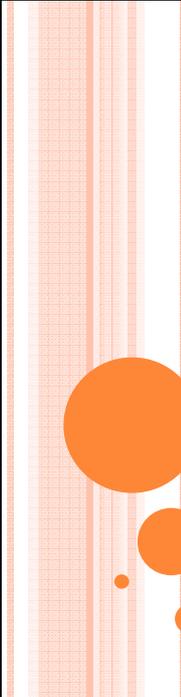


SINTAXE – PARTE 2

SCC5908 Tópicos em Processamento de Língua Natural

Thiago A. S. Pardo



PARSING

...CONTINUAÇÃO

MÉTODOS DE PARSING

○ Programação dinâmica

- Guarda em uma tabela (matriz) os constituintes já descobertos
 - Evita repetição de esforço
 - É possível recuperar todas as análises
- 2 métodos tradicionais
 - CKY (1965)
 - Earley (1970)

3

MÉTODOS DE PARSING

○ CKY: algoritmo de Cocke-Kasami-Younger

$S \rightarrow NP VP$
 $S \rightarrow XI VP$
 $XI \rightarrow Aux NP$
 $S \rightarrow book \mid include \mid prefer$
 $S \rightarrow Verb NP$
 $S \rightarrow X2 PP$
 $S \rightarrow Verb PP$
 $S \rightarrow VP PP$
 $NP \rightarrow I \mid she \mid me$
 $NP \rightarrow TWA \mid Houston$
 $NP \rightarrow Det Nominal$
 $Nominal \rightarrow book \mid flight \mid meal \mid money$
 $Nominal \rightarrow Nominal Noun$
 $Nominal \rightarrow Nominal PP$
 $VP \rightarrow book \mid include \mid prefer$
 $VP \rightarrow Verb NP$
 $VP \rightarrow X2 PP$
 $X2 \rightarrow Verb NP$
 $VP \rightarrow Verb PP$
 $VP \rightarrow VP PP$
 $PP \rightarrow Preposition NP$

+ léxico

	<i>Book</i>	<i>the</i>	<i>flight</i>	<i>through</i>	<i>Houston</i>
S, VP, Verb, Nominal, Noun	...		S, VP, X2	...	S, VP
		Det	NP	...	NP
			Nominal, Noun	...	Nominal
				Prep	PP
					NP, proper noun

MÉTODOS DE PARSING

○ Algoritmo de Earley

- Constrói um *chart* (mapa/gráfico) das derivações
- 1º passo: define um *chart* de N+1 posições, em que N é o número de palavras da sentença
 - Exemplo: *Book that flight.*

Chart



5

MÉTODOS DE PARSING

○ Algoritmo de Earley

- Constrói um mapa/gráfico das derivações
- 2º passo: em cada posição do mapa, tenta-se **prever as possíveis derivações**, representando-as em **estados de análise**
 - Cada estado é, na realidade, uma produção com a indicação do ponto da análise em que se está
 - Por exemplo

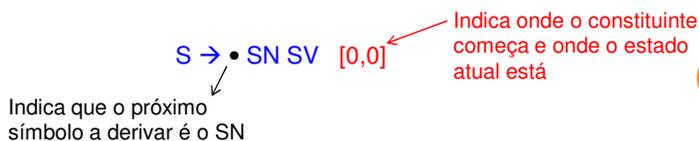
$S \rightarrow \bullet SN SV [0,0]$

6

MÉTODOS DE PARSING

○ Algoritmo de Earley

- Constrói um mapa/gráfico das derivações
- **2º passo:** em cada posição do mapa, tenta-se prever as possíveis derivações, representando-as em estados de análise
- Cada estado é, na realidade, uma produção com a indicação do ponto da análise em que se está
- Por exemplo

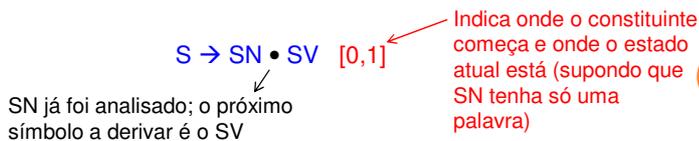


7

MÉTODOS DE PARSING

○ Algoritmo de Earley

- Constrói um mapa/gráfico das derivações
- **2º passo:** em cada posição do mapa, tenta-se prever as possíveis derivações, representando-as em estados de análise
- Cada estado é, na realidade, uma produção com a indicação do ponto da análise em que se está
- Por exemplo



8

MÉTODOS DE PARSING

○ Algoritmo de Earley

- Constrói um mapa/gráfico das derivações
- 2º passo: em cada posição do mapa, tenta-se prever as possíveis derivações, representando-as em estados de análise
- Cada estado é, na realidade, uma produção com a indicação do ponto da análise em que se está
- Por exemplo

SN começa na posição 0 da sentença, e a análise já está na posição 1

SN → determinante • substantivo [0,1]

Determinante já foi consumido; próximo símbolo a ser derivado é o substantivo

9

MÉTODOS DE PARSING

○ Algoritmo de Earley

- Constrói um mapa/gráfico das derivações
- 2º passo: em cada posição do mapa, tenta-se prever as possíveis derivações, representando-as em estados de análise
- Cada estado é, na realidade, uma produção com a indicação do ponto da análise em que se está
- Adiciona-se uma regra extra que gera o símbolo inicial da gramática
 - Quando o ponto passar para a direita do símbolo inicial, a análise termina

S' → • S [0,0]

10

MÉTODOS DE PARSING

○ Algoritmo de Earley

- Constrói um mapa/gráfico das derivações
 - 2º passo: em cada posição do mapa, tenta-se prever as possíveis derivações, representando-as em estados de análise
 - Cada estado é, na realidade, uma produção com a indicação do ponto da análise em que se está
 - Adiciona-se uma regra extra que gera o símbolo inicial da gramática
 - Quando o ponto passar para a direita do símbolo inicial, a análise termina

$S' \rightarrow S \bullet [0,10]$

Supondo que a sentença tem 10 palavras

11

EXEMPLO SIMPLES

- $S' \rightarrow S$
- $S \rightarrow SN SV$
- $SN \rightarrow$ artigo substantivo
- $SV \rightarrow$ verbo
- artigo \rightarrow o | ...
- substantivo \rightarrow menino | ...
- verbo \rightarrow chorou | ...

