

*PROCESSAMENTO DE LINGUAGEM NATURAL*



**BABY ON BOARD**

# O DISCURSO DE YODA



# O DISCURSO DE YODA

- O que há de “errado” no discurso do personagem Yoda?
  - “Errado” ou “diferente”?



- Qual a importância desse tipo de informação?
- Qual o impacto para o PLN?

# SINTAXE

- Um dos níveis mais fundamentais para a língua
- Há tendências já identificadas pela ciência
  - Sujeito antes do objeto
  - Falantes preferem dependências entre palavras mais curtas, em média
- ... mas há grande variedade entre as línguas

# CHEGA A SER TÓPICO DE PESQUISA

**Universal Dependencies in a galaxy far, far away...**

**What makes Yoda's English truly alien**

**Natalia Levshina**

Leipzig University

IPF 141199 Nikolaistraße 8-10

04109 Leipzig

natalia.levshina@uni-leipzig.de

## **Abstract**

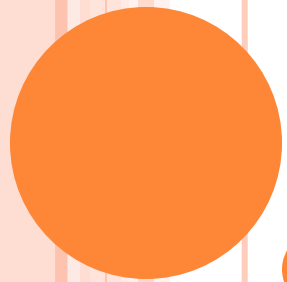
This paper investigates the word order used by Yoda, a character from the Star Wars universe. His clauses typically contain an Object, Oblique and/or non-finite part of the predicate followed by the subject and the finite predicate/auxiliary/copula, e.g. *Help you it will*. Using the sentences in Yodish from the scripts of the Star War films, this paper examines three cross-linguistically common tendencies, which can be explained by optimization of processing: the trade-off between entropy of S and O order and morphological cues, minimization of dependency lengths, and the tendency to place the verb in the end of a clause. For comparison, a standardized version of Yoda's sentences is used, as well as the Universal Dependencies corpora. The results of quantitative analyses indicate that Yodish is less adjusted to human processor's needs than standard English and other human languages.



# SINTAXE

*SCC5908 Introdução ao Processamento de Língua Natural*

*SCC0633 Processamento de Linguagem Natural*



# 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



# APLICAÇÃO

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

# CONSTITUINTES

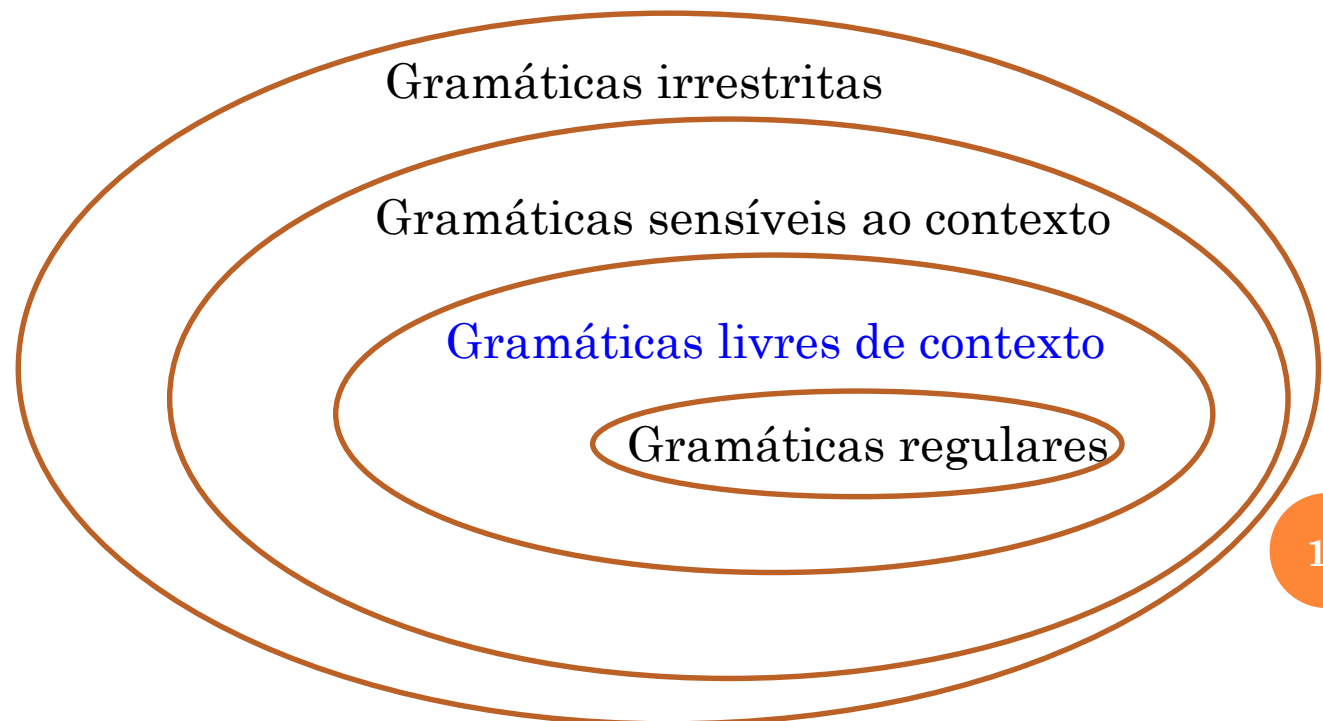
- Sintagma nominal
  - “Ela”, “João”, “a casa”, “o cavalo Pangaré”, “uma bela moça”
- Sintagma verbal
  - “Eu corri.”, “Ele precisa de uma passagem”, “João e Maria deram o livro para ela.”, “Faça!”
- Sintagma adjetival
  - “Ele é competente.”
- Sintagma adverbial
  - “Antigamente tudo era diferente.”,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

