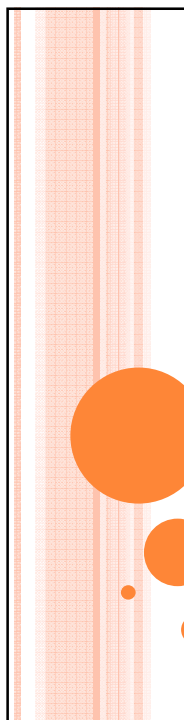


SINTAXE – PARTE 1

SCC5908 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.”
- 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

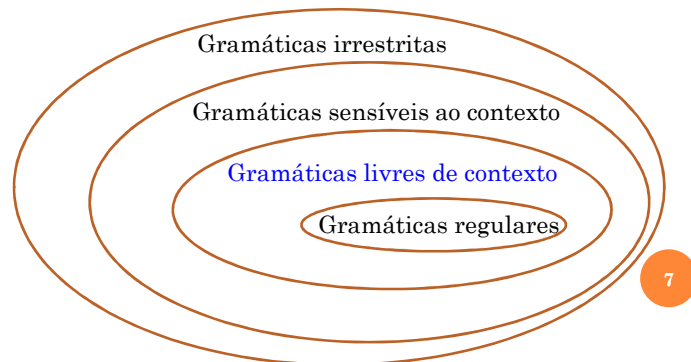
CONSTITUINTES

- 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]_{SN} [quebrou]_{SV}
 - [O vidro [de remédio]_{SP}]_{SN} [quebrou]_{SV}
 - [O vidro [de [remédio]_{SN}]_{SP}]_{SN} [quebrou]_{SV}
- **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

○ 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 | ...

9

GRAMÁTICA LIVRE DE CONTEXTO

○ Gramática pode ser usada para

- **Gerar** sentenças
- **Reconhecer** sentenças

○ **Derivação**

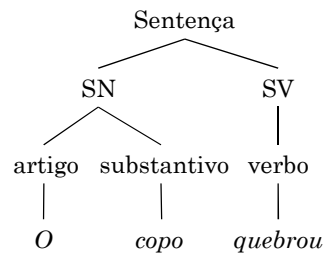
- Sequência de aplicação de regras da gramática
- Gera uma **árvore sintática** (*parse tree*)

10

GRAMÁTICA LIVRE DE CONTEXTO

○ Árvore sintática

- *O copo quebrou.*



• Terminologia

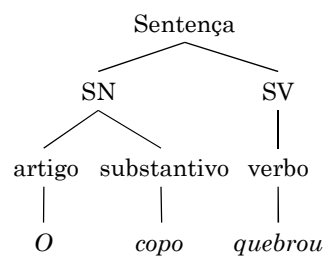
- Sentença **domina** todos os nós da árvore
- Sentença **domina imediatamente** SN e SV
- SN e SV são **filhos** de Sentença
- SN, SV, etc. são **descendentes** de Sentença

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

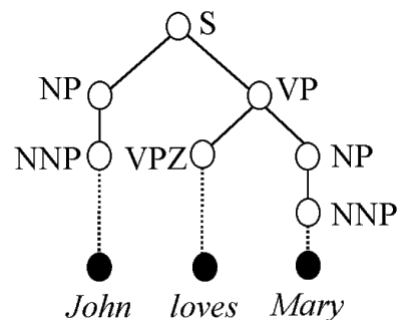
- Formato parentizado é comum
- Alguns treebanks contêm outras anotações
 - Predicados e argumentos
 - Funções gramaticais (sujeito, objeto, etc.)
 - Funções semânticas (local, tempo, etc.)
- Gramática da língua embutida nas análises
 - As árvores sintáticas podem ser a base para a construção de gramáticas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

