SINTAXE – PARTE 2

SCC5869 Tópicos em Processamento de Língua Natural

Thiago A. S. Pardo

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)

3

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.

4

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 U 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(LDR|LER)
  - S é o símbolo inicial da gramática

5

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 U 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(LDR|LER)
  - S é o símbolo inicial da gramática

6

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 U 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(LDR|LER)

$$\sum_{\beta} P(\alpha \rightarrow \beta) = 1$$

7

  - S é o símbolo inicial da gramática

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 U 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(LDR|LER)

$$\sum_{\beta} P(\alpha \rightarrow \beta) = 1$$

8

  - S é o símbolo inicial da gramática

GLCP

- A **gramática** é dita **consistente** se a soma das probabilidades de todas as sentenças da linguagem resultam em 1
  - Algumas recursões podem atrapalhar isso

9

GLCP

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

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

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

$$P(sentença, árvore) = P(árvore) \times P(sentença \mid árvore)$$
$$P(sentença, árvore) = P(árvore) \times 1$$
$$P(sentença, árvore) = P(árvore)$$

10

GLCP

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

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

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

$$P(sentença, árvore) = P(árvore) \times P(sentença \mid árvore)$$
$$P(sentença, árvore) = P(árvore) \times 1$$
$$P(sentença, árvore) = P(árvore)$$

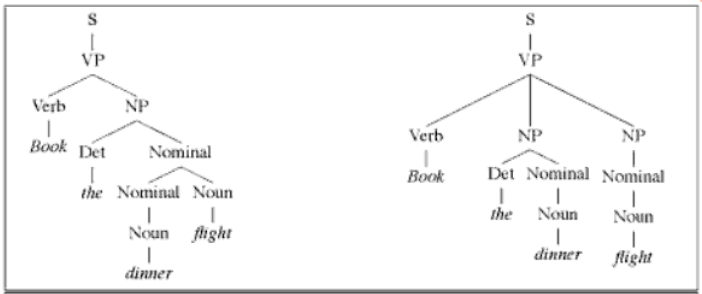
Como é possível?

11

GLCP: EXEMPLO

Qual a correta?

O que significam?

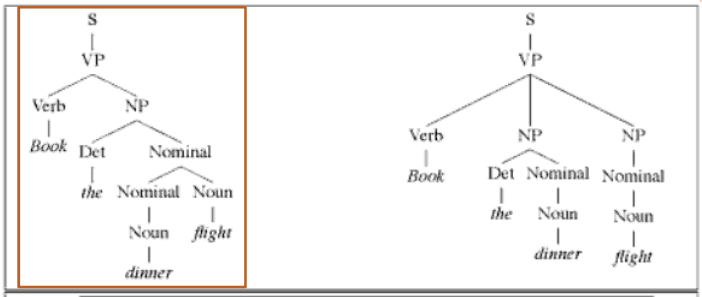


Rules			P	Rules			P
S	→	VP	.05	S	→	VP	.05
VP	→	Verb NP	.20	VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20	NP	→	Det Nominal	.20
Nominal	→	Nominal Noun	.20	NP	→	Nominal	.15
Nominal	→	Noun	.75	Nominal	→	Noun	.75
				Nominal	→	Noun	.75
Verb	→	book	.30	Verb	→	book	.30
Det	→	the	.60	Det	→	the	.60
Noun	→	dinner	.10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

GLCP: EXEMPLO

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

P(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			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

PARSING PROBABILÍSTICO

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

14

PARSING PROBABILÍSTICO

- É simples estender os métodos CKY ou de Earley para considerar probabilidades
  - Pode-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]

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

PARSING PROBABILÍSTICO

- Aprendizado de probabilidades
  - Alternativa 1: há um treebank

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

  - Exemplo hipotético

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

16

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

17

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]

18

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.

19

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]

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

Córpus	Córpus anotado
Ele morreu.	[[Ele$_{PRONOME}$]$_{SN}$ [morreu$_{VERBO}$]$_{SV}$]$_{S}$ → 0.5*0.33*1=0.165
A menina chorou.	[[A$_{ART}$ menina$_{SUBST}$]$_{SN}$ [chorou$_{VERBO}$]$_{SV}$]$_{S}$ → 0.5*0.33*1=0.165
Ela gritou.	[[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 houve mudança nos números, para-se; caso contrário, volta-se ao passo 2

23

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

24

GLCP: problemas

- 2 principais limitações
  - Suposições fracas de independência
  - Falta de informação lexical