# 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 entre elementos
  - Subcategorização dos verbos
    - Preferências sintáticas
    - Alternativa: predicado e argumentos

# 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

# 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

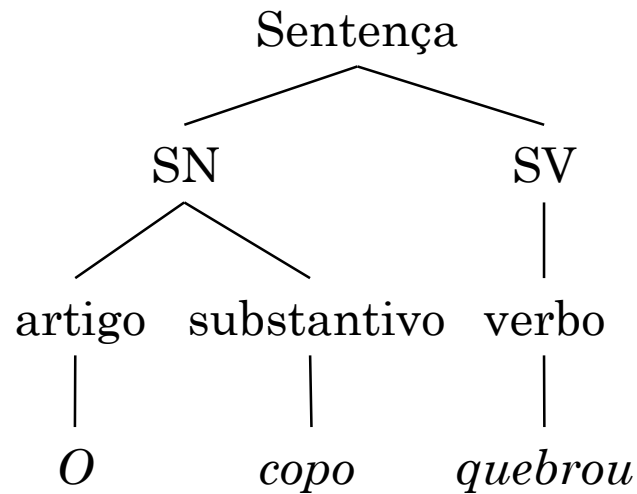
- artigo → o | a | os | ...
- pronome → eu | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

# 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*)

# GRAMÁTICA LIVRE DE CONTEXTO

- Árvore sintática
  - *O copo quebrou.*

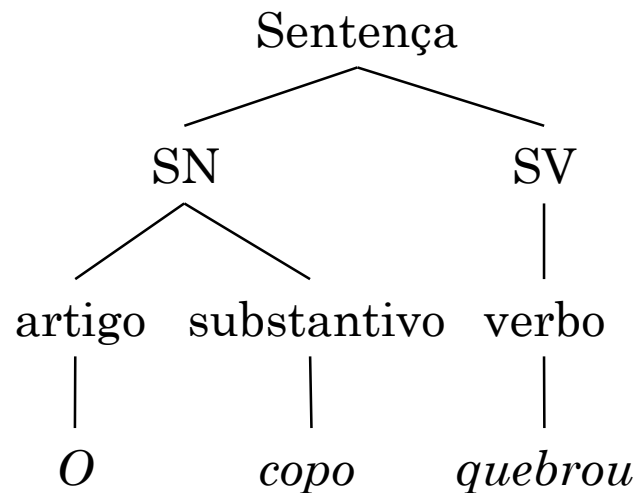


- Terminologia
  - Sentença **domina** todos os nós da árvore
  - Sentença **domina imediatamente** SN e SV
  - SN e SV são **filhos** de Sentença
  - SN, SV, etc. são **descendentes** de Sentença



# 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:  $[[O_{artigo} copo_{substantivo}]_{SN} [quebrou_{verbo}]_{SV}]_{Sentença}$

