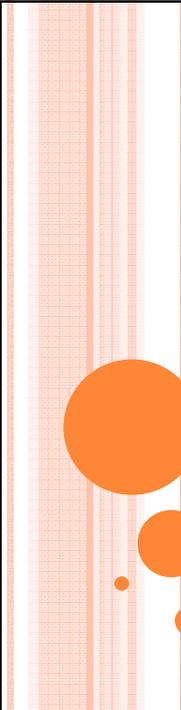


# SINTAXE – PARTE 1

*SCC5869 Tópicos em Processamento de Língua Natural*

Thiago A. S. Pardo



# SINTAXE E GRAMÁTICAS

## DEFINIÇÃO

- Forma como as **palavras se organizam** em uma sentença
  - Longa história: gramática do Sânscrito, com mais de 2.000 anos
  
- Questões envolvidas
  - Constituintes
  - Relações/funções gramaticais
  - Subcategorização e dependência

3

## APLICAÇÃO

- **Útil** para diversos fins em PLN
  - Revisão gramatical
  - Interpretação semântica
  - Sistemas de diálogo
  - Tradução automática
  - Sumarização de textos
  - Outros?

4

## CONSTITUINTES

- Sintagma nominal
  - “Ela”, “João”, “a casa”, “o cavalo Pangaré”, “uma bela moça”
- Sintagma verbal
  - “Eu corri.”, “Ele precisa de uma passagem”, “João e Maria deram o livro para ela.”, “Faça!”
- Sintagma adjetival
  - “Ele é competente.”, “Maria é muito desleal.”
- Sintagma adverbial
  - “Antigamente tudo era diferente.”, “Nós acordamos muito cedo.”
- Sintagma preposicional
  - “O armário da cozinha está trancado.”, “Ele queimou o livro de física.”

5

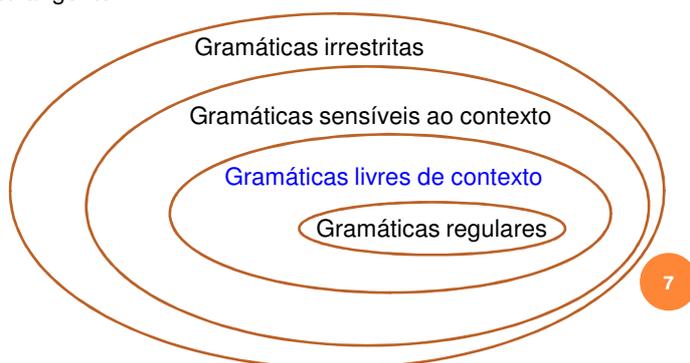
## CONSTITUINTES

- Um sintagma é do tipo de seu elemento nuclear
  - Substantivo, verbo, adjetivo, advérbio ou preposição
- É comum ter sintagma dentro de sintagma
  - [O vidro de remédio]<sub>SN</sub> [quebrou]<sub>SV</sub>
  - [O vidro [de remédio]<sub>SP</sub>]<sub>SN</sub> [quebrou]<sub>SV</sub>
  - [O vidro [de [remédio]<sub>SN</sub>]<sub>SP</sub>]<sub>SN</sub> [quebrou]<sub>SV</sub>
- **Pontos importantes**
  - Concordância de gênero e número entre elementos
  - Subcategorização dos verbos
    - Preferências sintáticas
    - Alternativa: predicado e argumentos

6

## SINTAXE

- Tradicionalmente representada por **gramáticas livres de contexto**
- Hierarquia de Chomsky
  - Uma gramática/linguagem de qualquer tipo também é do tipo mais abrangente



## SINTAXE

- Tradicionalmente representada por gramáticas livres de contexto
- Muito boas para fins computacionais
  - Poderosas, mas ainda assim eficientemente manipuladas
- Compostas por
  - Regras/produções
    - Indicam como os símbolos da linguagem podem ser agrupados
    - Podem ter recursões
  - Léxico
    - Palavras/símbolos da linguagem

8

## GRAMÁTICA LIVRE DE CONTEXTO

### o REGRAS

- Sentença → SN SV
- Sentença → SV
- SN → pronome
- SN → substantivo
- SN → artigo substantivo
- SV → verbo
- SV → verbo SN
- SV → verbo SN SP
- SP → preposição SN

### o LÉXICO

- artigo → o | a | os | ...
- pronome → eu | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

9

## GRAMÁTICA LIVRE DE CONTEXTO

### o Gramática pode ser usada para

- Gerar sentenças
- Reconhecer sentenças

### o Derivação

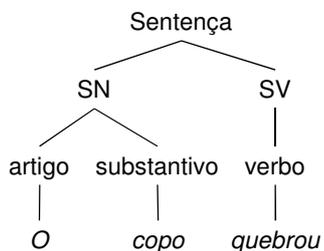
- Seqüência de aplicação de regras da gramática
- Gera uma *árvore sintática* (*parse tree*)

