiras de combinar estratégias top-down com estratégias bottom-up.
- A maioria usa uma estratégia como controle e a outra como filtro.
 - Exemplo: top-down parsing com bottom-up filtering

Aula 19 - 18/11/2008

4

Bottom-Up Filtering

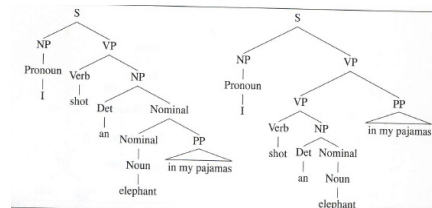


Aula 19 - 18/11/2008

5

Ambiguidade

- One morning I shot an elephant in my pajamas. How he got into my pajamas I don't know. (Groucho Marx)



Aula 19 - 18/11/2008

6

Tipos de ambiguidade

- Ambiguidade de ligação
 - Um agrupamento pode ser ligado à árvore em mais de um lugar.
 - No exemplo anterior, a PP poderia estar ligada à NP ou à VP.
 - Outro exemplo: "We saw the Eiffel Tower flying to Paris"
 - A VP "flying to Paris" pode estar ligada ao NP "the Eiffel Tower" ou a frase "We saw the Eiffel Tower"

Aula 19 - 18/11/2008

7

Tipos de Ambiguidade

- Ambiguidade de coordenação
 - Diferentes tipos de agrupamentos podem ser ligados através de conjunções como "and".
 - Exemplo: old men and women pode ser [[old men] and [women]] ou [old [men and women]]

Aula 19 - 18/11/2008

8

Muita ambiguidade

- VP → VP PP
- NP → NP PP
- "Show me the meal on flight 286 from SF to Denver"
- 14 parses!
 - Muitos com interpretações semânticas estranhas

Aula 19 - 18/11/2008

9

Muita ambiguidade

- Church and Patil (1982)
 - Número de parses de frases com PPs cresce exponencialmente com o número de PPs.

$$C(n) = \frac{1}{n+1} \binom{2n}{n}$$

PPs	Parses
1	2
2	5
3	14
4	132
5	469
6	1430

Aula 19 - 18/11/2008

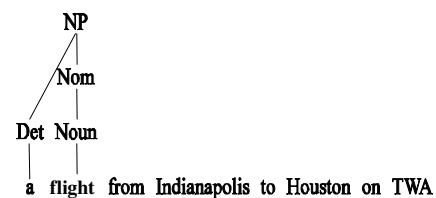
10

Evitando trabalho repetido

- Encontrar um parsing de maneira top-down ou bottom-up é um processo lento.
- Muitas vezes expansões de parte da árvore são repetidas muitas vezes.
- Considere fazer o parsing top-down do seguinte NP:
 - A flight from Indi to Houston on TWA

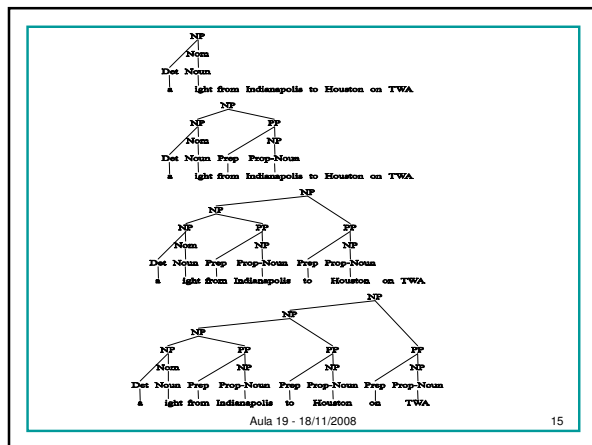
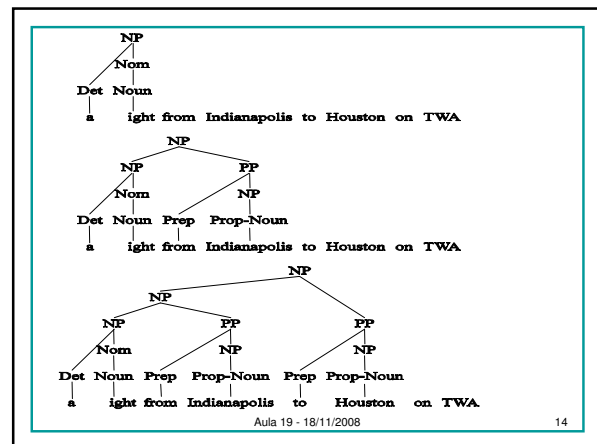
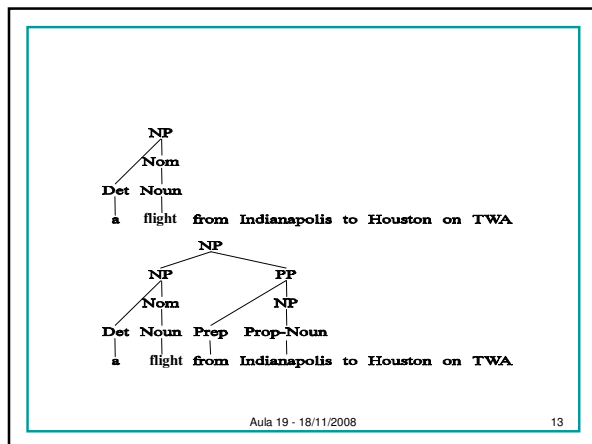
Aula 19 - 18/11/2008