# 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  $N \rightarrow (N \cup T)^*$

# 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

# 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

# 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?

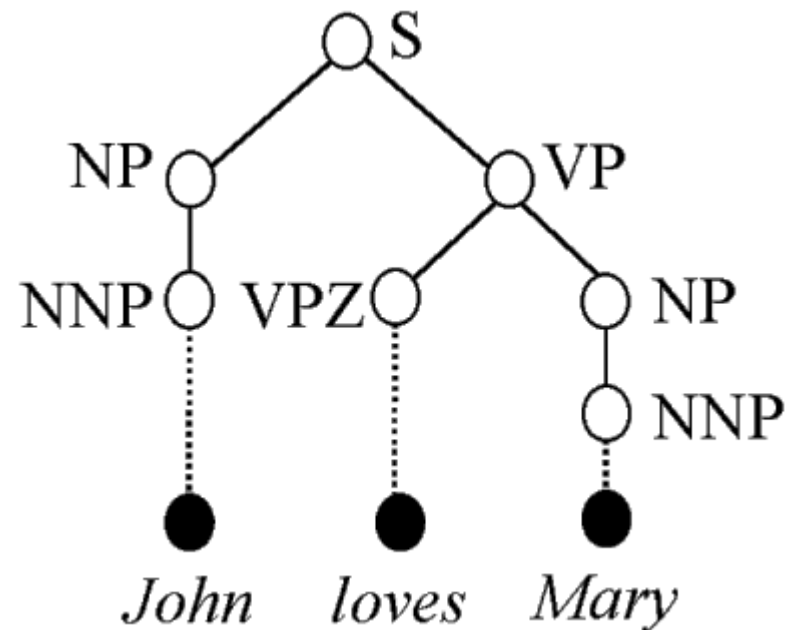
# TREEBANKS

- 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

# 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

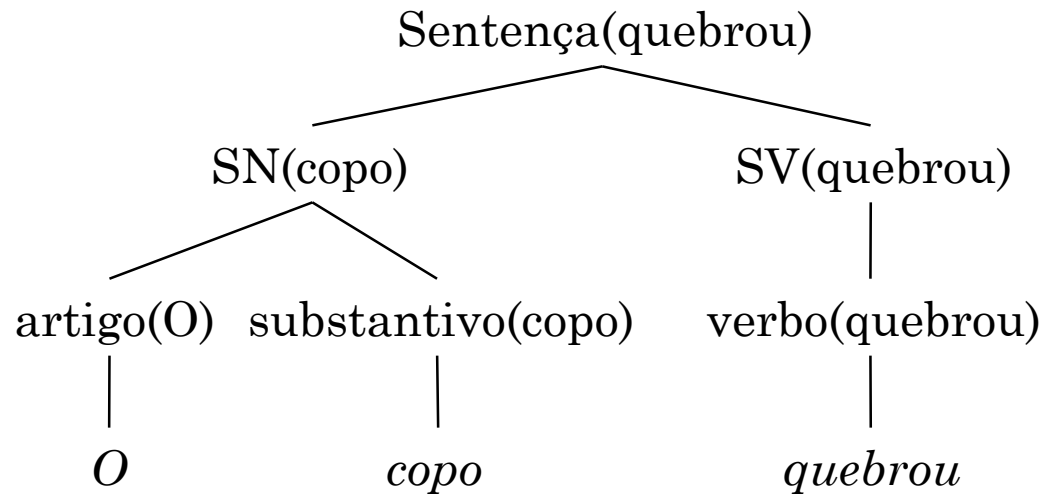
## ○ *Heads*

- 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

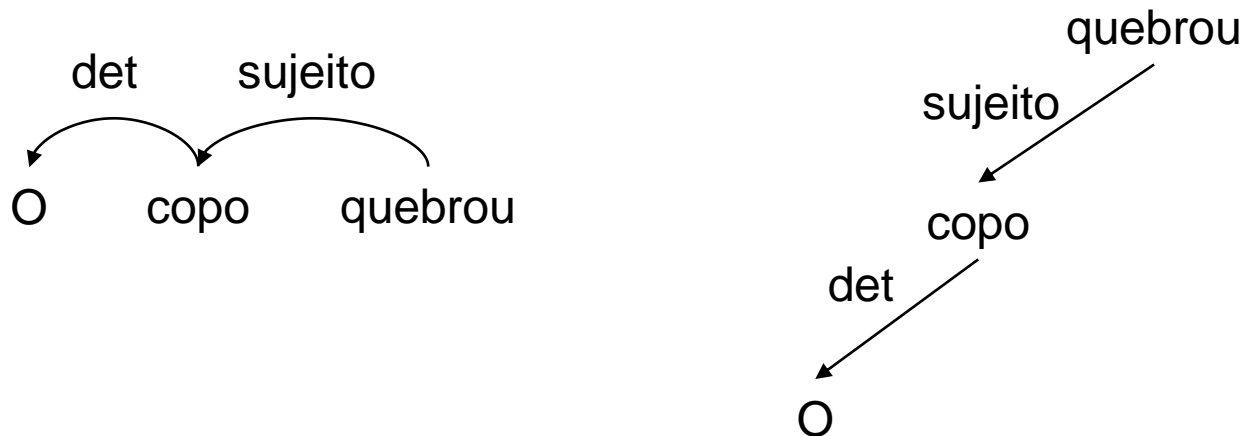
## ○ *Heads*

- Podem ser associadas a cada nó da árvore



# GRAMÁTICA DE DEPENDÊNCIA

- Alternativa à gramática de constituintes
- Foco nas **relações gramaticais**
  - Explicitamente rotuladas ou não
  - Sujeito, predicado, objetos, etc.



# GRAMÁTICA DE DEPENDÊNCIA

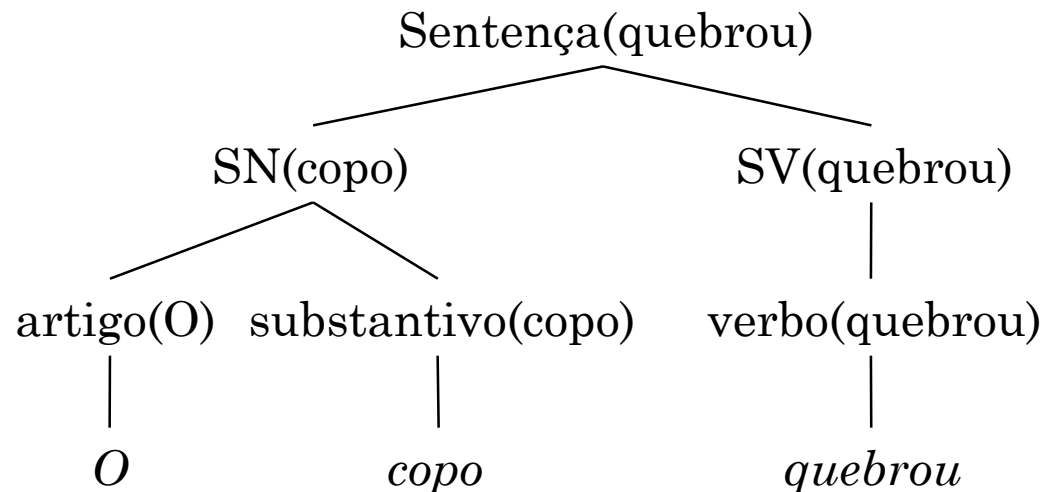
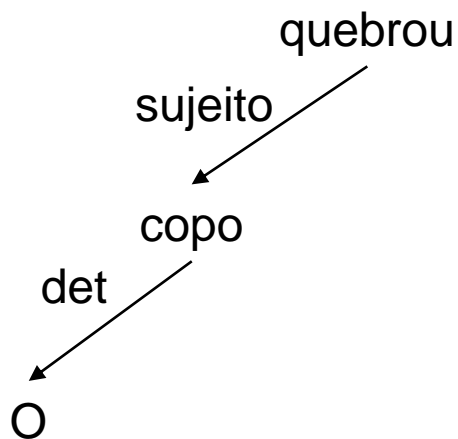
- Cada vez mais populares
- Vantagens?