O menino chorou

12

EXEMPLO SIMPLES

- S' → • S
- S → SN SV
- SN → artigo substantivo
- SV → verbo
- artigo → o | ...
- substantivo → menino | ...
- verbo → chorou | ...

- O menino chorou

13

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → artigo substantivo
- SV → verbo
- artigo → o | ...
- substantivo → menino | ...
- verbo → chorou | ...

- O menino chorou

14

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → • artigo substantivo
- SV → verbo
- artigo → o | ...
- substantivo → menino | ...
- verbo → chorou | ...

• O menino chorou

15

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → • artigo substantivo
- SV → verbo
- artigo → • o | ...
- substantivo → menino | ...
- verbo → chorou | ...

• O menino chorou

16

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → • artigo substantivo
- SV → verbo
- artigo → o • | ...
- substantivo → menino | ...
- verbo → chorou | ...

O • menino chorou

17

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → artigo • substantivo
- SV → verbo
- artigo → o • | ...
- substantivo → menino | ...
- verbo → chorou | ...

O • menino chorou

18

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → artigo • substantivo
- SV → verbo
- artigo → o • | ...
- substantivo → • menino | ...
- verbo → chorou | ...

O • menino chorou

19

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → artigo • substantivo
- SV → verbo
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou | ...

O menino • chorou

20

EXEMPLO SIMPLES

- S' → • S
- S → • SN SV
- SN → artigo substantivo •
- SV → verbo
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou | ...

O menino • chorou

21

EXEMPLO SIMPLES

- S' → • S
- S → SN • SV
- SN → artigo substantivo •
- SV → verbo
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou | ...

O menino • chorou

22

EXEMPLO SIMPLES

- S' → • S
- S → SN • SV
- SN → artigo substantivo •
- SV → • verbo
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou | ...

O menino • chorou

23

EXEMPLO SIMPLES

- S' → • S
- S → SN • SV
- SN → artigo substantivo •
- SV → • verbo
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → • chorou | ...

O menino • chorou

24

EXEMPLO SIMPLES

- S' → • S
- S → SN • SV
- SN → artigo substantivo •
- SV → • verbo
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou • | ...

O menino chorou •

25

EXEMPLO SIMPLES

- S' → • S
- S → SN • SV
- SN → artigo substantivo •
- SV → verbo •
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou • | ...

O menino chorou •

26

EXEMPLO SIMPLES

- S' → • S
- S → SN SV •
- SN → artigo substantivo •
- SV → verbo •
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou • | ...

O menino chorou •

27

EXEMPLO SIMPLES

- S' → S •
- S → SN SV •
- SN → artigo substantivo •
- SV → verbo •
- artigo → o • | ...
- substantivo → menino • | ...
- verbo → chorou • | ...

O menino chorou •

28

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Inicia-se pela regra inicial artificial

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

₀ O ₁ copo ₂ quebrou ₃

Chart

	0	1	2	3
S' → • S [0,0]				

MÉTODOS DE PARSING

- Algoritmo de Earley
 - A partir da regra inicial, adicionam-se todas as próximas regras possíveis para S

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

₀ O ₁ copo ₂ quebrou ₃

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0]				

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Recursivamente, adicionam-se as outras regras necessárias para SN e SV

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]				

MÉTODOS DE PARSING

- Algoritmo de Earley
 - O artigo casa com “o”

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]		artigo → o • [0,1]		

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Todas as regras que precisavam de um artigo avançam

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]		artigo → o • [0,1] SN → art • subst [0,1]		

MÉTODOS DE PARSING

- Algoritmo de Earley
 - O substantivo casa com “copo”

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]		artigo → o • [0,1] SN → art • subst [0,1]	subst → copo • [1,2]	

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Todas as regras que precisavam de um substantivo avançam

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]		artigo → o • [0,1] SN → art • subst [0,1]	subst → copo • [1,2] SN → art subst • [0,2]	

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Todas as regras que precisavam de um SN avançam

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]		artigo → o • [0,1] SN → art • subst [0,1]	subst → copo • [1,2] SN → art subst • [0,2] S → SN • SV [0,2]	

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Adicionam-se as próximas regras de SV necessárias

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]	artigo → o • [0,1] SN → art • subst [0,1]	subst → copo • [1,2] SN → art subst • [0,2] S → SN • SV [0,2] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]		

MÉTODOS DE PARSING

- Algoritmo de Earley
 - O verbo casa com “quebrou”

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

	0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]	artigo → o • [0,1] SN → art • subst [0,1]	subst → copo • [1,2] SN → art subst • [0,2] S → SN • SV [0,2] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]	verbo → quebrou • [2,3]	

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Todas as regras que precisavam de verbo avançam

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]	artigo → o • [0,1] SN → art • subst [0,1]	subst → copo • [1,2] SN → art subst • [0,2] S → SN • SV [0,2] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]	verbo → quebrou • [2,3] SV → verbo • [2,3] SV → verbo • SN [2,3] SV → verbo • SN SP [2,3]

MÉTODOS DE PARSING

- Algoritmo de Earley
 - Todas as regras que precisavam de um SV avançam

○ REGRAS

- S' → S
- S → SN SV
- S → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

○ LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ela | ...
- substantivo → copo | ...
- verbo → quebrou | ...
- preposição → de | em | ...