25

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

26

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

27

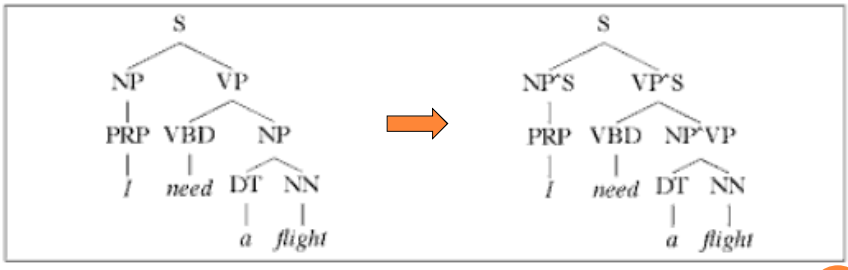
GLCP: problemas

- 2 principais limitações
  - Suposições fracas de independência
    - Solução possível: dividir as regras
      - $SN_{SUJEITO}$→pronome [0.91]
      - SN_{OBJETO}→pronome [0.34]
    - Forma de implementação: anexar a cada símbolo o símbolo de seu nó pai → nó_filho^nó_pai

28

GLCP: problemas

- 2 principais limitações
  - Suposições fracas de independência

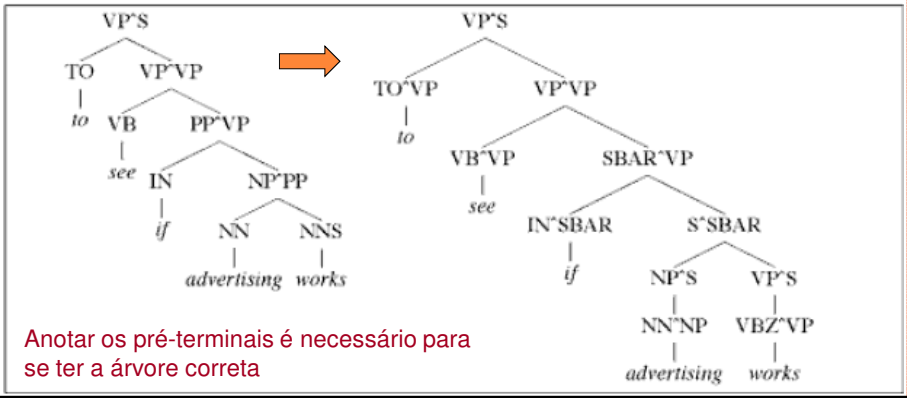


29

Sem anotar os pré-terminais (etiquetas morfossintáticas)

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*

31

GLCP: problemas

- 2 principais limitações
  - Suposições fracas de independência
    - Problemas dessa abordagem?

32

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

33

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.

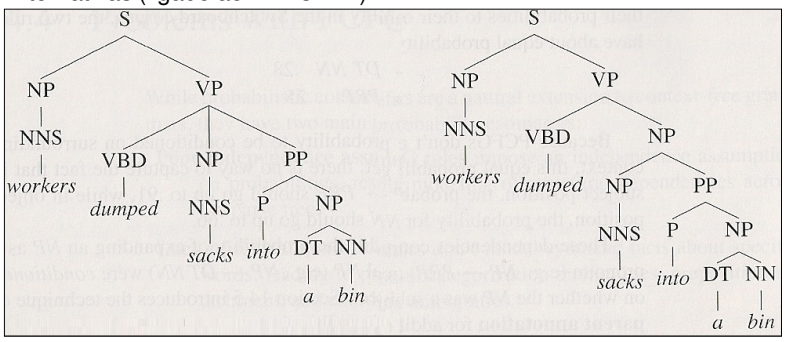
MAIS PROVÁVEL: *dumped* e *into* têm mais afinidade do que *sacks* e *into*

34

GLCP: problemas

- 2 principais limitações
  - Falta de informação lexical

Alternativas (ligado ao VP vs. NP)

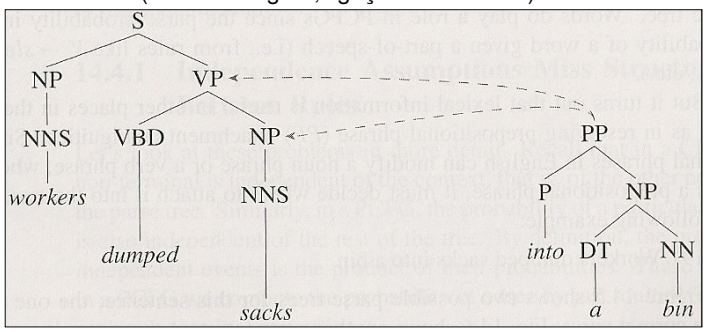


35

GLCP: problemas

- 2 principais limitações
  - Falta de informação lexical

Alternativas (mesmas regras, ligações diferentes)