# 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)

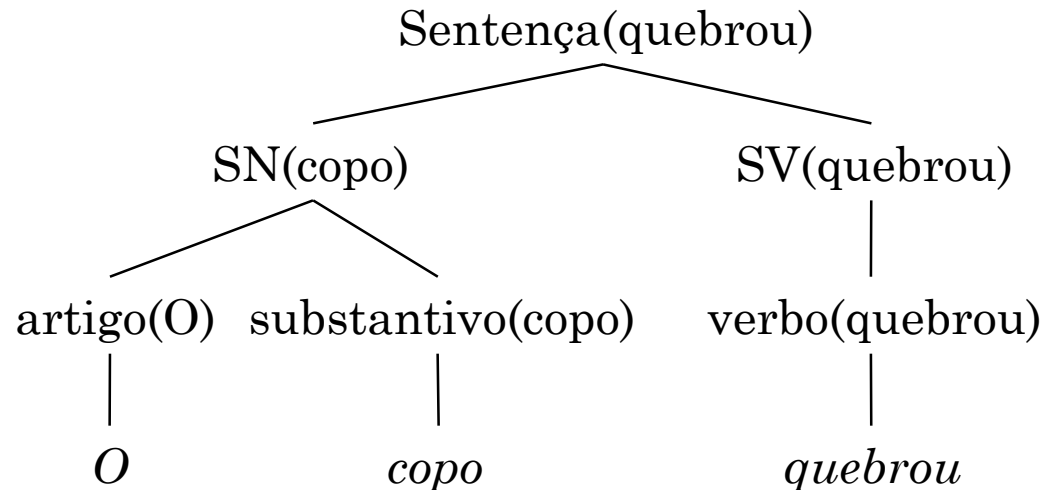
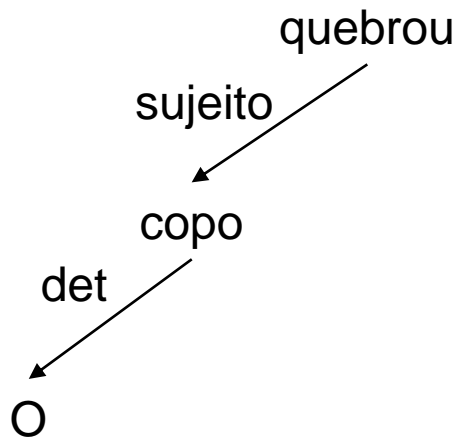
# GRAMÁTICA DE DEPENDÊNCIA

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



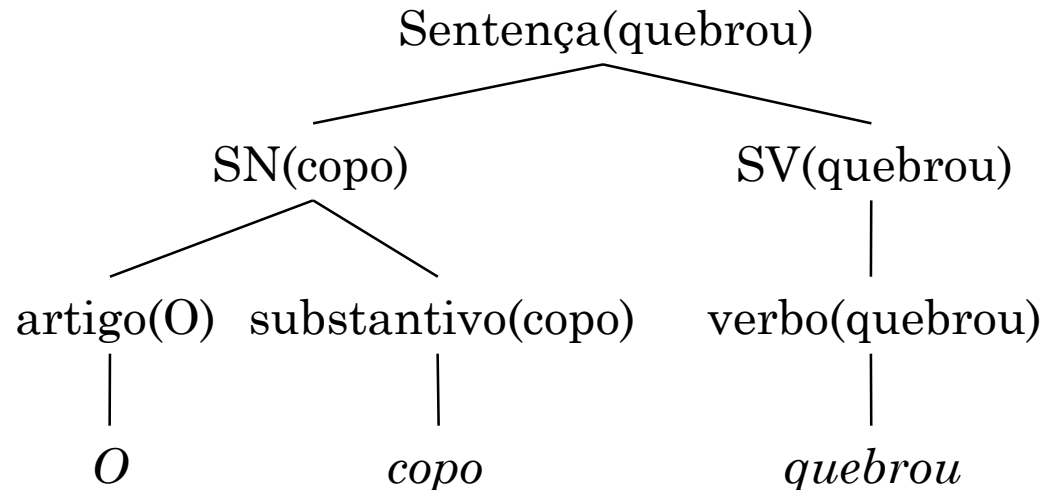
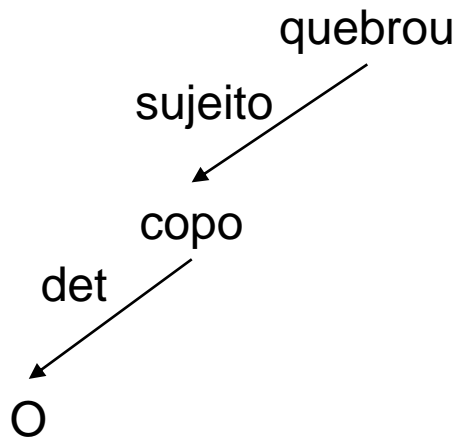
# GRAMÁTICA DE DEPENDÊNCIA

