

Processamento Estatístico da Linguagem Natural

Aula 17

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

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

Aula 17 - 04/11/2008

1

Aula de Hoje

- Cap. 12 – Jurafsky & Martin – Gramáticas Formais de Inglês
 - Seções 12.1, 12.2 e 12.3.1

Aula 17 - 04/11/2008

2

Sintaxe

- Do grego “syntaxis” = colocar junto, dispôr em ordem.
- Refere-se à maneira com a qual as palavras são dispostas no texto e as relações entre elas.
- Distinção:
 - Gramática prescritiva: como as pessoas devem falar.
 - Gramática descritiva: como as pessoas falam.
- O objetivo da sintaxe é modelar o conhecimento inconsciente que as pessoas têm da gramática da sua língua materna.

Aula 17 - 04/11/2008

3

Sintaxe

- Análise sintática é parte fundamental de sistemas de NLP como:
 - Correção gramatical de textos
 - Extração de informação
 - Tradução automática
 - Respostas a perguntas em linguagem natural

Aula 17 - 04/11/2008

4

Três idéias principais

- Agrupamento (“*constituency*”)
 - Grupos de palavras formam um elemento único.
- Relações gramaticais.
 - Formalização de idéias como “sujeito” e “predicado”.
- Subcategorizações e relações de dependência.
 - Relações entre palavras e expressões específicas.
 - Exemplo: regência verbal

Aula 17 - 04/11/2008

5

Agrupamento (*Constituency*)

- Locuções nominais = *Noun phrases* (NPs)
 - Three parties from Brooklyn
 - A high-class spot such as Mindy's
 - The Broadway coppers
 - They
 - Harry the Horse
 - The reason he comes into the Hot Box
- Como sabemos que formam uma unidade?
 - Todos podem aparecer antes de um verbo:
 - Three parties from Brooklyn *arrive*...
 - A high-class spot such as Mindy's *attracts*...
 - The Broadway coppers *love*...
 - They *sit*...

Aula 17 - 04/11/2008

6

Agrupamento (*Constituency*)

- Todos podem aparecer antes de um verbo:
 - Three parties from Brooklyn **arrive**...
 - A high-class spot such as Mindy's **attracts**...
 - The Broadway coppers **love**...
 - They **sit**
- Mas palavras individuais não podem aparecer antes de verbos:
 - *from **arrive**...
 - *as **attracts**...
 - *the **is**
 - *spot **is**...
- Deve ser possível fazer generalizações como:
 - **Locuções nominais ocorrem antes de verbos**

Aula 17 - 04/11/2008

7

Agrupamento (*Constituency*)

- Podemos mudar a posição da unidade como um todo:
 - On **September 17th**, I'd like to fly from Atlanta to Denver
 - I'd like to fly on **September 17th** from Atlanta to Denver
 - I'd like to fly from Atlanta to Denver on **September 17th**.
- Mas não dos pedaços separadamente:
 - *On **September**, I'd like to fly **17th** from Atlanta to Denver
 - *On I'd like to fly **September 17th** from Atlanta to Denver

Aula 17 - 04/11/2008

8

Gramáticas Livres de Contexto *Context-Free Grammars (CFG)*

- Capturam agrupamento e ordenação
 - Ordenação:
 - Quais são as **regras** que governam a ordenação de palavras e unidades maiores na linguagem?
 - Agrupamento:
 - Como as palavras se agrupam em **unidades** e como os tipos de unidades se comportam.

Aula 17 - 04/11/2008

9

Exemplo de CFG

- $S \rightarrow NP VP$
- $NP \rightarrow Det NOMINAL$
- $NOMINAL \rightarrow Noun$
- $VP \rightarrow Verb$
- $Det \rightarrow a$
- $Noun \rightarrow flight$
- $Verb \rightarrow left$

Símbolos em **preto** = não-terminais
Símbolos em **verde** = terminais

LÉXICO – associa os terminais às suas classes gramaticais

O item à esquerda da seta é um agrupamento formado pelos itens à direita

Aula 17 - 04/11/2008

10

Exemplo de CFG

- $S \rightarrow NP VP$
 - Diz que existem unidades chamadas S, NP, e VP nessa linguagem.
 - Diz que um S pode ser constituído de NP seguido imediatamente de VP
 - Não diz que esse é o único tipo de S
 - Não diz que esse é o único lugar onde NPs e VPs podem ocorrer

Aula 17 - 04/11/2008

11

Análise e Síntese

- Como com FSAs e FSTs podemos ver essas regras como máquinas de análise ou síntese
 - Síntese
 - Geram seqüências de palavras da linguagem
 - Rejeitam seqüências de palavras que não estão na linguagem
 - Análise
 - Impõem estruturas (árvores) a seqüências da linguagem.

