

Processamento Estatístico da Linguagem Natural

Aula 18

Professora Bianca
(Sala 302 – Bloco E)

bianca@ic.uff.br

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

Aula de Hoje

- Cap. 12 – Jurafsky & Martin – Gramáticas Formais de Inglês
 - Seções 12.4, 12.5 e 12.7
- Cap. 13 – Jurafsky & Martin – Parsing Sintático
 - Seções 13.1

Treebanks

- Um “treebank” é um corpus com anotações de sintaxe.
 - Tem um papel importante para o parsing e na investigação empírica de fenômenos sintáticos.
- O projeto “Penn Treebank” produziu (semi-automaticamente) treebanks para uma variedade de corpora.
 - Brown, Switchboard, ATIS e Wall Street Journal, além de corpora em outras línguas.

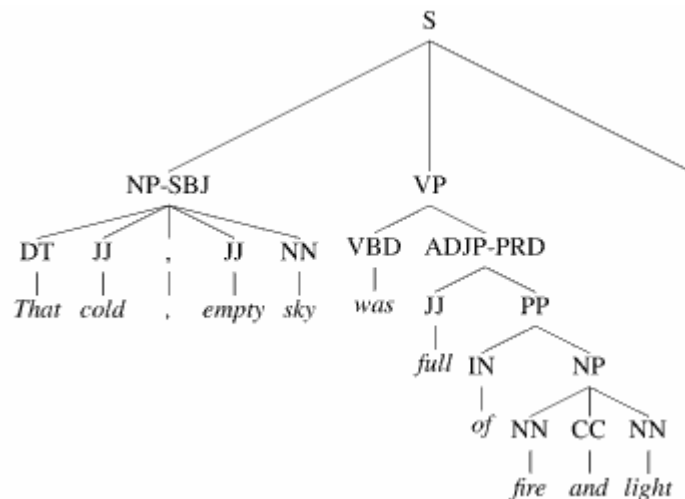
Exemplos de Treebanks

```
((S
  (NP-SBJ (DT That)
    (JJ cold) (, ,)
    (JJ empty) (NN sky) )
  (VP (VBD was)
    (ADJP-PRD (JJ full)
      (PP (IN of)
        (NP (NN fire)
          (CC and)
          (NN light) ))))
  (. .) ))
```

(a)

```
((S
  (NP-SBJ The/DT flight/NN )
  (VP should/MD
    (VP arrive/VB
      (PP-TMP at/IN
        (NP eleven/CD a.m/RB ))
      (NP-TMP tomorrow/NN )))))
```

(b)