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



# GRAMÁTICA DE DEPENDÊNCIA

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





# 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”
- Processo **semi-automático**: aprendidas automaticamente, revisadas por humanos

# EXERCÍCIO

- Extrair regras sintáticas e léxico da receita de bolinho de chuva
  - *Bata as claras. Junte as gemas. Acrescente açúcar, o sal, o fermento e a farinha de trigo. Adicione leite. Bata novamente. Aqueça o óleo. Frite pequenas porções. Sirva com açúcar e canela.*
- Se precisar de ajuda, consulte o LX-Parser
  - <http://lxcenter.di.fc.ul.pt/services/pt/LXParserPT.htm>

1

# 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.

# 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

# 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].

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

# PARSING

- DCG e Prolog

- Teste em ambiente online da gramática anterior

- <http://swish.swi-prolog.org/>

# FALA

- Questão: a gramática da fala é similar à do texto escrito?

# FALA

- Gramática

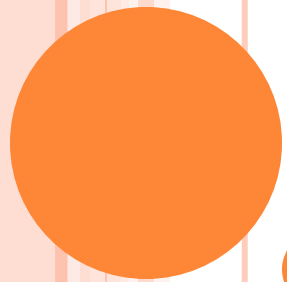
- Coisas em comum com língua escrita
- Muitos **outros fenômenos**
  - Pronomes são muito mais usados
  - Pequenos fragmentos de fala
  - Características próprias da fala
  - Disfluências: hesitação, pausa, reparo, recomeço, gagueira, etc.

- Também há treebanks importantes



# GRAMÁTICAS E PROCESSAMENTO HUMANO

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



# PARSING

# PARSING

## ○ Questão

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

# PARSING

## ○ Análise top-down

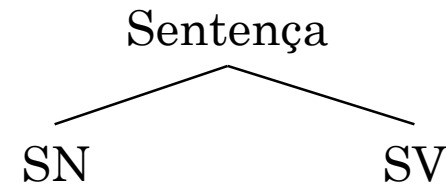
- Da raiz para as folhas (palavras)

## ○ 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 | ...



# PARSING

## ○ Análise top-down

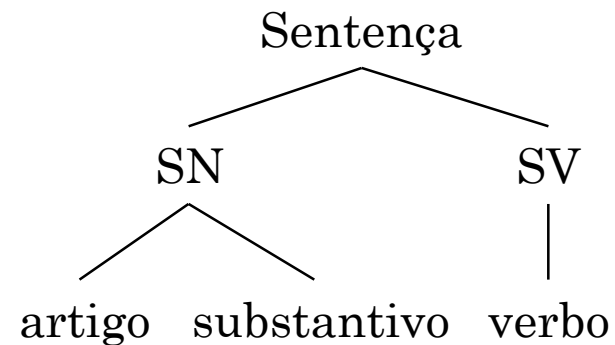
- Da raiz para as folhas (palavras)

## ○ 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 | ...



# PARSING

## ○ Análise top-down

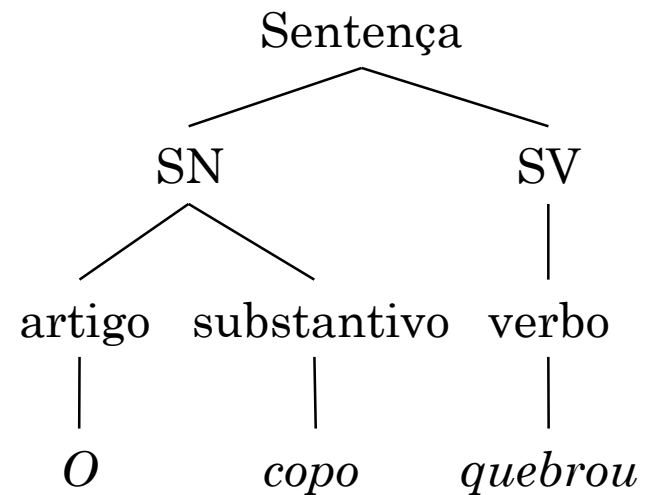
- Da raiz para as folhas (palavras)

### ○ 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 | ...



# PARSING

## ○ Análise top-down

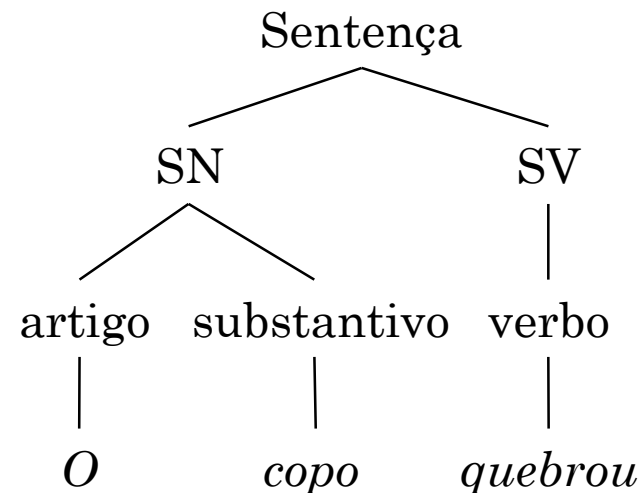
- Da raiz para as folhas (palavras)