10

## GRAMÁTICA LIVRE DE CONTEXTO

### ○ Árvore sintática

- *O copo quebrou.*



### • Terminologia

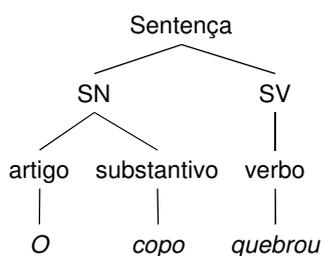
- Sentença **domina** todos os nós da árvore
- Sentença **domina imediatamente** SN e SV
- SN e SV são **filhos** de Sentença
- SN, SV, etc. são **descendentes** de Sentença

11

## GRAMÁTICA LIVRE DE CONTEXTO

### ○ Árvore sintática

- *O copo quebrou.*



### • Terminologia

- Sentença é o **símbolo inicial**
- Sentença, SN, SV, artigo, substantivo e verbo são **símbolos não terminais**
- O, copo e quebrou são **símbolos terminais**
- Se sentença é gerada/reconhecida pela gramática, é dita **gramatical**
- Notação parentizada:  $[[O_{artigo} copo_{substantivo}]_{SN} [quebrou_{verbo}]_{SV}]_{Sentença}$

12

## GRAMÁTICA LIVRE DE CONTEXTO

- Formalmente, uma gramática é uma **quádrupla**
  - $G = (N, T, P, S)$ 
    - N: conjunto de símbolos não terminais
    - T: conjunto de símbolos terminais
    - P: conjunto de regras de produção
    - S: símbolo inicial da gramática
  - Gramática livre de contexto
    - Regras da forma  $N \rightarrow (N \cup T)^*$

13

## PARSING

- Tarefa de mapear uma sentença em uma árvore sintática
  - Ferramenta: *parser*
- Atenção com o termo: parsing é muito genérico e pode significar outras coisas dependendo do contexto
  - Parser semântico
  - Parser discursivo
  - Tagger vs. parser

14

## TREEBANKS

- **Coleção de sentenças e suas árvores sintáticas**, normalmente construídas manualmente
  
- Exemplos
  - *Penn Treebank* para o inglês (há para outras línguas também)
    - *Wall Street Journal*, principalmente
  - Susanne para o inglês
  - *Prague Dependency Treebank* para o tcheco
  - Negra para o alemão
  - Floresta Sintá(c)tica para o português
  - Tycho Brahe para português histórico
  
- Linguagens de consulta a treebanks
  - TGrep e TGrep2

15

## TREEBANKS

- Formato parentizado é comum
  
- Alguns treebanks contêm outras anotações
  - Predicados e argumentos
  - Funções gramaticais (sujeito, objeto, etc.)
  - Funções semânticas (local, tempo, etc.)
  
- Para que treebanks?

16

## TREEBANKS

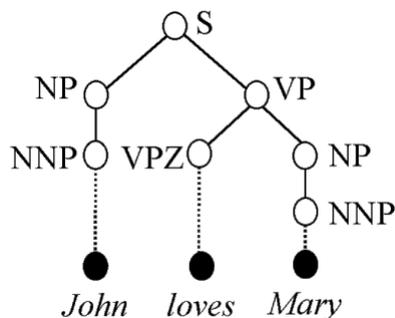
- Formato parentizado é comum
- Alguns treebanks contêm outras anotações
  - Predicados e argumentos
  - Funções gramaticais (sujeito, objeto, etc.)
  - Funções semânticas (local, tempo, etc.)
- Gramática da língua embutida nas análises
  - As árvores sintáticas podem ser a base para a construção de gramáticas
  - Pode ser a fonte de estudos de fenômenos lingüísticos

17

## TREEBANKS

- Exemplo do Penn Treebank

(S (NP (NNP *John*))  
 (VP (VPZ *loves*)  
 (NP (NNP *Mary*)))  
 (. .))



## TREEBANKS

### Exemplo da Floresta Sintá(c)tica

Queremos	[querer] <fmc> V PR 1P IND VFIN @FMV
que	[que] KS @SUB @#FS-<ACC
especialistas	[especialista] N M/F P @-<ACC
internacionais	[internacional] ADJ M/F P @N<
e	[e] <co-postnom> KC @CO
nacionais	[nacional] ADJ M/F P @N<
pensem	[pensar] V PR 3P SUBJ VFIN @FMV
em	[em] PRP @-<PIV
as	[o] <artd> DET F P @>N
possibilidades	[possibilidade] N F P @P<
que	[que] KS @SUB @#FS-<ACC
existem	[existir] V PR 3P IND VFIN @FMV
de	[de] PRP @-<ADVL
abordagem	[abordagem] N F S @P<
de	[de] PRP @N<
o	[o] <artd> DET M S @>N
tema	[tema] N M S @P<
em	[em] PRP @N<
o	[o] <artd> DET M S @>N
contexto	[contexto] N M S @P<
de	[de] <sam-> PRP @N<
a	[o] <sam> <artd> DET F S @>N
sociedade	[sociedade] N F S @P<
de	[de] <sam-> PRP @N<
a	[o] <sam> <artd> DET F S @>N
informação	[informação] N F S @P<

## TREEBANKS

### Heads

- Elemento lexical gramaticalmente mais importante de um constituinte
  - Por exemplo, um substantivo em um SN
- Noção importante para várias linhas de pesquisa, práticas ou teóricas
  - Treinamento automático de parsers
  - *Head-driven Phrase Structure Grammar* (HPSG)
- Nem sempre é trivial encontrar as *heads*
  - Pode ser necessária a aplicação de várias regras