11



Aula 19 - 18/11/2008

12



Programação Dinâmica

- Precisamos de um método que preencha uma tabela em tempo polinomial.
 - Esse método não repete trabalho desnecessariamente.
 - Todas as possíveis árvores estarão representadas em uma única tabela.
- Três métodos: CKY, **Earley**, Chart Parsing

Aula 19 - 18/11/2008 16

Earley Parsing

- É um método top-down.
- Preenche um vetor numa única passagem pela frase.
 - Vetor tem tamanho $N+1$; onde N é o número de palavras.
 - Cada posição i do vetor contém:
 - Agrupamentos completos e suas posições
 - Agrupamentos incompletos
 - Agrupamentos previstos

considerando que i palavras foram lidas.

Aula 19 - 18/11/2008 17

Estados de Earley

- As entradas no vetor são chamadas de estados e são representadas com regras e pontos.

$S \rightarrow \cdot VP$	Um VP é previsto
$NP \rightarrow Det \cdot Nominal$	Um NP em progresso
$VP \rightarrow V NP \cdot$	Um VP encontrado.

Aula 19 - 18/11/2008 18

Estados de Earley

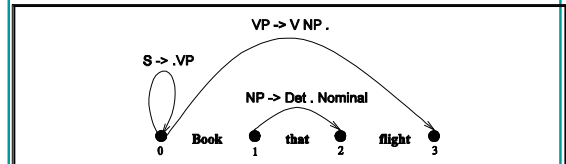
- Precisamos saber a posição na frase:

$S \rightarrow \cdot VP [0,0]$	Um VP previsto para o início da frase
$NP \rightarrow Det \cdot Nominal [1,2]$	Um NP em progresso com Det que vai de 1 a 2
$VP \rightarrow V NP \cdot [0,3]$	Um VP encontrado começando no 0 e terminando no 3.

Aula 19 - 18/11/2008

19

Graficamente



Aula 19 - 18/11/2008

20

Algoritmo de Earley

- Preencher o vetor da esquerda para a direita.
- Em cada passo, aplicar 1 de 3 operadores.
 - Predictor
 - Cria novos estados representando expectativas top-down.
 - Scanner
 - Cria previsões de palavras (regras com palavras depois do ponto) com palavras.
 - Completer
 - Quando um estado fica completo, verifica que regras precisavam daquele agrupamento completo.

Aula 19 - 18/11/2008

21

Predictor

- Dado um estado com um não-terminal à direita do ponto que **não** seja uma categoria POS.
 - Crie um novo estado para cada expansão do não-terminal.
 - Coloque esses estados na mesma entrada do vetor, começando e terminando onde o próprio estado gerador começa e termina.
 - Então o "predictor" olhando para
 - $S \rightarrow \cdot VP [0,0]$
 - vai resultar em:
 - $VP \rightarrow \cdot Verb [0,0]$
 - $VP \rightarrow \cdot Verb NP [0,0]$

Aula 19 - 18/11/2008

22

Scanner

- Dado um estado com um não-terminal à direita do ponto que seja uma categoria POS.
 - Se a próxima palavra for dessa categoria.
 - Criar um novo estado com o não-terminal à esquerda do ponto.
 - Exemplo:
 - $VP \rightarrow \cdot Verb NP [0,0]$
 - Se a próxima palavra for "book", adicionamos o estado:
 - $VP \rightarrow Verb \cdot NP [0,1]$
 - Adicione esse estado à próxima posição do vetor.

Aula 19 - 18/11/2008

23

Completer

- Aplicado a um estado quando o ponto chegou ao final.
 - O parser descobriu um agrupamento.
- Encontre e avance todos os estados anteriores que estavam esperando essa categoria.
 - Copiar estado, mover ponto, inserir no chart atual.
- Dado:
 - $NP \rightarrow Det Nominal \cdot [1,3]$
 - $VP \rightarrow Verb \cdot NP [0,1]$
- Adicionar
 - $VP \rightarrow Verb NP \cdot [0,3]$

Aula 19 - 18/11/2008

24

Earley: como sabemos que terminamos?