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*

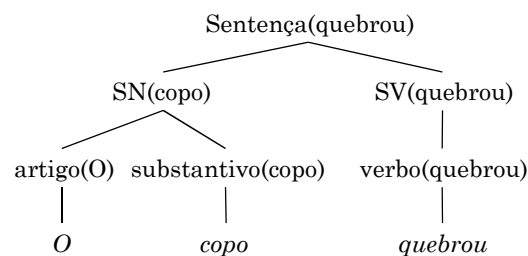
- Exemplo de regras para encontrar elemento mais importante de um SN para o inglês
 - Se última palavra é etiquetada como POS, retorne-a
Senão procure da direita para a esquerda pelo primeiro filho que é um NN, NNP, NNPS, NX, POS ou JJR
 - Senão procure da esquerda para a direita pelo primeiro filho 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*

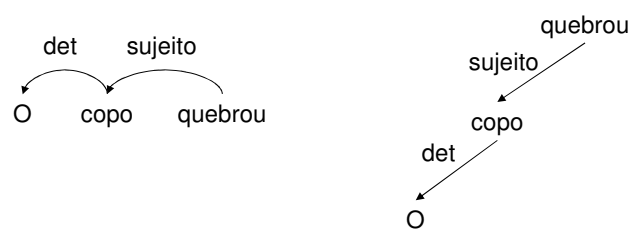
- Podem ser associadas a cada nó da árvore



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

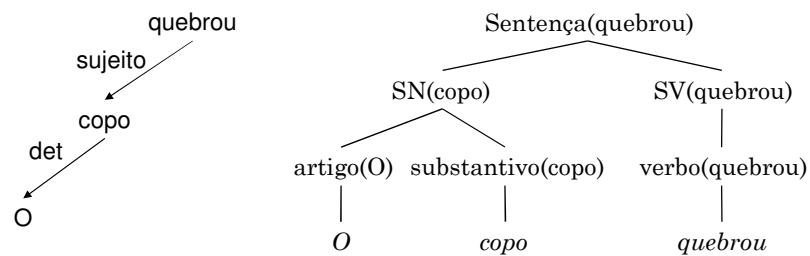
GRAMÁTICA DE DEPENDÊNCIA

- 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

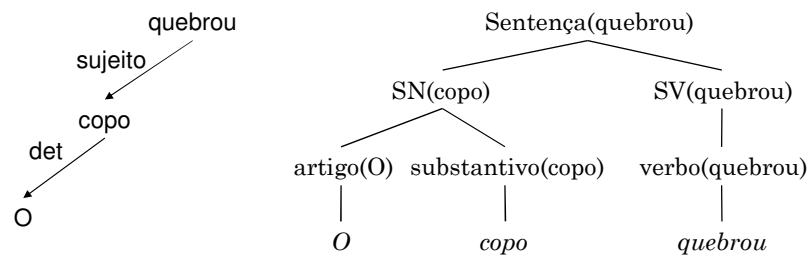
- Certa similaridade entre **análise de dependência** e *heads*



26

GRAMÁTICA DE DEPENDÊNCIA

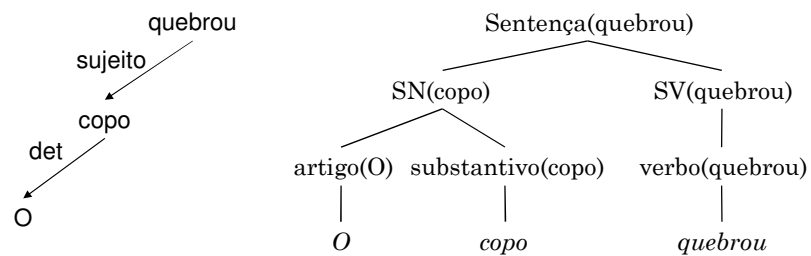
- É possível mapear uma estrutura de constituintes em uma de dependência? E o contrário?