## TREEBANKS

### ○ Heads

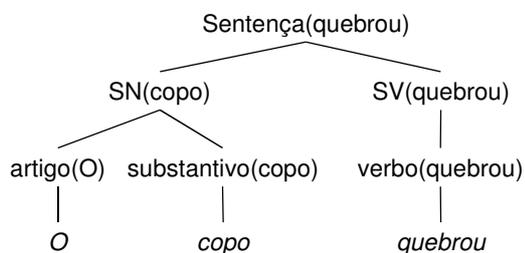
- Exemplo de regras para encontrar elemento mais importante de um SN para o inglês
  - Se última palavra é etiquetada como POS, retorne-a
    - Senão procure da direita para a esquerda pelo primeiro filho que é um NN, NNP, NNPS, NX, POS ou JJR
    - Senão procure da esquerda para a direita pelo primeiro filho que é um SN
    - Senão procure da direita para a esquerda pelo primeiro filho que é um \$, ADJP ou PRN
    - Senão procure da direita para a esquerda pelo primeiro filho que é um CD
    - Senão procure da direita para a esquerda pelo primeiro filho que é um JJ, JJS, RB ou QP
    - Senão retorne a última palavra

21

## TREEBANKS

### ○ Heads

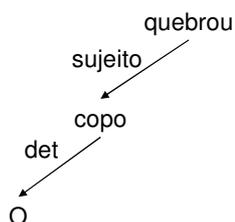
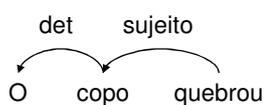
- Podem ser associadas a cada nó da árvore



22

## GRAMÁTICA DE DEPENDÊNCIA

- Alternativa à gramática de constituintes
- Foco nas **relações gramaticais**
  - Explicitamente rotuladas ou não



23

## GRAMÁTICA DE DEPENDÊNCIA

- Cada vez mais populares
- **Vantagens?**

24

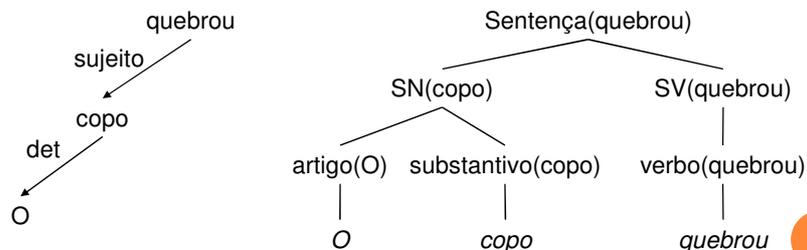
## GRAMÁTICA DE DEPENDÊNCIA

- Cada vez mais populares
- **Vantagens**
  - Maior poder preditivo das palavras em relação a seus dependentes
    - Saber a identidade de um verbo pode ajudar a decidir quem é seu sujeito
  - Lidam mais facilmente com línguas com ordenação livre de palavras (por exemplo, tcheco)
    - Na gramática de constituintes, seriam necessárias várias regras para montar os constituintes adequados
    - Na gramática de dependências, não (basta um link entre as palavras)

25

## GRAMÁTICA DE DEPENDÊNCIA

- Certa similaridade entre **análise de dependência** e **heads**
  - É possível mapear parcialmente uma estrutura de constituintes em uma de dependência



26

## FALA

- Gramática

- Coisas em comum com língua escrita
- Muitos **outros fenômenos**
  - Pronomes são muito mais usados
  - Pequenos fragmentos de fala
  - Características fonológicas, prosódicas e acústicas
  - Disfluências: hesitação, pausa, reparo, recomeço, gagueira, etc.

- Também há treebanks importantes

27

## GRAMÁTICAS E PROCESSAMENTO HUMANO

- Há evidências de que **sintagmas são mais do que um artefato sintático**
  - Representam uma unidade semântica, em geral