## ○ 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?**



# PARSING

## ○ Análise top-down

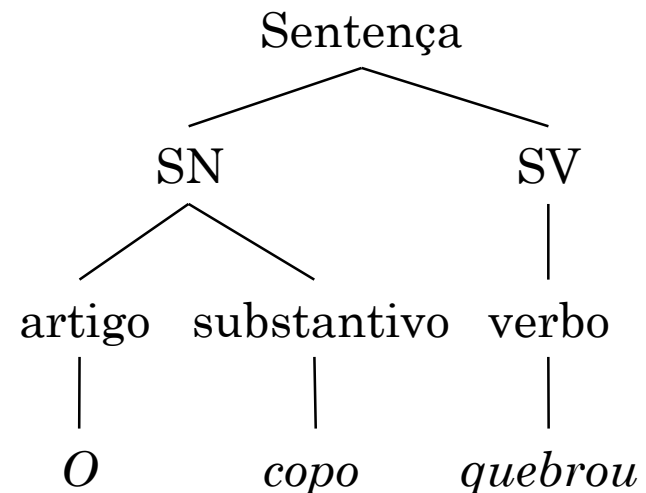
- Da raiz para as folhas (palavras)

## ○ 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?

# PARSING

## ○ Análise top-down

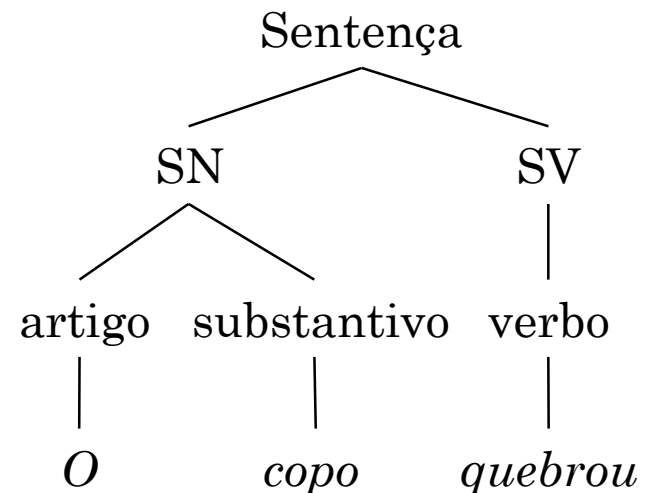
- Da raiz para as folhas (palavras)

## ○ 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!**

# 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

# PARSING

## ○ Análise top-down

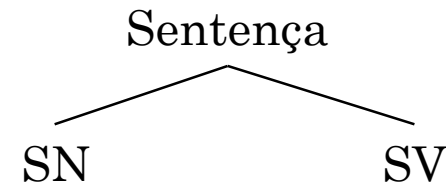
- Da raiz para as folhas (palavras)

## ○ 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 | ...



# PARSING

## ○ Análise top-down

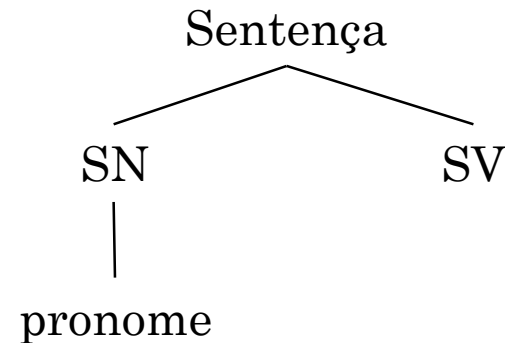
- Da raiz para as folhas (palavras)

## ○ 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 | ...



# PARSING

## ○ Análise top-down

- Da raiz para as folhas (palavras)

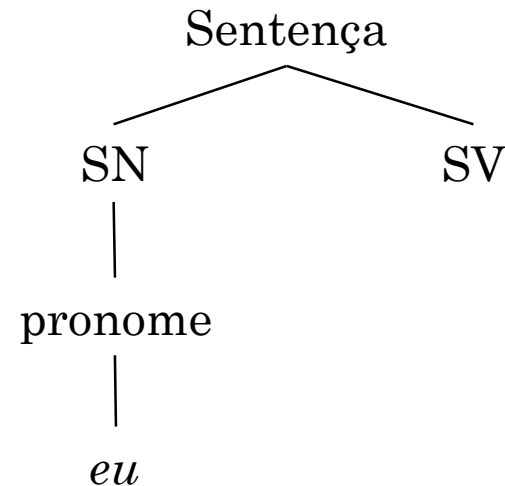
### ○ REGRAS

- 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!