0 O 1 copo 2 quebrou 3

Chart

0	1	2	3
S' → • S [0,0] S → • SN SV [0,0] S → • SV [0,0] SN → • pronome [0,0] SN → • subst [0,0] SN → • art subst [0,0] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]	artigo → o • [0,1] SN → art • subst [0,1]	subst → copo • [1,2] SN → art subst • [0,2] S → SN • SV [0,2] SV → • verbo [0,0] SV → • verbo SN [0,0] SV → • verbo SN SP [0,0]	verbo → quebrou • [2,3] SV → verbo • [2,3] SV → verbo • SN [2,3] SV → verbo • SN SP [2,3] S → SN SV • [0,3]

MÉTODOS DE PARSING

o Algoritmo de Earley

- Todas as regras que precisavam de um S avançam

→ Regra inicial completa!

→ Fim do processo

o REGRAS

- $S' \rightarrow S$
- $S \rightarrow SN SV$
- $S \rightarrow SV$
- $SN \rightarrow \text{pronome}$
- $SN \rightarrow \text{substantivo}$
- $SN \rightarrow \text{artigo substantivo}$
- $SV \rightarrow \text{verbo}$
- $SV \rightarrow \text{verbo SN}$
- $SV \rightarrow \text{verbo SN SP}$
- $SP \rightarrow \text{preposição SN}$

o LÉXICO

- artigo $\rightarrow o \mid a \mid os \mid \dots$
- pronome $\rightarrow eu \mid ela \mid \dots$
- substantivo $\rightarrow \text{copo} \mid \dots$
- verbo $\rightarrow \text{quebrou} \mid \dots$
- preposição $\rightarrow de \mid em \mid \dots$

$_0 O_1 \text{ copo}_2 \text{ quebrou}_3$

Chart

0	1	2	3
$S' \rightarrow \bullet S [0,0]$ $S \rightarrow \bullet SN SV [0,0]$ $S \rightarrow \bullet SV [0,0]$ $SN \rightarrow \bullet \text{pronome} [0,0]$ $SN \rightarrow \bullet \text{subst} [0,0]$ $SN \rightarrow \bullet \text{art subst} [0,0]$ $SV \rightarrow \bullet \text{verbo} [0,0]$ $SV \rightarrow \bullet \text{verbo SN} [0,0]$ $SV \rightarrow \bullet \text{verbo SN SP} [0,0]$	$\text{artigo} \rightarrow o \bullet [0,1]$ $SN \rightarrow \text{art} \bullet \text{subst} [0,1]$	$\text{subst} \rightarrow \text{copo} \bullet [1,2]$ $SN \rightarrow \text{art subst} \bullet [0,2]$ $S \rightarrow SN \bullet SV [0,2]$ $SV \rightarrow \bullet \text{verbo} [0,0]$ $SV \rightarrow \bullet \text{verbo SN} [0,0]$ $SV \rightarrow \bullet \text{verbo SN SP} [0,0]$	$\text{verbo} \rightarrow \text{quebrou} \bullet [2,3]$ $SV \rightarrow \text{verbo} \bullet [2,3]$ $SV \rightarrow \text{verbo} \bullet SN [2,3]$ $SV \rightarrow \text{verbo} \bullet SN SP [2,3]$ $S \rightarrow SN SV \bullet [0,3]$ $S' \rightarrow S \bullet [0,3]$

MÉTODOS DE PARSING

o Algoritmo de Earley

- O processo tem sucesso se consome a regra inicial
- A partir do chart, é possível recuperar todas as árvores sintáticas possíveis
 - o Cada constituinte consumido pode armazenar junto de si os filhos que contribuíram com ele
- Não é necessário remodelar a gramática (como acontecia com o CKY)
- Estilo top-down de análise

EXERCÍCIO

- Aplicar o algoritmo de Earley para a sentença em inglês *Book that flight*, com a gramática abaixo

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

43

PARSING PARCIAL

- Também chamado *shallow parsing*
 - Não se produzem árvores sintáticas completas
 - *Chunking*
 - Identificam-se os sintagmas que formam as sentenças
 - Grande variação: com ou sem recursão (mais comum)

[O v \hat{o} de S \hat{a} o Paulo]_{SN} [chegou]_{SV}
 [O v \hat{o} [de [S \hat{a} o Paulo]_{SN}]_{SP}]_{SN} [chegou]_{SV}
 [O v \hat{o}]_{SN} [de]_{SP} [S \hat{a} o Paulo]_{SN} [chegou]_{SV}

44

PARSING PARCIAL

- Também chamado *shallow parsing*
 - Não se produzem árvores sintáticas completas
 - *Chunking*
 - Identificam-se os sintagmas que formam as sentenças
 - Apenas alguns tipos de sintagmas

[O vôo]_{SN} de [São Paulo]_{SN} chegou

45

PARSING PARCIAL

- **Abordagens**
 - **Regras**
 - Em geral, aplicadas da esquerda para a direita, das maiores para as menores
 - Não garantem solução ótima
 - Exemplo
 - SN → artigo substantivo adjetivo
 - SN → artigo substantivo
 - SN → substantivo
 - SV → verbo_aux verbo
 - SV → verbo

46

PARSING PARCIAL

○ Abordagens

- **Aprendizado de máquina**

- Classificação seqüencial, da esquerda para a direita
- Exige treinamento: portanto, córpus anotado ou treebank
- Atributos (com janela de 2 palavras, normalmente)
 - Palavra a ser classificada, as duas palavras anteriores e as duas posteriores, as etiquetas morfossintáticas dessas palavras, *chunk* anterior

47

PARSING PARCIAL

○ Abordagens

- **Aprendizado de máquina**

- **Atenção:** podem-se aprender regras também

48

PARSING PARCIAL

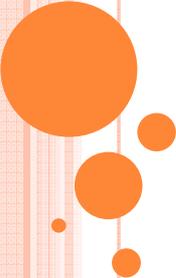
○ Esquema de anotação

- Esquema **IOB** para marcação de córpus (também usado para outros fins, por exemplo, *tagging*)
 - B = *Beginning*
 - I = *Internal*
 - O = *Outside*
- Exemplo

[O longo vôo]_{SN} de [São Paulo]_{SN} chegou

O	longo	vôo	de	São	Paulo	chegou
B_SN	I_SN	I_SN	O	B_SN	I_SN	O

49



PARSING PROBABILÍSTICO

ESTATÍSTICA

- Métodos anteriores são eficientes, mas não têm mecanismos para **escolher uma das possíveis análises sintáticas**
- Estatística pode ajudar a resolver isso
 - Ambigüidades, por exemplo, coordenações e ligação do SP
 - Modelagem lingüística
- Gramáticas livres de contexto probabilísticas (GLCP)

51

EXEMPLO DE GLCP

- REGRAS
 - Sentença → SN SV [0.80]
 - Sentença → SV [0.20]
 - SN → pronome [0.50]
 - SN → substantivo [0.15]
 - SN → artigo substantivo [0.35]
 - SV → verbo [0.40]
 - SV → verbo SN [0.40]
 - SV → verbo SN SP [0.20]
 - SP → preposição SN [1.00]
- LÉXICO
 - artigo → o [0.20] | a [0.20] | os [0.15] | ...
 - Etc.