28

## FORMALISMOS GRAMATICAIS

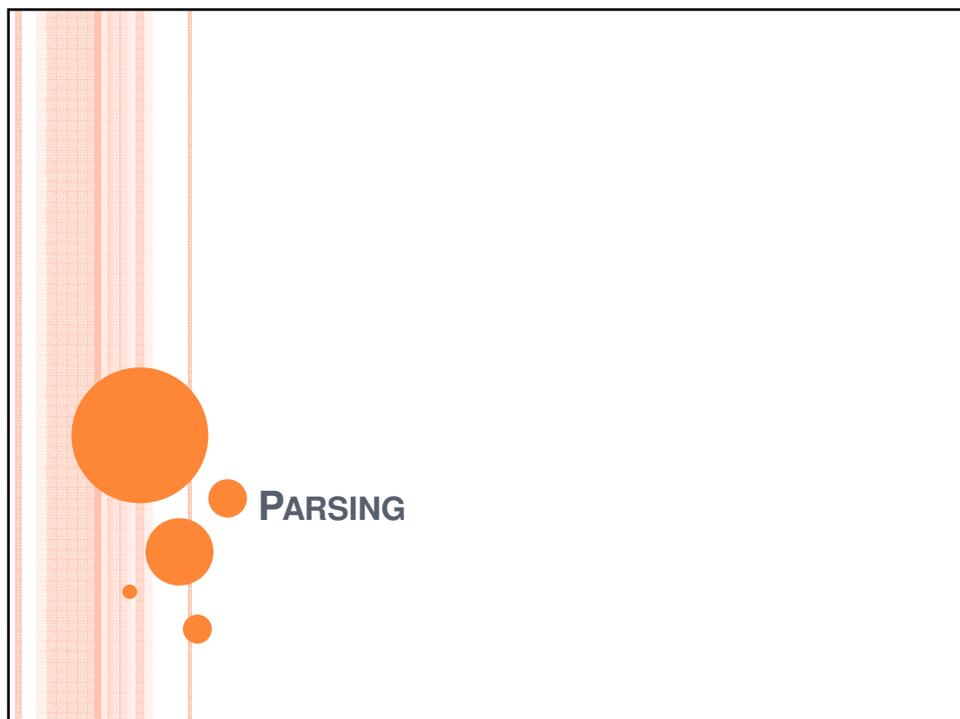
- DCG: *definite-clause grammar*
- LFG: *lexical functional grammar*
- GPSG: *generalized phrase structure grammar*
- HPSG: *head-driven phrase structure grammar*
- TAG: *tree adjoining grammar*
  - Árvores em vez de regras
- Etc.

29

## GRAMÁTICAS

- Regras
  - Escritas manualmente
    - Demandam tempo, sujeitas a erros e inconsistências humanas
    - Podem ter pouca cobertura
  - Aprendidas automaticamente
    - Rápido, mas sujeitas a erros e inconsistências por *overfitting* ou *underfitting*
    - Podem não fazer sentido
      - *Experiência com Penn Treebank*: regras longas, árvores mais “achatadas”
- Aprendidas automaticamente, revisadas por humanos

30



## PARSING

### o **Questão**

- Dada uma gramática, como analisar uma sentença para produzir sua árvore sintática?
  - o *Top-down*, ou descendente
  - o *Bottom-up*, ou ascendente

## PARSING

### ○ Análise top-down

- Da raiz para as folhas (palavras)

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

Sentença

33

## PARSING

### ○ Análise top-down

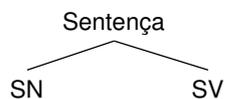
- Da raiz para as folhas (palavras)

#### ○ REGRAS

- **Sentença → SN SV**
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



34

## PARSING

### ○ Análise top-down

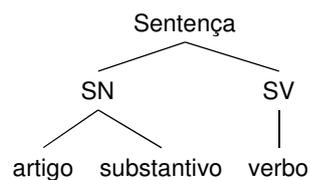
- Da raiz para as folhas (palavras)

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



35

## PARSING

### ○ Análise top-down

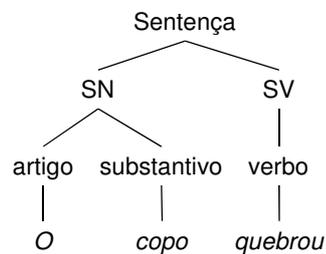
- Da raiz para as folhas (palavras)

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- **substantivo → casa | brinquedo | copo | abacaxi | ...**
- **verbo → correu | comprou | faça | quebrou | deu | ...**
- preposição → de | para | em | ...



36

## PARSING

### ○ Análise top-down

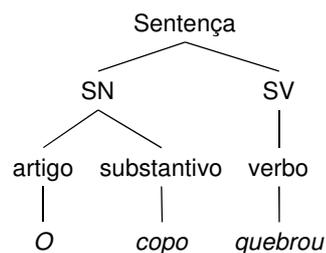
- Da raiz para as folhas (palavras)

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



37

Sentença gramatical!  
Mas se chega diretamente a ela?

## PARSING

### ○ Análise top-down

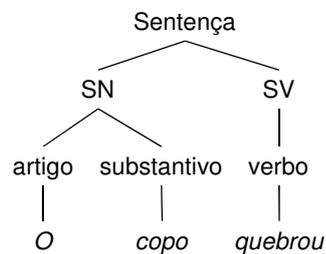
- Da raiz para as folhas (palavras)

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



38

**Não!**  
Qual o problema?

## PARSING

### ○ Análise top-down

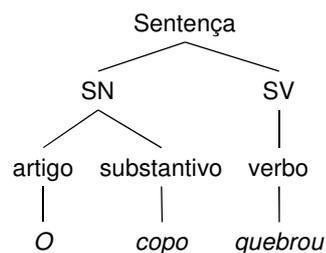
- Da raiz para as folhas (palavras)

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



Várias regras são testadas.  
Pode haver **backtracking!**

39

MUNDO REAL

## PARSING

### ○ Análise top-down

- Da raiz para as folhas (palavras)

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

Sentença

40

## MUNDO REAL

## PARSING

## ○ Análise top-down

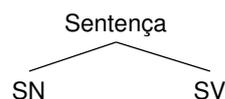
- Da raiz para as folhas (palavras)

## ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



41

## MUNDO REAL

## PARSING

## ○ Análise top-down

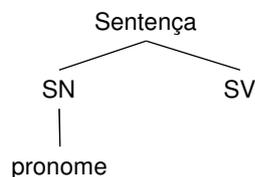
- Da raiz para as folhas (palavras)

## ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



42

## MUNDO REAL

## PARSING

## ○ Análise top-down

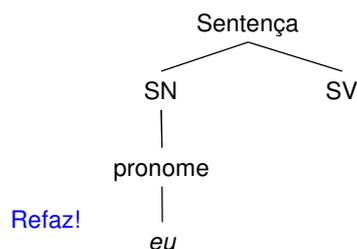
- Da raiz para as folhas (palavras)

## ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



43

## MUNDO REAL

## PARSING

## ○ Análise top-down

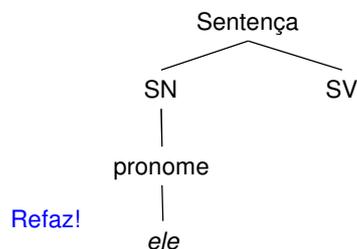
- Da raiz para as folhas (palavras)

## ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



44

## MUNDO REAL

## PARSING

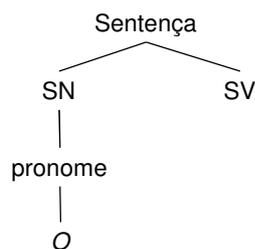
- Análise top-down
  - Da raiz para as folhas (palavras)

## ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



45

## MUNDO REAL

## PARSING

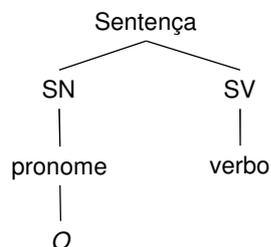
- Análise top-down
  - Da raiz para as folhas (palavras)

## ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



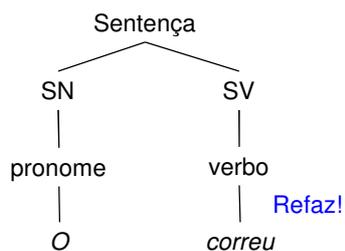
46

## PARSING

- Análise top-down
  - Da raiz para as folhas (palavras)

- REGRAS
  - Sentença → SN SV
  - Sentença → 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 | ele | ela | ...
  - substantivo → casa | brinquedo | copo | abacaxi | ...
  - verbo → correu | comprou | faça | quebrou | deu | ...
  - preposição → de | para | em | ...



47

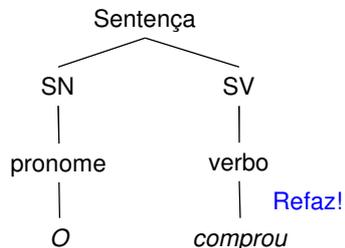
## PARSING

- Análise top-down
  - Da raiz para as folhas (palavras)

- REGRAS
  - Sentença → SN SV
  - Sentença → 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 | ele | ela | ...
  - substantivo → casa | brinquedo | copo | abacaxi | ...
  - verbo → correu | comprou | faça | quebrou | deu | ...
  - preposição → de | para | em | ...

Refaz várias vezes!  
 → Não vai achar regra que cubra a sentença  
 → Vai testar a segunda opção para SN e recomeçar o processo  
 → Muito esforço repetido



48

## PARSING

### ○ Análise **bottom-up**

- Das folhas (palavras) para a raiz

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

O      copo      quebrou

49

## PARSING

### ○ Análise **bottom-up**

- Das folhas (palavras) para a raiz

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- **substantivo** → casa | brinquedo | **copo** | abacaxi | ...
- **verbo** → correu | comprou | faça | **quebrou** | deu | ...
- preposição → de | para | em | ...

artigo    substantivo    verbo  
|                    |                    |  
O            copo            quebrou

50

## PARSING

### ○ Análise bottom-up

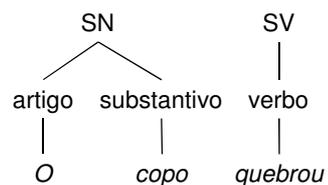
- Das folhas (palavras) para a raiz

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



51

## PARSING

### ○ Análise bottom-up

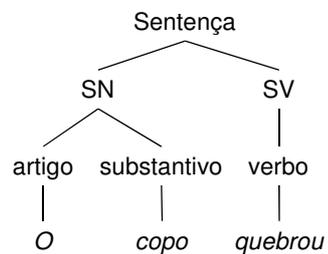
- Das folhas (palavras) para a raiz

#### ○ REGRAS

- **Sentença → SN SV**
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



52

## PARSING

### ○ Análise bottom-up

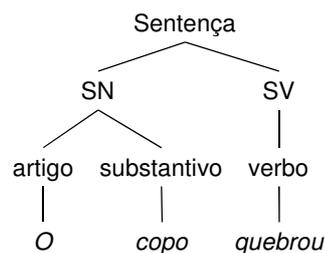
- Das folhas (palavras) para a raiz

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



Tem problemas?

53

## PARSING

### ○ Análise bottom-up

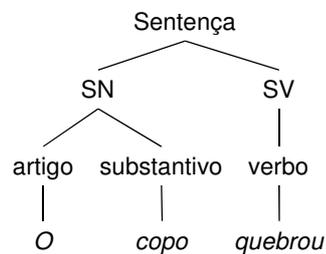
- Das folhas (palavras) para a raiz

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...