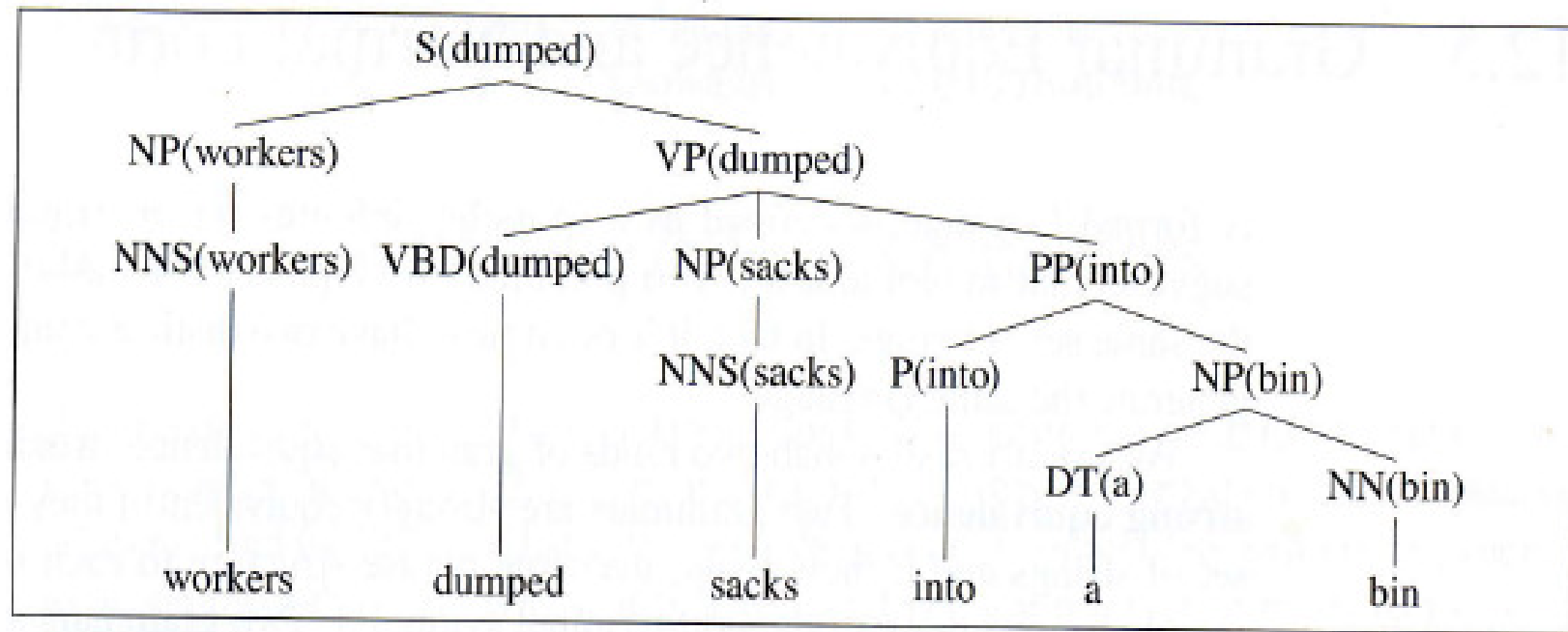
Outro exemplo

```
( (S ( ' ' ' ' )
  (S-TPC-2
    (NP-SBJ-1 (PRP We) )
    (VP (MD would)
      (VP (VB have)
        (S
          (NP-SBJ (-NONE- *-1) )
          (VP (TO to)
            (VP (VB wait)
              (SBAR-TMP (IN until)
                (S
                  (NP-SBJ (PRP we) )
                  (VP (VBP have)
                    (VP (VBN collected)
                      (PP-CLR (IN on)
                        (NP (DT those)(NNS assets))))))))))))))
    ( , , ) ( ' ' ' ' )
    (NP-SBJ (PRP he) )
    (VP (VBD said)
      (S (-NONE- *T*-2) ))
    ( . . ) ))
```

Cabeças Léxicas

- Cada agrupamento pode ser associado com uma cabeça léxica.
 - A cabeça é a palavra ou conjunto de palavras gramaticalmente mais importante.
 - N é a cabeça de um NP
 - V é a cabeça de um VP.
- Cabeças são úteis para alguns tipos de parsing.
 - Parsing lexicalizado.
- Treebanks atuais usam um conjunto de regras fixas para determinar as cabeças.

Exemplo: Cabeças Léxicas



Equivalência de Gramáticas e Forma Normal de Chomsky

- Duas gramáticas são equivalentes
 - De maneira fraca: se elas geram as mesmas seqüências de palavras.
 - De maneira forte: se elas geram as mesmas seqüências de palavras, com as mesmas estruturas.
- Alguns algoritmos de parsing necessitam que a gramática esteja em uma forma normal.
 - Exemplo: CNF = Forma Normal de Chomsky

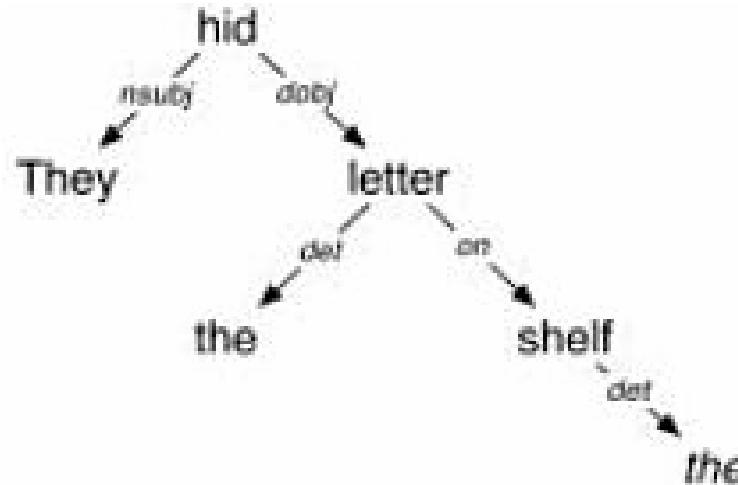
Forma Normal de Chomsky

- Uma gramática está em CNF se todas regras são da forma $A \rightarrow BC$ ou $A \rightarrow a$.
- Gramáticas em CNF são gramáticas de ramificação binária, isto é, elas geram árvores sintáticas binárias (até o nível pré-léxico).
- Toda gramática pode ser convertida em CNF.