# PARSING

## ○ Análise top-down

- Da raiz para as folhas (palavras)

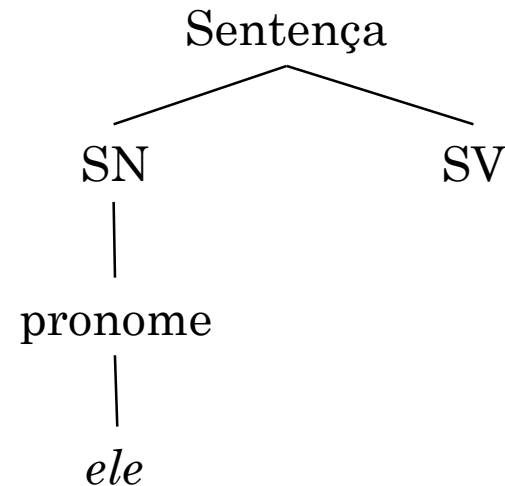
### ○ REGRAS

- 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!



# PARSING

## ○ Análise top-down

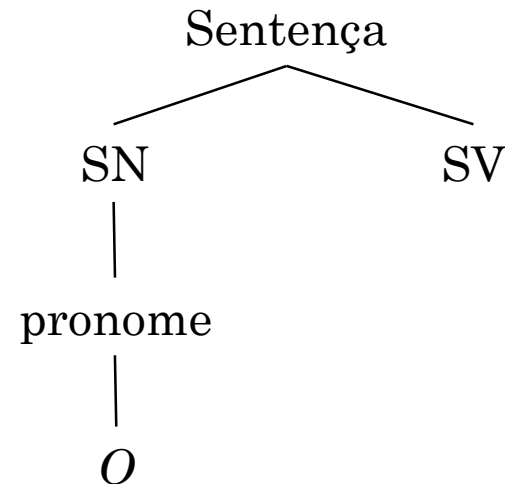
- Da raiz para as folhas (palavras)

## ○ 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 | ...





# PARSING

## ○ Análise top-down

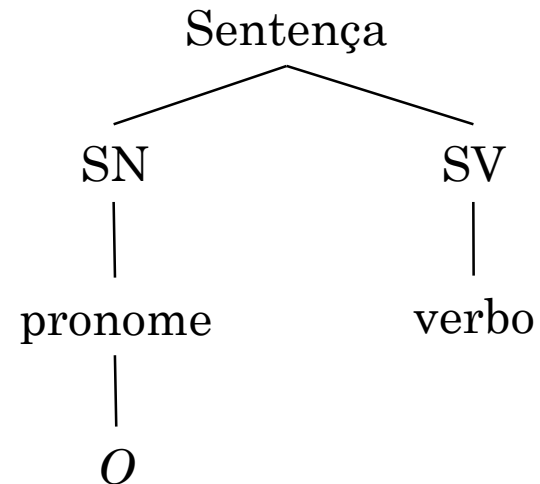
- Da raiz para as folhas (palavras)

## ○ 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 | ...



# PARSING

## ○ Análise top-down

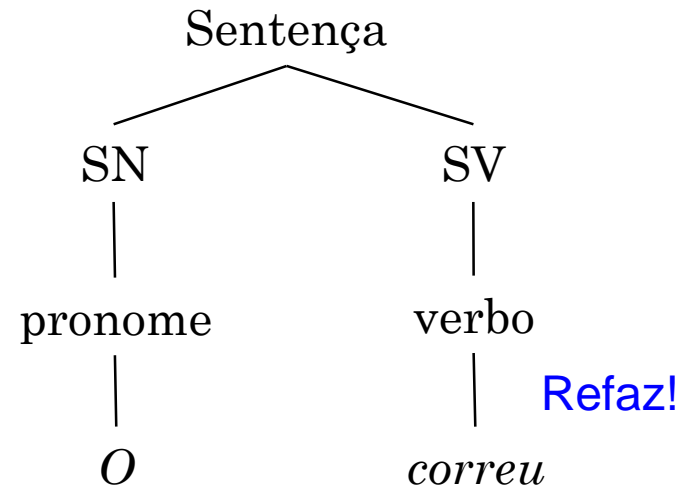
- Da raiz para as folhas (palavras)

### ○ 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 | ...



# 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

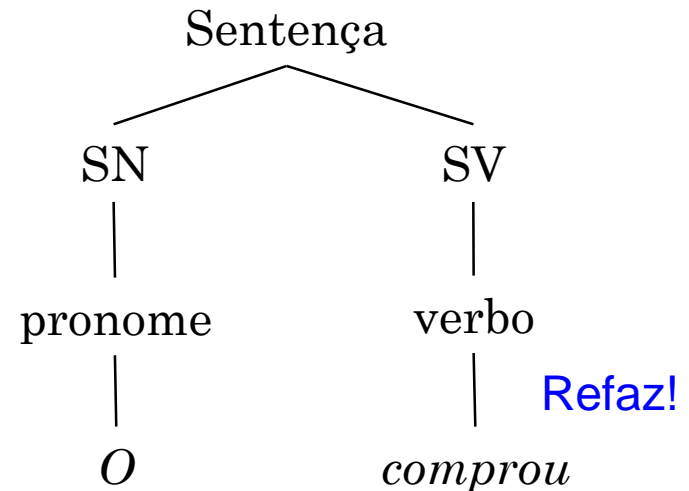
- artigo → o | a | os | ...
- pronome → eu | ele | ela | ...
- substantivo → casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