Também há várias regras, que podem produzir árvores que nunca cheguem ao símbolo inicial

54

## PARSING

### o DCG e PROLOG

- Gramática para gerar ou reconhecer sentenças (top-down)

```

sentenca --> sintagma_nominal, sintagma_verbal.
sintagma_nominal --> artigo, substantivo.
sintagma_verbal --> verbo, sintagma_nominal.
artigo --> [o].
substantivo --> [gato].
substantivo --> [rato].
verbo --> [matou].

```

- Notação
  - o Símbolos separados por vírgula
  - o Regras terminadas por ponto
  - o Palavras indicadas entre colchetes

55

## PARSING

### o DCG e PROLOG

- Gramática para gerar ou reconhecer sentenças (top-down)

```

sentenca --> sintagma_nominal, sintagma_verbal.
sintagma_nominal --> artigo, substantivo.
sintagma_verbal --> verbo, sintagma_nominal.
artigo --> [o].
substantivo --> [gato].
substantivo --> [rato].
verbo --> [matou].

```

- Para reconhecer sentença
  - o `sentenca([o,gato,matou,o,rato],[ ])`.
- Para gerar sentenças
  - o `sentenca(S,[ ])`.

56

## PARSING

### ○ Exercício para casa

- Testar gramática abaixo

#### ○ REGRAS

- Sentença → SN SV
- Sentença → 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 | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

57

## AMBIGÜIDADE

- Há vários **tipos de ambigüidades** que afetam o parsing

- **Etiquetas morfossintáticas**

- *Book the flight!*
- Substantivo vs. verbo

- **Funções gramaticais**

- *He gave her his book*
- Objeto direto vs. indireto

58

## AMBIGÜIDADE

- Há vários tipos de ambigüidades que afetam o parsing
  - **Estrutural**
    - Ligação
      - *Ele viu a Torre Eiffel voando para Paris.*
        - viu ← voando para Paris
        - Torre Eiffel ← voando para Paris
    - Coordenação
      - Ele chamou amigos e amigas legais.
        - [amigos] e [amigas legais]
        - [amigos e amigas] legais

59

## AMBIGÜIDADE

- Há vários tipos de ambigüidades que afetam o parsing
  - Geram diferentes árvores sintáticas
    - Um parser pode relatar todas (podem ser muitas!!!)
  - Em geral, faz-se necessária a desambiguação sintática
    - Escolha da melhor análise
      - Utilizando critérios estatísticos, semânticos ou pragmáticos
        - **Exemplos desses critérios?**

60

## AMBIGÜIDADE

- Há vários tipos de ambigüidades que afetam o parsing
  - **Ligação do SP**
    - Um dos maiores problemas para a língua inglesa
    - Responsável pela grande maioria dos erros atuais dos parsers
    - Acredita-se que somente a semântica pode ajudar

61

## MÉTODOS DE PARSING

- **Aplicação seqüencial de regras gramaticais**
  - Estilos *top-down* ou *bottom-up* clássicos
    - Várias desvantagens

62

## 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)

63

## MÉTODOS DE PARSING

### ○ CKY: algoritmo de Cocke-Kasami-Younger

- Primeiro passo: converter gramática livre de contexto para a Forma Normal de Chomsky (FNC)
  - Somente produções da forma  $N \rightarrow N N$  e  $N \rightarrow T$
  - Procedimentos simples
    - Produções novas
      - $A \rightarrow a B \rightarrow A \rightarrow X B, X \rightarrow a$
      - $A \rightarrow B C D \rightarrow A \rightarrow X C, X \rightarrow B C$
    - União de produções
      - $A \rightarrow B C, B \rightarrow D, D \rightarrow E \rightarrow A \rightarrow E C$
    - Aplicação sistemática/recursiva dos procedimentos acima

64

## MÉTODOS DE PARSING

- Exemplo de gramática para o inglês

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

65

## MÉTODOS DE PARSING

- Exemplo de gramática para o inglês

- Conversão da gramática para CNF

- Léxico não precisa ser convertido (não é mostrado)

$\mathcal{L}_1$ Grammar	$\mathcal{L}_1$ in CNF
$S \rightarrow NP VP$	$S \rightarrow NP VP$
$S \rightarrow Aux NP VP$	$S \rightarrow X1 VP$
$S \rightarrow VP$	$X1 \rightarrow Aux NP$
	$S \rightarrow book \mid include \mid prefer$
	$S \rightarrow Verb NP$
	$S \rightarrow X2 PP$
	$S \rightarrow Verb PP$
	$S \rightarrow VP PP$
$NP \rightarrow Pronoun$	$NP \rightarrow I \mid she \mid me$
$NP \rightarrow Proper-Noun$	$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det Nominal$	$NP \rightarrow Det Nominal$
$Nominal \rightarrow Noun$	$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal Noun$	$Nominal \rightarrow Nominal Noun$
$Nominal \rightarrow Nominal PP$	$Nominal \rightarrow Nominal PP$
$VP \rightarrow Verb$	$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb NP$	$VP \rightarrow Verb NP$
$VP \rightarrow Verb NP PP$	$VP \rightarrow X2 PP$
	$X2 \rightarrow Verb NP$
$VP \rightarrow Verb PP$	$VP \rightarrow Verb PP$
$VP \rightarrow VP PP$	$VP \rightarrow VP PP$
$PP \rightarrow Preposition NP$	$PP \rightarrow Preposition NP$

## MÉTODOS DE PARSING

- CKY: algoritmo de Cocke-Kasami-Younger
  - Primeiro passo: converter gramática livre de contexto para a Forma Normal de Chomsky (FNC)
    - Por que converter? Qual a vantagem?