Aula 17 - 04/11/2008

12

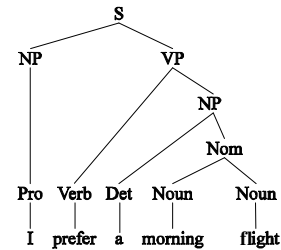
Derivações

- Uma derivação é uma seqüência de aplicações de regras que gera uma determinada seqüência de palavras que pertence à linguagem.
 - Cobre todas as palavras da seqüência.
 - Cobre somente as palavras da seqüência.

Aula 17 - 04/11/2008

13

Derivações podem ser representadas por árvores

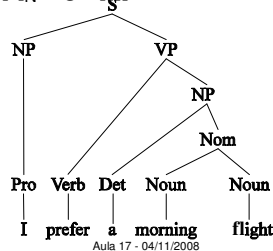


Aula 17 - 04/11/2008

14

Notação com colchetes

- $[S [NP [PRO I] [VP [V prefer [NP [NP [Det a] [Nom [N morning] [N flight]]]]]]]$



Aula 17 - 04/11/2008

15

Parsing

- Parsing é o processo de pegar uma seqüência de palavras e uma gramática e retornar uma (ou mais?) árvore(s) para aquela seqüência.

Aula 17 - 04/11/2008

16

Por que “livre de contexto”?

- Porque o não-terminal do lado esquerdo da regra fica “sozinho” (livre de contexto)

$$A \rightarrow B C$$
 Significa que podemos reescrever **A** como **um B seguido de um C** não importando o contexto em que **A** é encontrado

Aula 17 - 04/11/2008

17

Definição formal de CFGs

Σ : a set of non-terminal symbols (or variables)
 Σ_T : a set of terminal symbols (all equal from Σ)
 P : a set of rules or productions, each of the form $A \rightarrow \alpha$, where A is a non-terminal,
 α is a string of symbols from the terminal set of strings $(\Sigma \cup \Sigma_T)^*$
 S : a designated start symbol

Convenções:

Capital letters like A, B , and S	Non-terminals
S	The start symbol
Lower-case Greek letters like α, β , and γ	Strings drawn from $(\Sigma \cup N)^*$
Lower-case Roman letters like u, v , and w	Strings of terminals

Aula 17 - 04/11/2008

18

Agrupamentos Principais (Inglês)

- Frases (= *sentences*)
- Locuções nominais (= *noun phrases*)
- Locuções verbais (= *verb phrases*)
- Locuções preposicionais (= *prepositional phrases*)

Aula 17 - 04/11/2008

19

Tipos de Frases

- Declarativas: **A plane left**
S -> NP VP
- Imperativas: **Leave!**
S -> VP
- Perguntas Sim-Não: **Did the plane leave?**
S -> Aux NP VP
- Perguntas WH: **When did the plane leave?**
S -> Wh-NP Aux NP VP

Aula 17 - 04/11/2008

20

NPs

- NP → Pronoun
– I came, you saw it, they conquered
- NP → Proper-Noun
– **Los Angeles** is west of **Texas**
– **John Hennessey** is the president of **Stanford**
- NP → Det Noun
– **The president**
- NP → Nominal
- Nominal → Noun Noun
– A **morning flight** to Denver

Aula 17 - 04/11/2008

21

PPs

- PP → Preposition NP
– From LA
– To Boston
– On Tuesday
– With lunch

Aula 17 - 04/11/2008

22

Recursão

- Temos que lidar com regras como as seguintes onde o não-terminal da esquerda também aparece do lado direito.
NP → NP PP **[[The flight] [to Boston]]**
VP → VP PP **[[departed Miami] [at noon]]**

Aula 17 - 04/11/2008

23

Recursão

- Exemplo:
flights from Denver
Flights from Denver to Miami
Flights from Denver to Miami in February
Flights from Denver to Miami in February on a Friday
Flights from Denver to Miami in February on a Friday under \$300
Flights from Denver to Miami in February on a Friday under \$300 with lunch

Aula 17 - 04/11/2008

24

Recursão

Exemplo:

[[flights] [from Denver]]
[[[Flights] [from Denver]] [to Miami]]
[[[[Flights] [from Denver]] [to Miami]] [in February]]
[[[[[Flights] [from Denver]] [to Miami]] [in February]]
[on a Friday]]
Etc.

Aula 17 - 04/11/2008

25

Implicação de ser livre-de-contexto

- Se temos uma regra do tipo
– VP -> V NP

– A única restrição é que o elemento depois do verbo seja um NP. Não precisamos saber como esse NP é constituído.

Aula 17 - 04/11/2008

26

Exemplo

- VP -> V NP
- I hate
flights from Denver
Flights from Denver to Miami
Flights from Denver to Miami in February
Flights from Denver to Miami in February on a Friday
Flights from Denver to Miami in February on a Friday under \$300
Flights from Denver to Miami in February on a Friday under \$300
with lunch

Aula 17 - 04/11/2008

27