52

GLCP

- Formalmente definida como uma quádrupla
 - Símbolos não terminais N
 - Símbolos terminais T
 - Conjunto de regras R da forma $\alpha \rightarrow \beta$ [p], em que
 - α pertence a N
 - β pertence a $(N \cup T)^*$
 - p é a probabilidade condicional entre 0 e 1 de se ter $P(\beta | \alpha)$
 - Probabilidade de β ser gerado por α
 - Probabilidade do Lado Direito da Regra (LDR) ser gerado pelo Lado Esquerdo da Regra (LER)
 - $P(\alpha \rightarrow \beta)$
 - $P(\alpha \rightarrow \beta | \alpha)$
 - $P(\text{LDR} | \text{LER})$
- S é o símbolo inicial da gramática

53

GLCP

- Formalmente definida como uma quádrupla
 - Símbolos não terminais N
 - Símbolos terminais T
 - Conjunto de regras R da forma $\alpha \rightarrow \beta$ [p], em que
 - α pertence a N
 - β pertence a $(N \cup T)^*$
 - p é a probabilidade condicional entre 0 e 1 de se ter $P(\beta | \alpha)$
 - Probabilidade de β ser gerado por α
 - Probabilidade do Lado Direito da Regra (LDR) ser gerado pelo Lado Esquerdo da Regra (LER)
 - $P(\alpha \rightarrow \beta)$
 - $P(\alpha \rightarrow \beta | \alpha)$
 - $P(\text{LDR} | \text{LER})$
 - S é o símbolo inicial da gramática

54

GLCP

- Formalmente definida como uma quádrupla
 - Símbolos não terminais N
 - Símbolos terminais T
 - Conjunto de regras R da forma $\alpha \rightarrow \beta$ [p], em que
 - α pertence a N
 - β pertence a $(N \cup T)^*$
 - p é a probabilidade condicional entre 0 e 1 de se ter $P(\beta | \alpha)$
 - Probabilidade de β ser gerado por α
 - Probabilidade do Lado Direito da Regra (LDR) ser gerado pelo Lado Esquerdo da Regra (LER)
 - $P(\alpha \rightarrow \beta)$
 - $P(\alpha \rightarrow \beta | \alpha)$
 - $P(\text{LDR} | \text{LER})$
- $$\sum_{\beta} P(\alpha \rightarrow \beta) = 1$$
- S é o símbolo inicial da gramática

55

GLCP

- Formalmente definida como uma quádrupla
 - Símbolos não terminais N
 - Símbolos terminais T
 - Conjunto de regras R da forma $\alpha \rightarrow \beta$ [p], em que
 - α pertence a N
 - β pertence a $(N \cup T)^*$
 - p é a probabilidade condicional entre 0 e 1 de se ter $P(\beta | \alpha)$
 - Probabilidade de β ser gerado por α
 - Probabilidade do Lado Direito da Regra (LDR) ser gerado pelo Lado Esquerdo da Regra (LER)
 - $P(\alpha \rightarrow \beta)$
 - $P(\alpha \rightarrow \beta | \alpha)$
 - $P(\text{LDR} | \text{LER})$
$$\sum_{\beta} P(\alpha \rightarrow \beta) = 1$$
 - S é o símbolo inicial da gramática

56

GLCP

- Como usar a gramática para computar a probabilidade de uma árvore?
 - Consideram-se as probabilidades de cada “pedaço” da árvore, ou seja, de cada regra usada na árvore

57

GLCP

- Como usar a gramática para computar a probabilidade de uma árvore?

$$P(\text{sentença}, \text{árvore}) = \prod_{i=1}^n P(LDR_i \mid LER_i)$$

58

GLCP

- Como usar a gramática para computar a probabilidade de uma árvore?

$$P(\text{sentença}, \text{árvore}) = \prod_{i=1}^n P(\text{LDR}_i \mid \text{LER}_i)$$

- Além de ser a probabilidade conjunta da sentença e da árvore, também é a probabilidade da árvore

$$P(\text{sentença}, \text{árvore}) = P(\text{árvore}) \times P(\text{sentença} \mid \text{árvore})$$

$$P(\text{sentença}, \text{árvore}) = P(\text{árvore}) \times 1$$

$$P(\text{sentença}, \text{árvore}) = P(\text{árvore})$$

59

GLCP

- Como usar a gramática para computar a probabilidade de uma árvore?

$$P(\text{sentença}, \text{árvore}) = \prod_{i=1}^n P(\text{LDR}_i \mid \text{LER}_i)$$

- Além de ser a probabilidade conjunta da sentença e da árvore, também é a probabilidade da árvore

$$P(\text{sentença}, \text{árvore}) = P(\text{árvore}) \times P(\text{sentença} \mid \text{árvore})$$

$$P(\text{sentença}, \text{árvore}) = P(\text{árvore}) \times 1$$

$$P(\text{sentença}, \text{árvore}) = P(\text{árvore})$$

Como é possível?

60

GLCP: EXEMPLO

Qual a correta?

O que significam?