67

## MÉTODOS DE PARSING

- CKY: algoritmo de Cocke-Kasami-Younger
  - Primeiro passo: converter gramática livre de contexto para a Forma Normal de Chomsky (FNC)
    - Por que converter? Qual a vantagem?
      - A árvore sintática será binária, ou seja, cada nó pode ter no máximo dois filhos
        - Exatamente o que precisamos se vamos lidar com TABELAS

68

## MÉTODOS DE PARSING

- CKY: algoritmo de Cocke-Kasami-Younger
  - Segundo passo: construir uma tabela/matriz quadrada de  $(N+1)$  linhas por  $(N+1)$  colunas
    - $N$  é o número de palavras da sentença a se analisar
    - Por exemplo, posição  $[0,1]$  indica palavra entre 0 e 1: *Book*
      - $_0$  *Book*  $_1$  *the*  $_2$  *flight*  $_3$  *through*  $_4$  *Houston*  $_5$

69

## MÉTODOS DE PARSING

- CKY: algoritmo de Cocke-Kasami-Younger

- Somente a parte de cima da tabela é usada

	<i>Book</i>	<i>the</i>	<i>flight</i>	<i>through</i>	<i>Houston</i>

## MÉTODOS DE PARSING

- CKY: algoritmo de Cocke-Kasami-Younger
  - Terceiro passo: recheiar a tabela usando a gramática e o léxico
    - Passo a passo, da esquerda para a direita, de baixo para cima
      - Cada célula verifica as células que domina
    - Todas as opções até que se chegue no canto superior direito da tabela, que seria a raiz da árvore
    - Atenção: deve-se relacionar segmentos contínuos e completos

71

## 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

Tabela vazia inicial

	<i>Book</i>	<i>the</i>	<i>flight</i>	<i>through</i>	<i>Houston</i>

## MÉTODOS DE PARSING

### o 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

Começando pela diagonal principal  
→ etiquetas possíveis das palavras

	Book	the	flight	through	Houston
S, VP, Verb, Nominal, Noun					
	Det				
			Nominal, Noun		
				Prep	
					NP, proper noun

## MÉTODOS DE PARSING

### o 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

Próxima diagonal  
→ primeiros constituintes

	Book	the	flight	through	Houston
S, VP, Verb, Nominal, Noun	---				
	Det	NP			
			Nominal, Noun	---	
				Prep	PP
					NP, proper noun

## MÉTODOS DE PARSING

o CKY: algoritmo de Cocke-Kasami-Younger

*S* → *NP VP*  
*S* → *X1 VP*  
*X1* → *Aux NP*  
*S* → *book | include | prefer*  
*S* → *Verb NP*  
*S* → *X2 PP*  
*S* → *Verb PP*  
*S* → *VP PP*  
*NP* → *I | she | me*  
*NP* → *TWA | Houston*  
*NP* → *Det Nominal*  
*Nominal* → *book | flight | meal | money*  
*Nominal* → *Nominal Noun*  
*Nominal* → *Nominal PP*  
*VP* → *book | include | prefer*  
*VP* → *Verb NP*  
*VP* → *X2 PP*  
*X2* → *Verb NP*  
*VP* → *Verb PP*  
*VP* → *VP PP*  
*PP* → *Preposition NP*

+ léxico

Próxima diagonal

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

## MÉTODOS DE PARSING

o CKY: algoritmo de Cocke-Kasami-Younger

*S* → *NP VP*  
*S* → *X1 VP*  
*X1* → *Aux NP*  
*S* → *book | include | prefer*  
*S* → *Verb NP*  
*S* → *X2 PP*  
*S* → *Verb PP*  
*S* → *VP PP*  
*NP* → *I | she | me*  
*NP* → *TWA | Houston*  
*NP* → *Det Nominal*  
*Nominal* → *book | flight | meal | money*  
*Nominal* → *Nominal Noun*  
*Nominal* → *Nominal PP*  
*VP* → *book | include | prefer*  
*VP* → *Verb NP*  
*VP* → *X2 PP*  
*X2* → *Verb NP*  
*VP* → *Verb PP*  
*VP* → *VP PP*  
*PP* → *Preposition NP*

+ léxico

Próxima diagonal

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

## MÉTODOS DE PARSING

### o 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

Última diagonal  
→ raiz da árvore

	<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

### o CKY: algoritmo de Cocke-Kasami-Younger

- O processo tem sucesso se chega ao símbolo inicial da gramática no canto direito superior
- A partir da tabela, é possível recuperar todas as árvores sintáticas possíveis
  - o Cada constituinte encontrado pode armazenar junto de si os filhos que lhe deram origem
- É possível pós-processar a árvore para remodelar a gramática para a gramática original, antes de se tornar CNF
  - o Mais natural
- Estilo bottom-up de análise
- Pode ser on-line/por demanda, ou seja, analisar a sentença conforme as palavras aparecem