27

GRAMÁTICA DE DEPENDÊNCIA

- É possível mapear uma estrutura de constituintes em uma de dependência? E o contrário?
 - Apenas parcialmente



28

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

29

GRAMÁTICAS E PROCESSAMENTO HUMANO

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

30

FORMALISMOS GRAMATICAIS

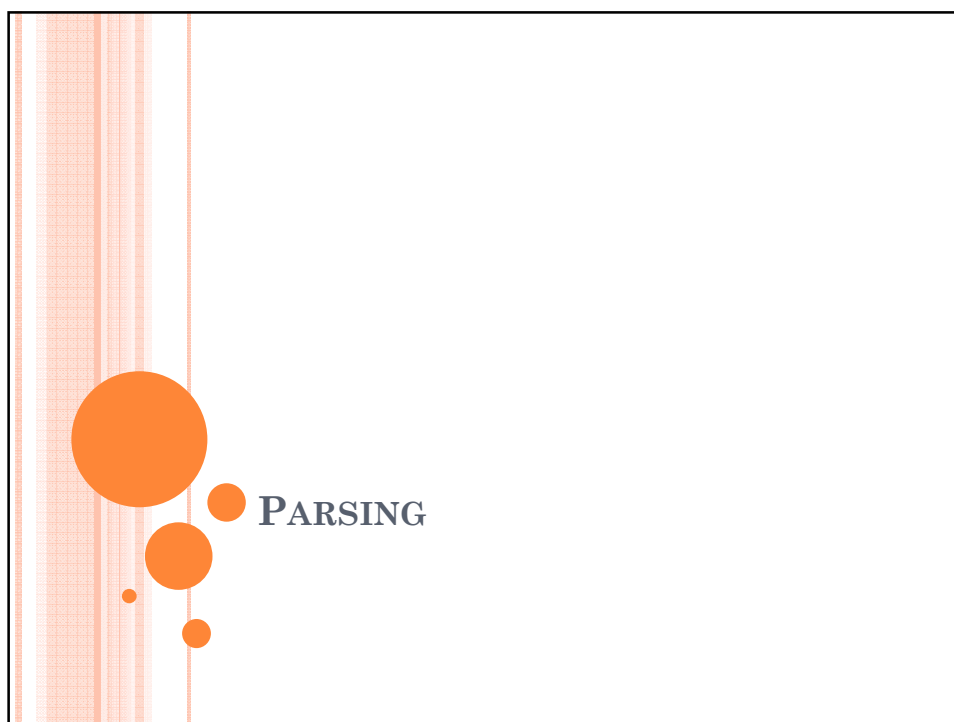
- DCG: *definite-clause grammar*
- LFG: *lexical functional grammar*
- GPSG: *generalized phrase structure grammar*
- HPSG: *head-driven phrase structure grammar*
- TAG: *tree adjoining grammar*
 - Árvores em vez de regras
- Etc.

31

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

32



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

35

PARSING

○ Análise top-down

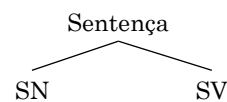
- Da raiz para as folhas (palavras)

○ 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

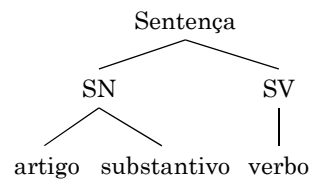
- Da raiz para as folhas (palavras)

○ 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

PARSING

○ Análise top-down

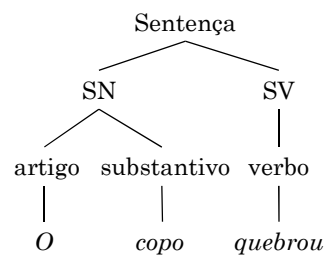
- Da raiz para as folhas (palavras)

○ 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

PARSING

○ Análise top-down

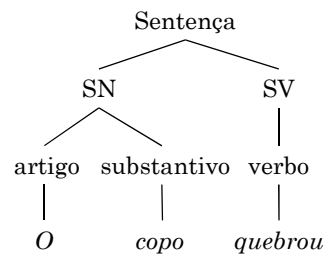
- Da raiz para as folhas (palavras)

○ 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 gramatical!
Mas se chega diretamente a ela?

39

PARSING

○ Análise top-down

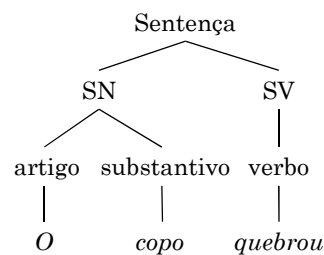
- Da raiz para as folhas (palavras)

○ 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 | ...