Rules			Rules		
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	
Verb → book	.30		Nominal → Noun	.75	
Det → the	.60		Verb → book	.30	
Noun → dinner	.10		Det → the	.60	
Noun → flights	.40		Noun → dinner	.10	
			Noun → flights	.40	

GLCP: EXEMPLO

$$P(\text{esq}) = 0.05 * 0.2 * 0.2 * 0.2 * 0.75 * 0.3 * 0.6 * 0.1 * 0.4 = 2.2 * 10^{-6}$$

$$P(\text{dir}) = 0.05 * 0.1 * 0.2 * 0.15 * 0.75 * 0.75 * 0.3 * 0.6 * 0.1 * 0.4 = 6.1 * 10^{-7}$$

Rules			Rules		
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	
Verb → book	.30		Nominal → Noun	.75	
Det → the	.60		Verb → book	.30	
Noun → dinner	.10		Det → the	.60	
Noun → flights	.40		Noun → dinner	.10	
			Noun → flights	.40	

PARSING PROBABILÍSTICO

- É simples estender os métodos CKY ou de Earley para considerar probabilidades
 - Podem-se guardar todas ou somente as melhores análises

63

PARSING PROBABILÍSTICO

- É simples estender os métodos CKY ou de Earley para considerar probabilidades
 - Podem-se guardar todas ou somente as melhores análises

Trecho de uma gramática

S → NS VP	[0.8]
NP → Det N	[0.3]
VP → V NP	[0.2]
V → includes	[0.05]
Det → the	[0.4]
Det → a	[0.4]
N → meal	[0.01]
N → flight	[0.02]

	<i>The</i>	<i>flight</i>	<i>includes</i>	<i>a</i>	<i>meal</i>
Det: 0.4	NP: 0.3 * 0.4 * 0.02 = 0.0024				
	N: 0.02	...			
			V: 0.05		
				...	

PARSING PROBABILÍSTICO

○ Lindo! Mas...

- **Exercício em duplas:** como conseguir as probabilidades das regras da gramática?

65

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 1: há um **treebank**

$$P(\alpha \rightarrow \beta | \alpha) = \frac{\text{Número}(\alpha \rightarrow \beta)}{\text{Número}(\alpha)}$$

- Exemplo hipotético

$$P(SV \rightarrow V | SV) = \frac{\text{Número}(SV \rightarrow V)}{\text{Número}(SV)} = \frac{5}{10} = 50\%$$

66

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: **não há** um treebank
 - Geram-se todas as árvores sintáticas das sentenças com um parser disponível (não probabilístico), assumindo-se que todas as regras têm igual probabilidade

Parser convencional

S → SN SV	SN → art subst
S → SV	SN → subst
SV → verbo	SN → pronome

67

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: **não há** um treebank
 - Geram-se todas as árvores sintáticas das sentenças com um parser disponível (não probabilístico), assumindo-se que todas as regras têm igual probabilidade

Parser convencional estendido → **prob. uniformes**

S → SN SV	[0.50]	SN → art subst	[0.33]
S → SV	[0.50]	SN → subst	[0.33]
SV → verbo	[1.00]	SN → pronome	[0.33]

68

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: não há um treebank
 - Geram-se todas as árvores sintáticas das sentenças com um parser disponível (não probabilístico), assumindo-se que todas as regras têm igual probabilidade

Parser convencional estendido → prob. uniformes

S → SN SV	[0.50]	SN → art subst	[0.33]
S → SV	[0.50]	SN → subst	[0.33]
SV → verbo	[1.00]	SN → pronome	[0.33]

Córpus

Ele morreu.
A menina chorou.
Ela gritou.

69

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: não há um treebank
 - Geram-se todas as árvores sintáticas das sentenças com um parser disponível (não probabilístico), assumindo-se que todas as regras têm igual probabilidade

Parser convencional estendido → prob. uniformes

S → SN SV	[0.50]	SN → art subst	[0.33]
S → SV	[0.50]	SN → subst	[0.33]
SV → verbo	[1.00]	SN → pronome	[0.33]

Córpus

Ele morreu.
A menina chorou.
Ela gritou.



Córpus anotado

[[Ele_{PRONOME}]SN [morreu_{VERBO}]SV]S → 0.5*0.33*1=0.165
 [[A_{ART} menina_{SUBST}]SN [chorou_{VERBO}]SV]S → 0.5*0.33*1=0.165
 [[Ela_{PRONOME}]SN [gritou_{VERBO}]SV]S → 0.5*0.33*1=0.165

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: não há um treebank
 - Estimam-se novas probabilidades para as regras
 - Prob(regra) = soma das prob. das árvores em que ocorreram
 - Normalização posterior

Parser convencional com **novas probabilidades**

S → SN SV [0.165*3]	SN → art subst [0.165]
S → SV [0]	SN → subst [0]
SV → verbo [0.165*3]	SN → pronome [0.165*2]

Cópus

Ele morreu.
A menina chorou.
Ela gritou.



Cópus anotado

[[Ele _{PRONOME}] _{SN} [morreu _{VERBO}] _{SV}] _S → 0.5*0.33*1=0.165
[[A _{ART} menina _{SUBST}] _{SN} [chorou _{VERBO}] _{SV}] _S → 0.5*0.33*1=0.165
[[Ela _{PRONOME}] _{SN} [gritou _{VERBO}] _{SV}] _S → 0.5*0.33*1=0.165

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: não há um treebank
 - Estimam-se novas probabilidades para as regras
 - Prob(regra) = soma das prob. das árvores em que ocorreram
 - Normalização posterior

Parser convencional com **novas probabilidades**

S → SN SV [1.00]	SN → art subst [0.33]
S → SV [0]	SN → subst [0]
SV → verbo [1.00]	SN → pronome [0.66]

Cópus

Ele morreu.
A menina chorou.
Ela gritou.