36

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*

37

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]*

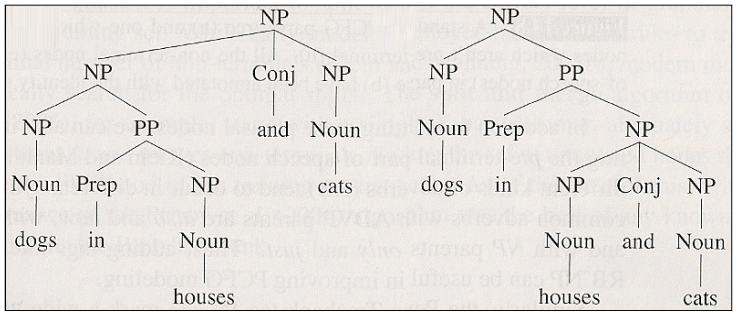
MAIS PROVÁVEL: *dogs* e *cats* são mais afins... e *dogs* não cabem dentro de *cats*

38

GLCP: problemas

- 2 principais limitações
  - Falta de informação lexical

Alternativas



39

GLCP: problemas

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

40

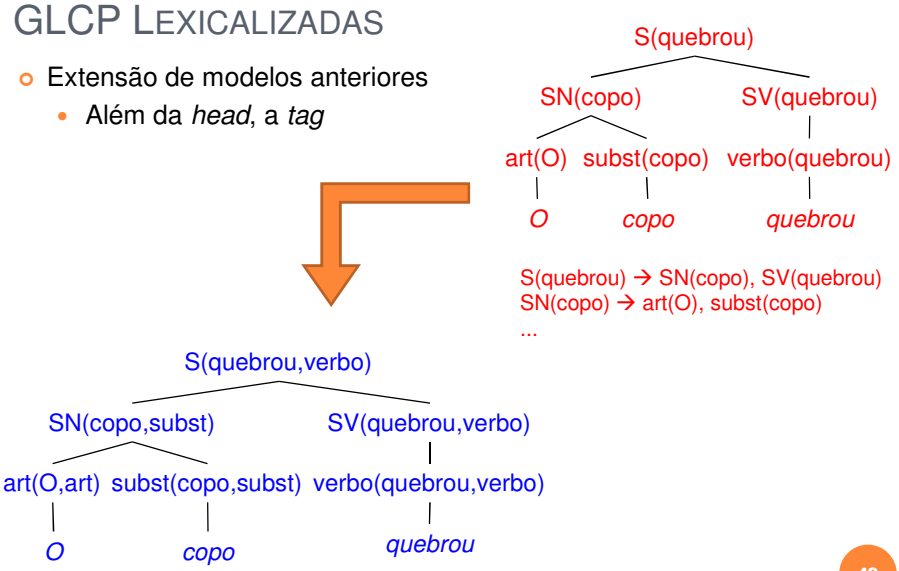
GLCP LEXICALIZADAS

- 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

41

GLCP LEXICALIZADAS

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



42

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

43

GLCP LEXICALIZADAS

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

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

44

GLCP LEXICALIZADAS

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

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

    - Qual o problema?

45

GLCP LEXICALIZADAS

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

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

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

46

GLCP LEXICALIZADAS

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

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

    - Qual o problema?
      - Regras muito mais específicas
      - Dados mais esparsos ainda
        - Maioria das probabilidades será zero!
    - Solução: mais suposições de independência!

47

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 → $E_N\ E_{N-1}\ ...\ E_1$ *head* $D_1\ ...\ D_{M-1}\ D_M$
  - 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*

48

GLCP LEXICALIZADAS

- Estimativas de probabilidades
  - Exemplo
    - S(quebrou,verbo) → SN(copo,subst), SV(quebrou,verbo)
      - P(regra) = P_{HEAD}(SV(quebrou,verbo)|S(quebrou,verbo)) * P_{ESQ}(SN(copo,subst)|S,SV(quebrou,verbo))
    - Mais simples de se calcular, como menos dados esparsos

49

GLCP LEXICALIZADAS

- Estimativas de probabilidades
  - Variações 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.

50

GLCP LEXICALIZADAS

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

51

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

52

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

53

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

54

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos
  - Estruturas e palavras mais previsíveis (prováveis) são lidas mais rapidamente por humanos
    - Como se mede isso?

55

PROCESSAMENTO HUMANO & PROBABILIDADE

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

56

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos
  - Humanos desambiguam 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.*

57

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos
  - Humanos desambiguam 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.*

58