Gramáticas de Dependência

- Tipo de formalismo diferente das gramáticas livres de contexto.
 - Agrupamentos e regras sobre a estrutura da frase não têm um papel fundamental.
 - Ao invés disso, a estrutura sintática é descrita puramente em termos das palavras e relações binárias sintáticas ou semânticas entre elas.

Gramáticas de Dependência



Argument Dependencies	Description
nsubj	nominal subject
csbj	clausal subject
dobj	direct object
iobj	indirect object
pobj	object of preposition
Modifier Dependencies	Description
tmod	temporal modifier
appos	appositional modifier
det	determiner
prep	prepositional modifier

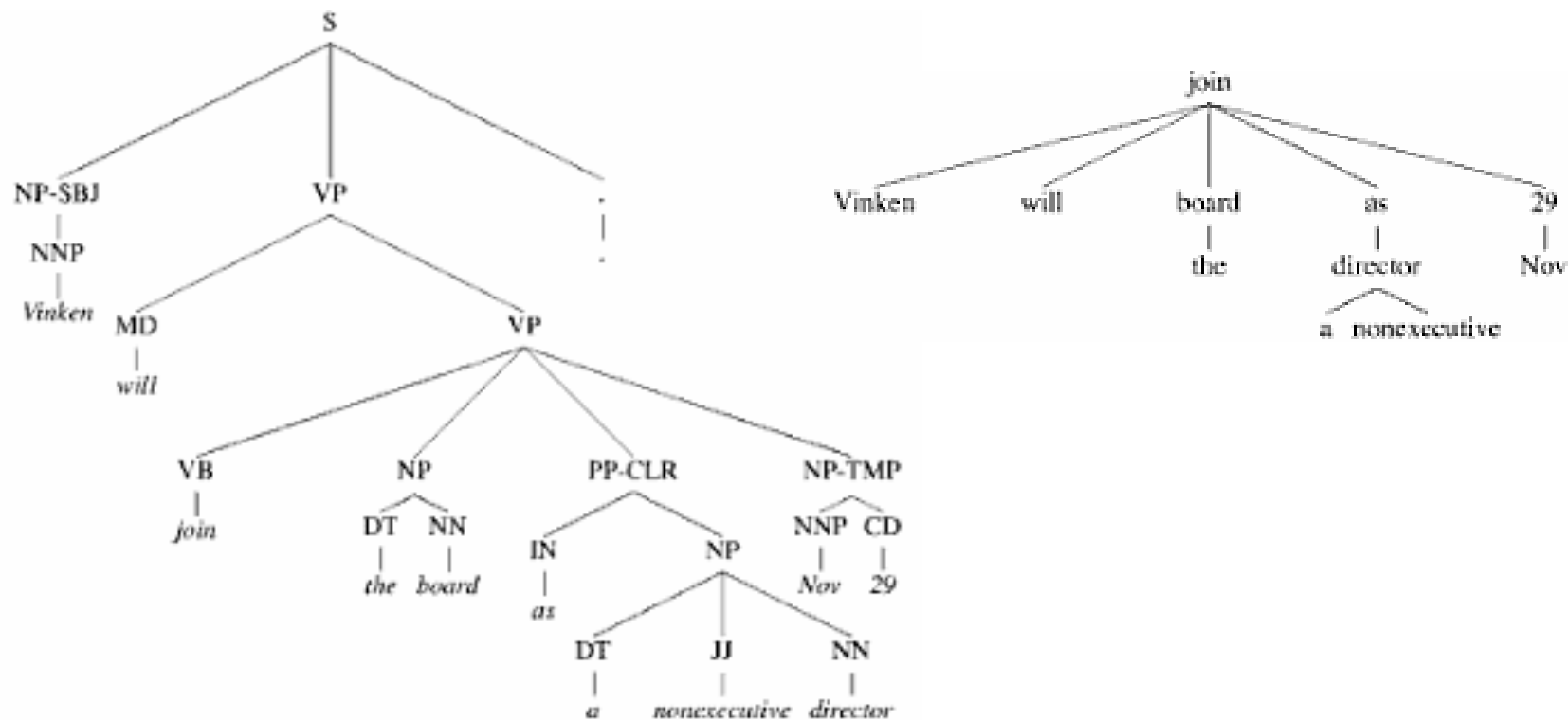
Gramáticas de Dependência

- Vantagens:
 - Poder preditivo que as palavras exibem em relação aos seus dependentes, o que facilita o parsing.
 - Saber a identidade do verbo ajuda a decidir que substantivo é o sujeito e qual é o objeto.
 - Capacidade de lidar com línguas que tenham uma liberdade maior na ordem das palavras.
 - Uma CFG teria que listar todas as possibilidades.