Não!
Qual o problema?

40

PARSING

○ Análise top-down

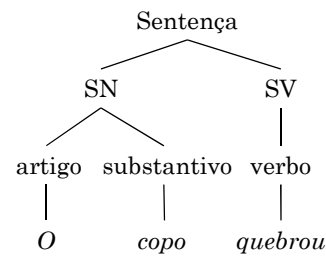
- Da raiz para as folhas (palavras)

○ 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!**

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 | ...

Sentença

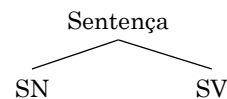
42

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 | ...



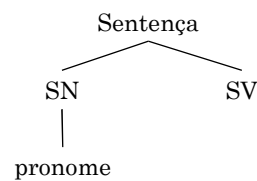
43

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

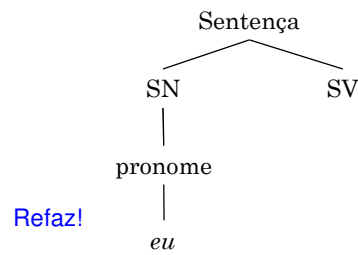
- Da raiz para as folhas (palavras)

○ 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

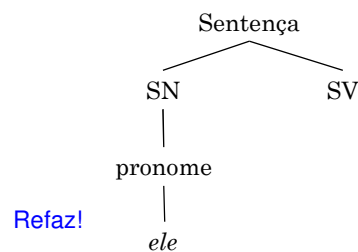
- Da raiz para as folhas (palavras)

○ 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

MUNDO REAL

PARSING

○ Análise top-down

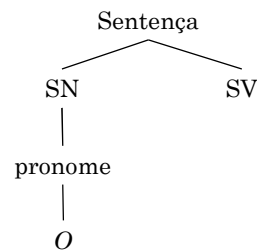
- Da raiz para as folhas (palavras)

○ 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 | ...



47

MUNDO REAL

PARSING

○ Análise top-down

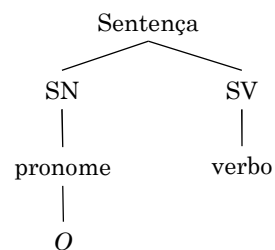
- Da raiz para as folhas (palavras)

○ 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 | ...



48

MUNDO REAL

PARSING

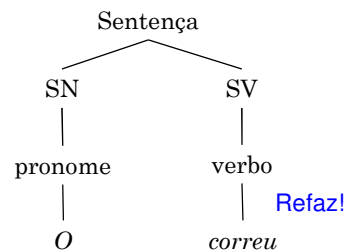
- Análise top-down
 - Da raiz para as folhas (palavras)

○ 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 | ...



49

MUNDO REAL

PARSING

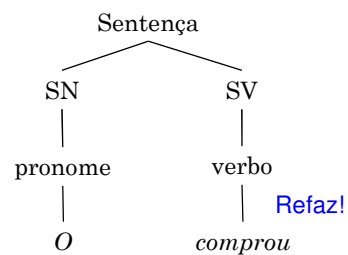
- Análise top-down
 - Da raiz para as folhas (palavras)

○ 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 | ...



50

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

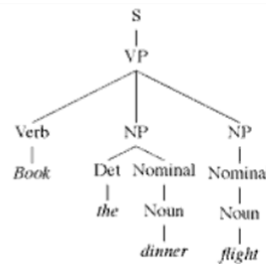
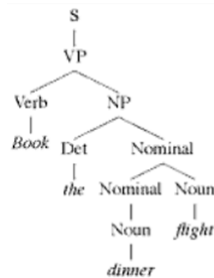
MUNDO REAL

PARSING

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

PIOR:

- mais de uma árvore é possível para uma mesma sentença
- uma mesma regra pode ser aplicada várias vezes



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 | ...