Cópus anotado

[[Ele _{PRONOME}] _{SN} [morreu _{VERBO}] _{SV}] _S → 0.5*0.33*1=0.165
[[A _{ART} menina _{SUBST}] _{SN} [chorou _{VERBO}] _{SV}] _S → 0.5*0.33*1=0.165
[[Ela _{PRONOME}] _{SN} [gritou _{VERBO}] _{SV}] _S → 0.5*0.33*1=0.165

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: não há um treebank
 - Repete-se o processo até os números convergirem
 - Geram-se árvores sintáticas com novas probabilidades
 - Estimam-se novas probabilidades para as regras
 - Método conhecido como *Expectation-Maximization* (EM)
 1. Tudo começa igual, com a mesma probabilidade
 2. Estimam-se probabilidades dos dados reais
 3. Maximizam-se parâmetros/probabilidades
 4. Se não houve mudança nos números, para-se; caso contrário, volta-se ao passo 2

73

PARSING PROBABILÍSTICO

○ Aprendizado de probabilidades

- Alternativa 2: não há um treebank
 - **Atenção:** se parser convencional gerasse uma única árvore para cada sentença, as contas seriam tão simples quanto na alternativa 1

74

GLCP: problemas

○ 2 principais limitações

- Suposições fracas de independência
- Falta de informação lexical

75

GLCP: problemas

○ 2 principais limitações

- Suposições fracas de independência
 - A probabilidade de uma regra independe de onde ela é usada
 - SN → art subst [0.28]
 - SN → pronome [0.25]
 - Sabe-se que **isso não é verdade**
 - **Pronomes** são muito mais prováveis de acontecerem como **sujeito** → recuperam o tópico ou a informação antiga
 - Sintagmas nominais não pronominais são mais prováveis como objeto → introduzem informação nova

76

GLCP: problemas

○ 2 principais limitações

- Suposições fracas de independência

- Estudo para o inglês (Francis et al., 1999)

	Pronome	Não pronome
Sujeito	91%	9%
Objeto	34%	66%

- Para representar tal fenômeno, faz-se necessário ter a informação do pai do elemento sendo expandido

77

GLCP: problemas

○ 2 principais limitações

- Suposições fracas de independência

- Solução possível: dividir as regras

- $SN_{SUJEITO} \rightarrow \text{pronome}$ [0.91]

- $SN_{OBJETO} \rightarrow \text{pronome}$ [0.34]

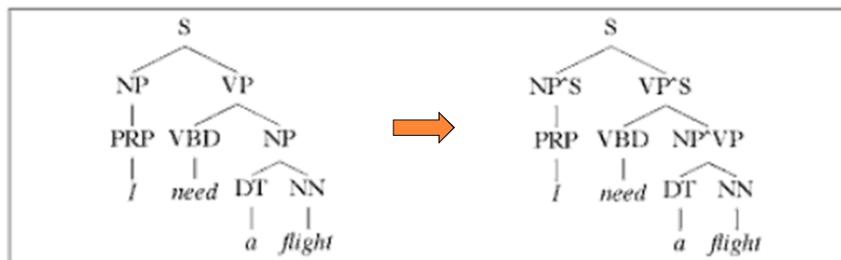
- Forma de implementação: anexar a cada símbolo o símbolo de seu nó pai $\rightarrow \text{nó_filho}^{\text{nó_pai}}$

78

GLCP: problemas

- 2 principais limitações

- Suposições fracas de independência



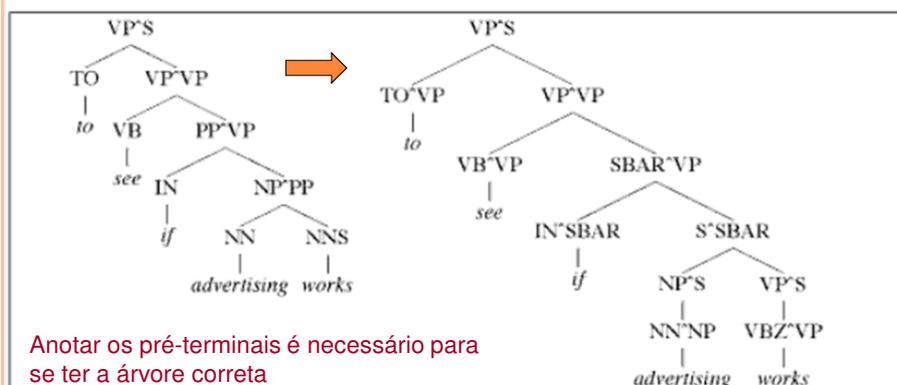
Sem anotar os pré-terminais (etiquetas morfossintáticas)

79

GLCP: problemas

- 2 principais limitações

- Suposições fracas de independência



Anotar os pré-terminais é necessário para se ter a árvore correta

GLCP: problemas

- 2 principais limitações

- Suposições fracas de independência

- Anotar os pré-terminais permite representar mais fenômenos

- Por exemplo, *SVs* são comuns com o **advérbio**^{SV} *não* e **SNs** são comuns com os **advérbios**^{SN} *apenas* e *somente*

81

GLCP: problemas

- 2 principais limitações

- Suposições fracas de independência

- Problemas dessa abordagem?

82

GLCP: problemas

○ 2 principais limitações

- Suposições fracas de independência

○ Problemas dessa abordagem?

- Aumento do tamanho da gramática
- Dados mais esparsos

→ há procedimentos automáticos para se achar o nível ótimo de anotação

83

GLCP: problemas

○ 2 principais limitações

- Falta de informação lexical

- Informação lexical é determinante para se decidir onde ligar sintagmas preposicionais

Workers **dumped** sacks **into** a bin.



VS.

Workers dumped **sacks** **into** a bin.