Relações entre dependências e cabeças

- Um grafo de dependência não-rotulado pode ser automaticamente derivado de um parse livre-de-contexto.
- Algoritmo:
 - Marque a cabeça de cada nó da árvore do parse.
 - Na estrutura de dependência, faça a cabeça de cada filho (que não seja cabeça) depender da cabeça de seu pai.

Exemplo



Parsing Sintático

- Parsing sintático = tarefa de reconhecer uma frase e atribuir uma estrutura sintática a ela.
- Capítulo 13: três algoritmos baseados em gramáticas livre-de-contexto para fazer o parsing.
 - Cocke-Kasami-Younger (CKY)
 - Earley
 - Chart Parsing

Parsing como Busca

- O parsing pode ser visto como uma busca no espaço de árvores de parse para uma determinada frase.
 - O espaço de busca é restringido pela gramática e pelas palavras da frase.
- Estratégias de busca:
 - Top-down
 - Bottom-up

Gramática L_1

Grammar	Lexicon
$S \rightarrow NP VP$	$Det \rightarrow that \mid this \mid a$
$S \rightarrow Aux NP VP$	$Noun \rightarrow book \mid flight \mid meal \mid money$
$S \rightarrow VP$	$Verb \rightarrow book \mid include \mid prefer$
$NP \rightarrow Pronoun$	$Pronoun \rightarrow I \mid she \mid me$
$NP \rightarrow Proper-Noun$	$Proper-Noun \rightarrow Houston \mid NWA$
$NP \rightarrow Det Nominal$	$Aux \rightarrow does$
$Nominal \rightarrow Noun$	$Preposition \rightarrow from \mid to \mid on \mid near \mid through$
$Nominal \rightarrow Nominal Noun$	
$Nominal \rightarrow Nominal PP$	
$VP \rightarrow Verb$	
$VP \rightarrow Verb NP$	
$VP \rightarrow Verb NP PP$	
$VP \rightarrow Verb PP$	
$VP \rightarrow VP PP$	
$PP \rightarrow Preposition NP$	

Figure 13.1 The $\mathcal{L}_1$ miniature English grammar and lexicon.