## MÉTODOS DE PARSING

### o 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

On-line, palavra a palavra

	Book	the	flight	through	Houston
S, VP, Verb, Nominal, Noun					

## MÉTODOS DE PARSING

### o 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

On-line, palavra a palavra

	Book	the	flight	through	Houston
S, VP, Verb, Nominal, Noun	---				
	Det				

## 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

On-line, palavra a palavra

	Book	the	flight	through	Houston
S, VP, Verb, Nominal, Noun	---		S, VP, X2		
	Det		NP		
			Nominal, Noun		

## 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

On-line, palavra a palavra

	Book	the	flight	through	Houston
S, VP, Verb, Nominal, Noun	---		S, VP, X2	---	
	Det		NP	---	
			Nominal, Noun	---	
				Prep	

## MÉTODOS DE PARSING

### o 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

On-line, palavra a palavra

	<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

## EXERCÍCIO

- o Em duplas, pelo método CKY, analisar a sentença "O copo quebrou."

### o REGRAS

- Sentença  $\rightarrow$  SN SV
- Sentença  $\rightarrow$  SV
- SN  $\rightarrow$  pronome
- SN  $\rightarrow$  substantivo
- SN  $\rightarrow$  artigo substantivo
- SV  $\rightarrow$  verbo
- SV  $\rightarrow$  verbo SN
- SV  $\rightarrow$  verbo SN SP
- SP  $\rightarrow$  preposição SN

### o LÉXICO

- artigo  $\rightarrow$  o | a | os | ...
- pronome  $\rightarrow$  eu | ele | ela | ...
- substantivo  $\rightarrow$  casa | brinquedo | copo | abacaxi | ...
- verbo  $\rightarrow$  correu | comprou | faça | quebrou | deu | ...
- preposição  $\rightarrow$  de | para | em | ...

## MÉTODOS DE PARSING

### o Algoritmo de Earley

- Constrói um *chart* (mapa/gráfico) das derivações
  - o 1º passo: define um chart de N+1 posições, em que N é o número de palavras da sentença
    - o Mesmo raciocínio: posição [0,1] indica palavra entre 0 e 1
      - Exemplo: *Book that flight.*

Chart



85

## MÉTODOS DE PARSING

### o Algoritmo de Earley

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

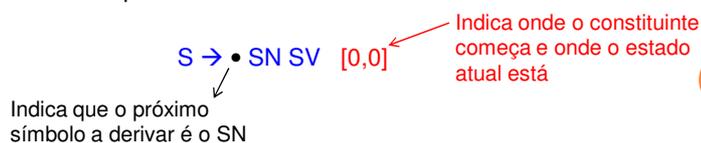
$S \rightarrow \bullet SN SV$  [0,0]

86

## 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

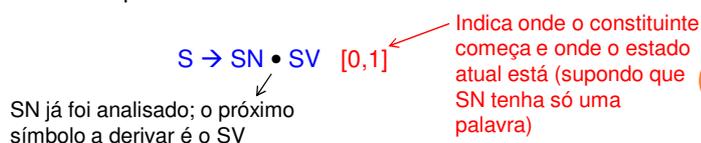


87

## 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



88

## MÉTODOS DE PARSING

### o Algoritmo de Earley

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

89

## MÉTODOS DE PARSING

### o Algoritmo de Earley

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

90

## 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 | ...

<sub>0</sub> O <sub>1</sub> copo <sub>2</sub> quebrou <sub>3</sub>

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 | ...

<sub>0</sub> O <sub>1</sub> copo <sub>2</sub> quebrou <sub>3</sub>

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 | ...

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

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 | ...

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

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 | ...

<sub>0</sub> O <sub>1</sub> copo <sub>2</sub> quebrou <sub>3</sub>

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 | ...

<sub>0</sub> O <sub>1</sub> copo <sub>2</sub> quebrou <sub>3</sub>

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
- Regra inicial completa!
- Fim do processo

○ 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] <b>S → SN SV • [0,3]</b> <b>S' → S • [0,3]</b>

## MÉTODOS DE PARSING

### ○ Algoritmo de Earley

- O processo tem sucesso se consome a regra inicial
- A partir do chart, é possível recuperar todas as árvores sintáticas possíveis
  - 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

103

## EXERCÍCIO PARA CASA

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

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

104

## 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 voo de São Paulo]<sub>SN</sub> [chegou]<sub>SV</sub>  
 [O voo [de [São Paulo]<sub>SN</sub>]<sub>SP</sub>]<sub>SN</sub> [chegou]<sub>SV</sub>  
 [O voo]<sub>SN</sub> [de]<sub>SP</sub> [São Paulo]<sub>SN</sub> [chegou]<sub>SV</sub>

105

## 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 voo]<sub>SN</sub> de [São Paulo]<sub>SN</sub> chegou

106

## 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

107

## 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

108

## PARSING PARCIAL

### ○ Abordagens

- **Aprendizado de máquina**

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

109

## 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]<sub>SN</sub> de [São Paulo]<sub>SN</sub> chegou

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

110