MAIS PROVÁVEL: *dumped* e *into* têm mais afinidade do que *sacks* e *into*

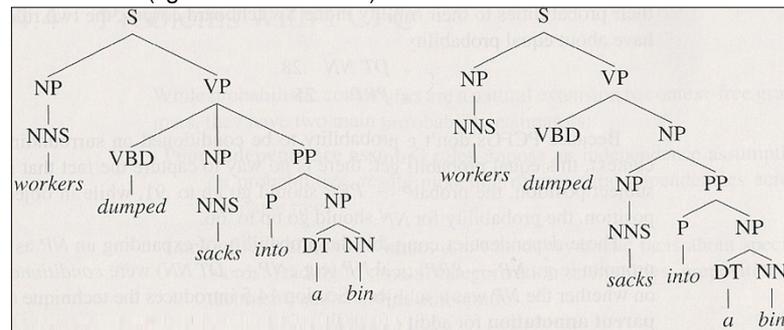
84

GLCP: problemas

- 2 principais limitações

- Falta de informação lexical

Alternativas (ligado ao VP vs. NP)



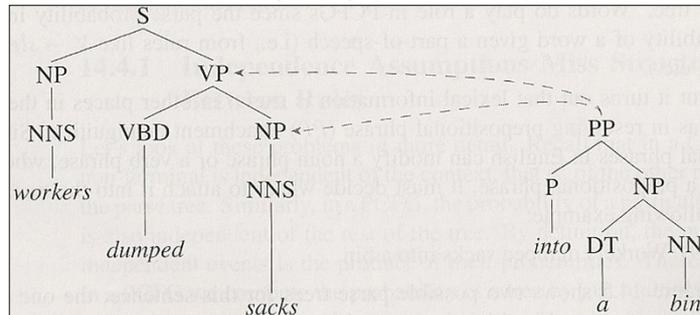
85

GLCP: problemas

- 2 principais limitações

- Falta de informação lexical

Alternativas (mesmas regras, ligações diferentes)



86

GLCP: problemas

○ 2 principais limitações

- Falta de informação lexical
 - Informação lexical é determinante para se decidir onde ligar sintagmas preposicionais

Fishermen caught **tons of** herring.



VS.

Fishermen **caught** tons of herring.



MAIS PROVÁVEL: *tons* e *of* têm mais afinidade do que *caught* e *of*

87

GLCP: problemas

○ 2 principais limitações

- Falta de informação lexical
 - Informação lexical é determinante resolver coordenações

dogs in houses and cats

- [*dogs in houses*] and [*cats*]

- *dogs in [houses and cats]*

MAIS PROVÁVEL: *dogs* e *cats* são mais afins... e *dogs* não cabem dentro de *cats*

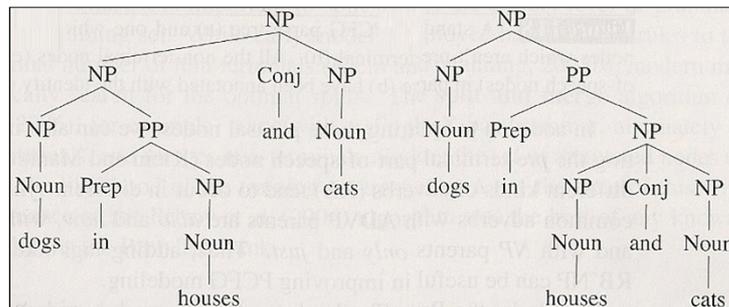
88

GLCP: problemas

- 2 principais limitações

- Falta de informação lexical

Alternativas



89

GLCP: problemas

- 2 principais limitações

- Falta de informação lexical
 - É necessário estender as GLCPs para lidar com dependências lexicais

90

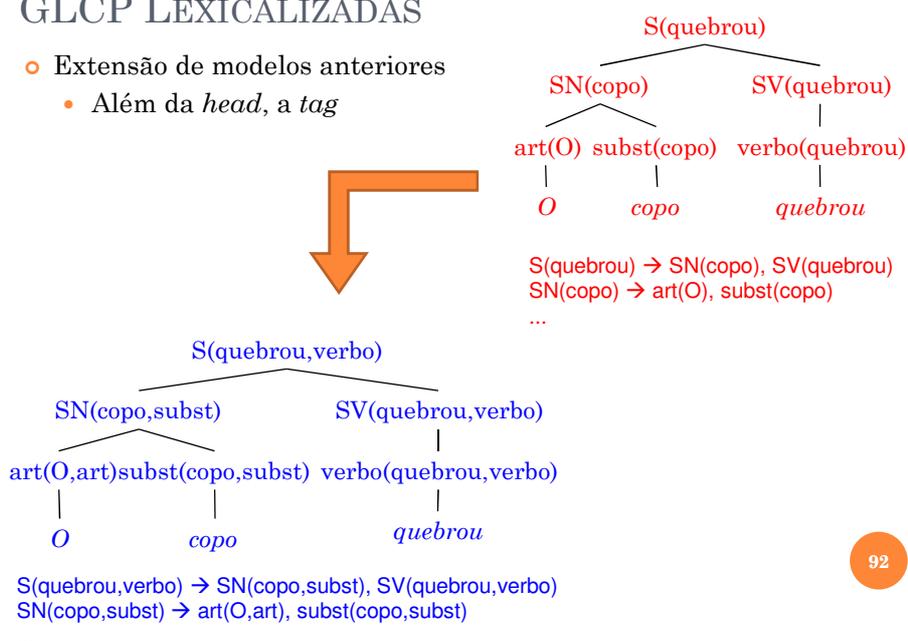
GLCP LEXICALIZADAS

- Modelos mais utilizados hoje
 - Parsers de Collins (1999) e Charniak (1997)
- Vantagens
 - Alternativa para a divisão de regras
 - Considera dependência lexical
 - Em vez de se alterarem as regras, altera-se o modelo probabilístico

91

GLCP LEXICALIZADAS

- Extensão de modelos anteriores
 - Além da *head*, a *tag*



92

GLCP LEXICALIZADAS

○ Dois tipos de regras

- Regras lexicais
 - subst(copo,subst)→copo
 - Atenção: probabilidade 1, pois não há outra opção (o terminal está explícito)
- Regras internas
 - S(quebrou,verbo) → SN(copo,subst), SV(quebrou,verbo)
 - Probabilidades precisam ser estimadas

93

GLCP LEXICALIZADAS

○ Estimativas de probabilidades

- Regras internas
 - S(quebrou,verbo) → SN(copo,subst), SV(quebrou,verbo)

$$P(\text{regra}) = \frac{\text{Número}(S(\text{quebrou, verbo}) \rightarrow \text{SN}(\text{copo, subst}), \text{SV}(\text{quebrou, verbo}))}{\text{Número}(S(\text{quebrou, verbo}))}$$

94

GLCP LEXICALIZADAS

o Estimativas de probabilidades

- Regras internas
 - o S(quebrou,verbo) → SN(copo,subst), SV(quebrou,verbo)

$$P(\text{regra}) = \frac{\text{Número}(S(\text{quebrou, verbo}) \rightarrow \text{SN}(\text{copo, subst}), \text{SV}(\text{quebrou, verbo}))}{\text{Número}(S(\text{quebrou, verbo}))}$$

- o Qual o problema?