Exemplo: Árvore de Parse

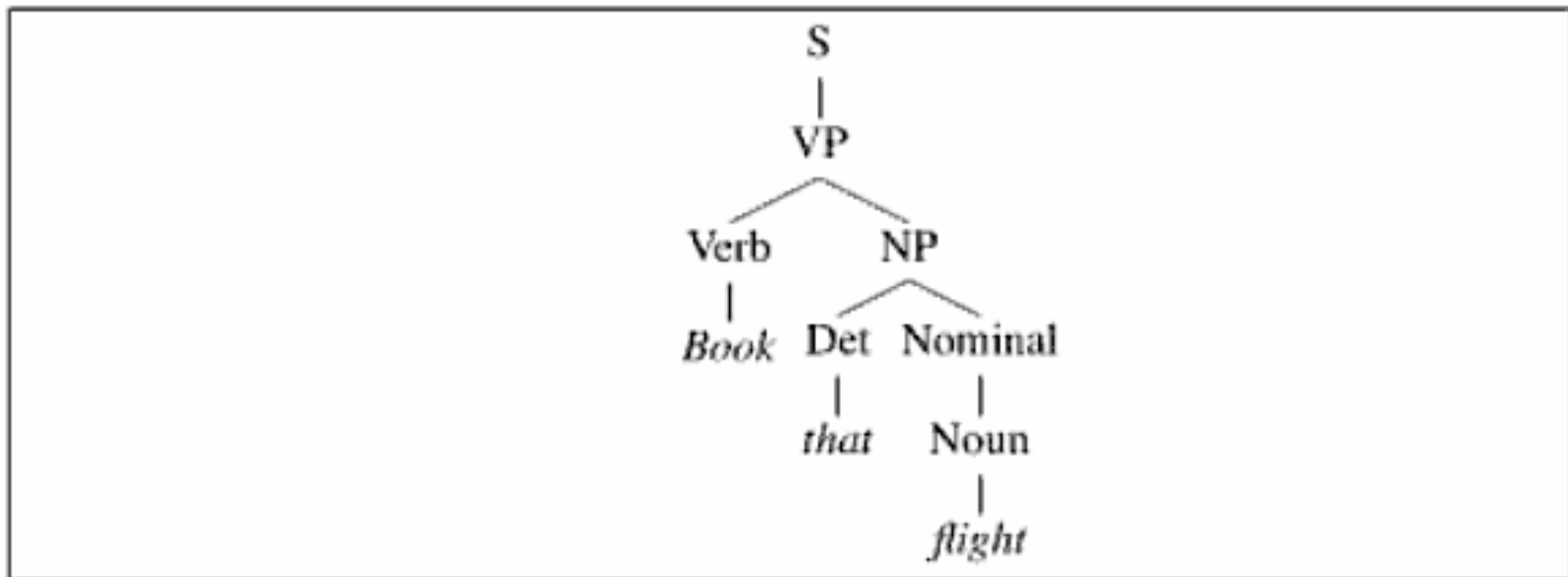
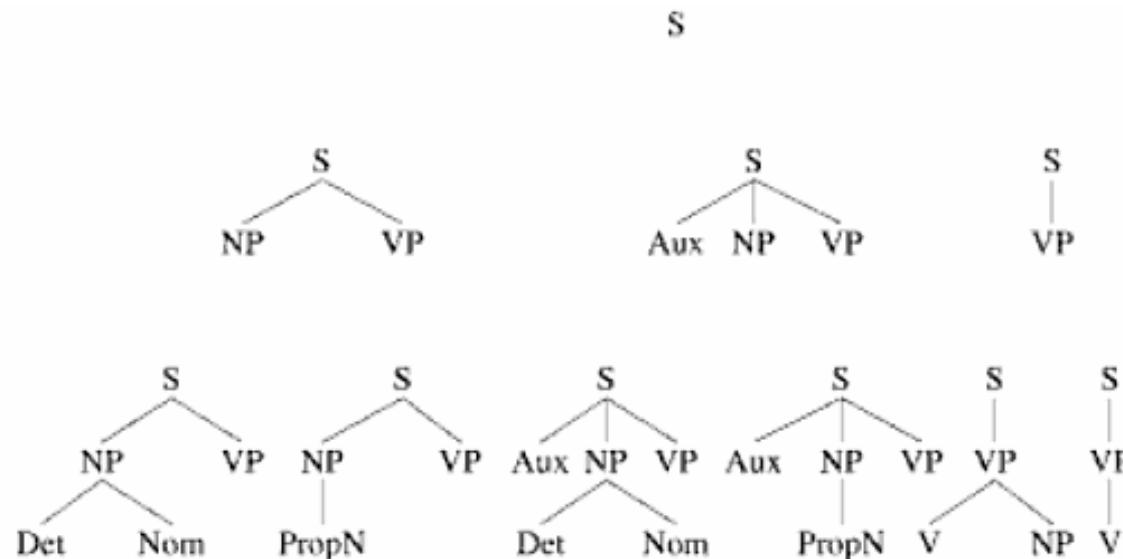


Figure 13.2 The parse tree for the sentence *Book that flight* according to grammar $\mathcal{L}_1$.

Parsing Top-Down

- Começa a partir do nó raiz S e desce em direção às folhas.
- Tenta todas as possibilidades até chegar às folhas.
 - Mantém a árvore que tiver as folhas de acordo com a frase em questão.



Parsing Bottom-Up

- Começa com as palavras da frase e tenta construir a árvore a partir das palavras subindo em direção à raiz.
- O parse tem sucesso se conseguimos chegar à raiz S.

Book that flight

Noun Det Noun Verb Det Noun
| | | | | |
Book that flight Book that flight

Nominal Nominal Nominal
| | |
Noun Det Noun Verb Det Noun
| | | | | |
Book that flight Book that flight

NP NP VP Nominal NP NP
/ \ / \ | | / \ / \ / \ / \
Nominal Noun Det Nominal Verb Det Noun Verb Det Noun
| | | | | | | | | | | |
Noun Book that Noun Book that flight Book that flight
| |
Book flight

VP NP VP NP
| / \ | / \ | / \ | / \
Verb Det Nominal Verb Det Nominal
| | | | | | | | |
Book that flight Book that flight Book that flight

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.