O copo quebrou

52

PARSING

○ Análise bottom-up

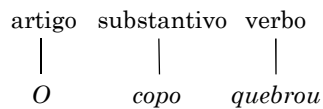
- Das folhas (palavras) para a raiz

○ 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 | ...



53

PARSING

○ Análise bottom-up

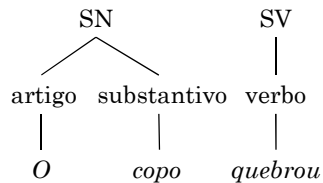
- Das folhas (palavras) para a raiz

○ 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 | ...



54

PARSING

○ Análise bottom-up

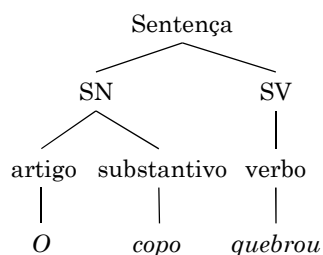
- Das folhas (palavras) para a raiz

○ 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 | ...



55

PARSING

○ Análise bottom-up

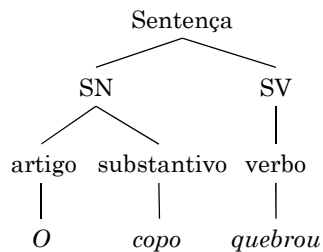
- Das folhas (palavras) para a raiz

○ 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?

56

PARSING

Mesmos problemas anteriores: regras improdutivas, mais de uma árvore, regras aplicadas várias vezes

○ Análise bottom-up

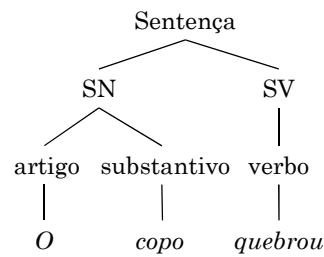
- Das folhas (palavras) para a raiz

○ 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 | ...



57

PARSING

○ 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

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

58

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,[,]).

59

PARSING

o Exercício para casa

- Testar gramática abaixo

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 | ...

60

AMBIGUIDADE

- Há vários tipos de ambiguidades que afetam o parsing
 - Etiquetas morfossintáticas
 - Book the flight!
 - Substantivo vs. verbo
 - Funções gramaticais
 - He gave her his book
 - Objeto direto vs. indireto

61

AMBIGUIDADE

- Há vários tipos de ambiguidades 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

62

AMBIGUIDADE

- Há vários tipos de ambiguidades 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?**

63

AMBIGUIDADE

- Há vários tipos de ambiguidades 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

64

MÉTODOS DE PARSING

- Aplicação sequencial de regras gramaticais
 - Estilos *top-down* ou *bottom-up* clássicos
 - Várias questões com que lidar

65

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)

66

MÉTODOS DE PARSING

o CKY: algoritmo de Cocke-Kasami-Younger

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

67

MÉTODOS DE PARSING

o Exemplo de gramática para o inglês

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

68

MÉTODOS DE PARSING

- Exemplo de gramática para o inglês

- Conversão da gramática para FNC
- Léxico não precisa ser convertido (não é mostrado)

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

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?

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

71

MÉTODOS DE PARSING

- CKY: algoritmo de Cocke-Kasami-Younger
 - **Segundo passo:** construir uma tabela/matriz quadrada de N linhas por N colunas
 - N é o número de palavras da sentença a se analisar
 - Cada posição indica uma palavra
 - *Book₁ the₂ flight₃ through₄ Houston₅*

72

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

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

Tabela vazia inicial

	Book	the	flight	through	Houston

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 → *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
→ primeiros constituintes

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

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

- 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
 - 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 FNC
 - Mais natural
 - Estilo bottom-up de análise
 - Pode ser on-line/por demanda, ou seja, analisar a sentença conforme as palavras aparecem

81

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					

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

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	---		S, VP, X2		
	Det		NP		
			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	---		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

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

EXERCÍCIO

- Pelo método CKY, analisar a sentença “O copo quebrou.”

○ 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 | ...