95

GLCP LEXICALIZADAS

o Estimativas de probabilidades

- Regras internas
 - o S(quebrou,verbo) → SN(copo,subst), SV(quebrou,verbo)

$$P(\text{regra}) = \frac{\text{Número}(S(\text{quebrou, verbo}) \rightarrow \text{SN}(\text{copo, subst}), \text{SV}(\text{quebrou, verbo}))}{\text{Número}(S(\text{quebrou, verbo}))}$$

- o Qual o problema?
 - o Regras muito mais específicas
 - o Dados mais esparsos ainda
 - Maioria das probabilidades será zero!

96

- o Solução: ?

GLCP LEXICALIZADAS

○ Estimativas de probabilidades

- Regras internas
 - S(quebrou, verbo) → SN(copo, subst), SV(quebrou, verbo)

$$P(\text{regra}) = \frac{\text{Número}(S(\text{quebrou, verbo}) \rightarrow \text{SN}(\text{copo, subst}), \text{SV}(\text{quebrou, verbo}))}{\text{Número}(S(\text{quebrou, verbo}))}$$

- Qual o problema?
 - Regras muito mais específicas
 - Dados mais esparsos ainda
 - Maioria das probabilidades será zero!

97

- Solução: mais suposições de independência!

GLCP LEXICALIZADAS

○ Estimativas de probabilidades

- Modelo 1 do parser de Collins
 - Lado Direito da Regra (LDR): uma *head* + *símbolos que precedem a head* + *símbolos que seguem a head*
 - LER → $E_N E_{N-1} \dots E_1 \textit{head} D_1 \dots D_{M-1} D_M$
- Cálculo das probabilidades
 - Dado o lado esquerda da regra, computa-se a probabilidade se gerar a *head*
 - A partir da *head* e do lado esquerdo, gera-se cada um dos símbolos que precedem e seguem a *head*, individualmente
 - Deve-se controlar quando parar de gerar símbolos à esquerda e à direita da *head*

98

GLCP LEXICALIZADAS

○ Estimativas de probabilidades

• Exemplo

- S(quebrou,verbo) → SN(copo,subst), SV(quebrou,verbo)

- $P(\text{regra}) = \frac{P_{\text{HEAD}}(\text{SV}(\text{quebrou,verbo}) \mid \text{S}(\text{quebrou,verbo})) * P_{\text{ESQ}}(\text{SN}(\text{copo,subst}) \mid \text{S,SV}(\text{quebrou,verbo}))}{P_{\text{ESQ}}(\text{SN}(\text{copo,subst}) \mid \text{S,SV}(\text{quebrou,verbo}))}$

- Mais simples de se calcular, com menos dados esparsos

99

GLCP LEXICALIZADAS

○ Estimativas de probabilidades

• Varições dos modelos de Collins

- Distância entre elementos
- Subcategorização de verbos, identificando argumentos e adjuntos
- Somente a *tag* em vez da *head* e da *tag*
- Palavras “curinga”
- Etc.

100

GLCP LEXICALIZADAS

- Collins (2003)
 - Extensão do CKY, incluindo as probabilidades e as lexicalizações

101

RE-RANQUEAMENTO DE ANÁLISES

- Modelos gerativos como os anteriores são **muito bons**
 - Relativamente fácil calcular probabilidades
 - Bons resultados
- Mas é **difícil incorporar conhecimento externo**
 - Por exemplo
 - Árvores sintáticas tendem a “pender para a direita”
 - Constituintes mais longos acontecem no fim da árvore
 - Certos falantes/escritores têm preferências por estruturas sintáticas particulares → questões de estilo de escrita

102

RE-RANQUEAMENTO DE ANÁLISES

○ Possível solução

• Re-ranqueamento discriminativo

- Produz-se um ranque com as N melhores (mais prováveis) árvores sintáticas
 - Chamada *N-best list*
- Novo ranqueamento com base em um conjunto de atributos relevantes
 - Por exemplo, probabilidade, regras aplicadas, número de ocorrências de cada constituinte, bigramas de não terminais adjacentes na árvore, etc.
- Escolhe-se a melhor árvore

103

RE-RANQUEAMENTO DE ANÁLISES

○ Possível solução

• Re-ranqueamento discriminativo

- **Atenção:** a **qualidade do método** depende diretamente da **qualidade da *N-best list***
- Se a análise correta não estiver na lista ou estiver muito mal ranqueada, o método será provavelmente ruim

104

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Estruturas e palavras mais previsíveis (prováveis) são lidas mais rapidamente por humanos
 - Como se mede isso?

105

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Estruturas e palavras mais previsíveis (prováveis) são lidas mais rapidamente por humanos
 - Medidas empíricas: por exemplo, entropia vs. rastreamento do movimento dos olhos

106

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Humanos desambigam análises, preferindo análises mais prováveis
 - Sentenças ***garden-path***: temporariamente ambíguas
 - *The students forgot the solution was in the back of the book.*
 - *The horse raced past the barn fell.*
 - *The complex houses married and single students and their families.*

107

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Humanos desambigam análises, preferindo análises mais prováveis
 - Sentenças ***garden-path***: temporariamente ambíguas
 - *Por mais que Jorge continuasse lendo as histórias aborreciam as crianças da creche.*
 - *Maria beijou João e o irmão dele arregalou os olhos de espanto.*

108