- Encontrar um estado S na coluna final que vá de 0 a N+1 e esteja completo.
- Forma geral do estado final:
 - $S \rightarrow \alpha \cdot [0, n+1]$

Aula 19 - 18/11/2008

25

Earley

- Mais especificamente
 1. Crie todos os estados iniciais
 2. Leia uma palavra
 1. Extenda todos os estados de acordo com a palavra lida.
 2. Adicione novas previsões
 3. Vá para 2
 3. Olhe no chart N+1 pra ver se existe um estado final.

Aula 19 - 18/11/2008

26

Exemplo

- Book that flight
- Temos que encontrar um S de 0 a 3 que seja um estado completo.

Aula 19 - 18/11/2008

27

Exemplo

Chart[0]		
$\gamma \Rightarrow S$	[0,0]	Dummy start state
$S \Rightarrow NP VP$	[0,0]	Predictor
$NP \Rightarrow Det NOMINAL$	[0,0]	Predictor
$NP \Rightarrow Proper-Noun$	[0,0]	Predictor
$S \Rightarrow Aux NP VP$	[0,0]	Predictor
$S \Rightarrow VP$	[0,0]	Predictor
$VP \Rightarrow Verb$	[0,0]	Predictor
$VP \Rightarrow Verb NP$	[0,0]	Predictor

Aula 19 - 18/11/2008

28

Exemplo

Chart[0]		
$\gamma \Rightarrow S$	[0,0]	Dummy start state
$S \Rightarrow NP VP$	[0,0]	Predictor
$NP \Rightarrow Det NOMINAL$	[0,0]	Predictor
$NP \Rightarrow Proper-Noun$	[0,0]	Predictor
$S \Rightarrow Aux NP VP$	[0,0]	Predictor
$S \Rightarrow VP$	[0,0]	Predictor
$VP \Rightarrow Verb$	[0,0]	Predictor
$VP \Rightarrow Verb NP$	[0,0]	Predictor

Chart[1]		
$Verb \Rightarrow book$	[0,1]	Scanner
$VP \Rightarrow Verb$	[0,1]	Completer
$S \Rightarrow VP$	[0,1]	Completer
$VP \Rightarrow Verb NP$	[0,1]	Completer
$NP \Rightarrow Det NOMINAL$	[1,1]	Predictor
$NP \Rightarrow Proper-Noun$	[1,1]	Predictor

Aula 19 - 18/11/2008

29

Exemplo

Chart[1]		
$Verb \Rightarrow book$	[0,1]	Scanner
$VP \Rightarrow Verb$	[0,1]	Completer
$S \Rightarrow VP$	[0,1]	Completer
$VP \Rightarrow Verb NP$	[0,1]	Completer
$NP \Rightarrow Det NOMINAL$	[1,1]	Predictor
$NP \Rightarrow Proper-Noun$	[1,1]	Predictor

Chart[2]		
$Det \Rightarrow that$	[1,2]	Scanner
$NP \Rightarrow Det NOMINAL$	[1,2]	Completer
$NOMINAL \Rightarrow Noun$	[2,2]	Predictor
$NOMINAL \Rightarrow Noun NOMINAL$	[2,2]	Predictor

Aula 19 - 18/11/2008

30

Convertendo Earley a um Parser

- O Earley como visto até agora serve para reconhecer a linguagem da gramática, mas não gera a árvore de parse.
- Com a adição de alguns ponteiros teremos um parser.
- Temos que fazer o “Completer” apontar de onde veio o agrupamento.

Aula 19 - 18/11/2008

31

Augmenting the chart with structural information

Chart[1]			
S8	<i>Verb</i> ■ <i>book</i> ■	[0,1]	Scanner
S9	<i>VP</i> ■ <i>Verb</i> ■	[0,1]	Completer
S10	<i>S</i> ■ <i>VP</i> ■	[0,1]	Completer
S11	<i>VP</i> ■ <i>Verb</i> ■ <i>NP</i>	[0,1]	Completer
S12	<i>NP</i> ■ ■ <i>Det</i> <i>NOMINAL</i>	[1,1]	Predictor
S13	<i>NP</i> ■ ■ <i>Proper-Noun</i>	[1,1]	Predictor

Chart[2]			
<i>Det</i> ■ <i>that</i> ■	[1,2]	Scanner	
<i>NP</i> ■ <i>Det</i> ■ <i>NOMINAL</i>	[1,2]	Completer	
<i>NOMINAL</i> ■ ■ <i>Noun</i>	[2,2]	Predictor	
<i>NOMINAL</i> ■ ■ <i>Proper-Noun</i>	[2,2]	Predictor	

32

Construindo as árvores de parse

- Todos os parses da frase estarão na tabela.
- Temos que ler todos os ponteiros de cada S completo.
- Encontrar todos os $S \rightarrow X$. [0,N+1]
- Siga os ponteiros para trás.
- O tempo para fazer isso não será polinomial porque pode haver um número exponencial de árvores.

Aula 19 - 18/11/2008

33