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

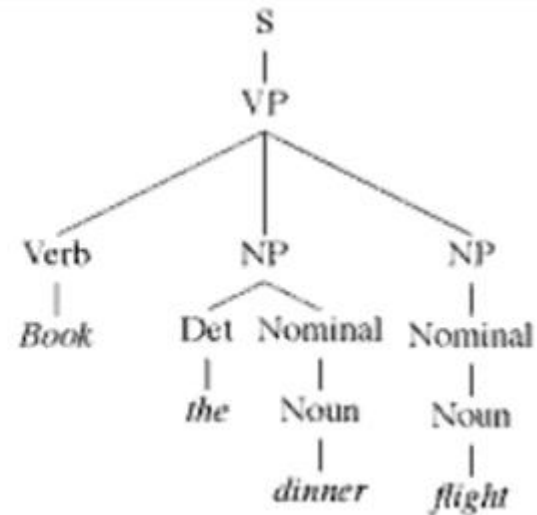
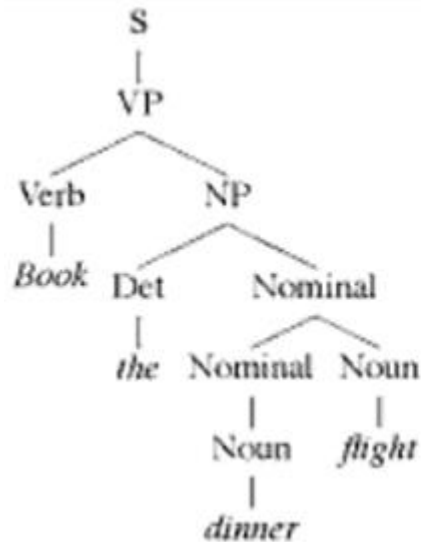


# PARSING

- 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



# 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*

# 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
<i>O</i>	<i>copo</i>	<i>quebrou</i>

# PARSING

## ○ Análise bottom-up

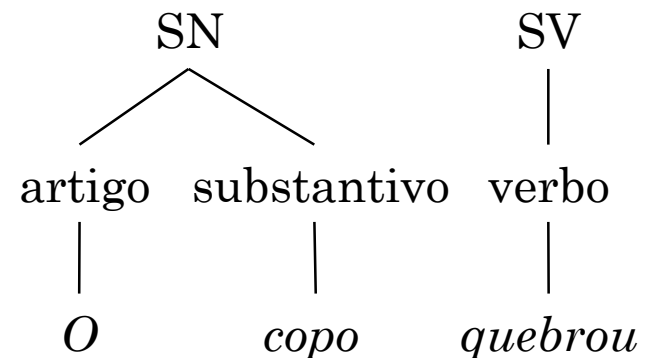
- Das folhas (palavras) para a raiz

## ○ 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 | ...



# PARSING

## ○ Análise bottom-up

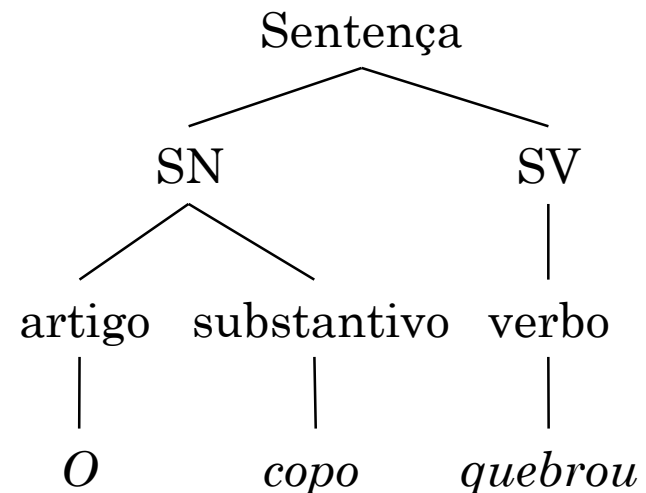
- Das folhas (palavras) para a raiz

## ○ 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 | ...





# PARSING

## ○ Análise bottom-up

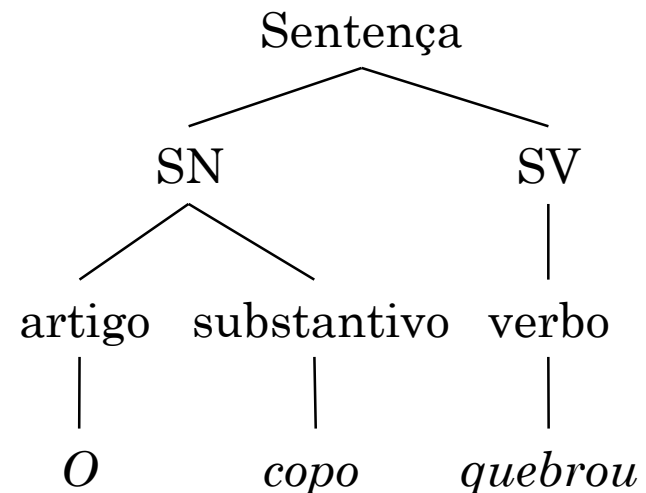
- Das folhas (palavras) para a raiz

## ○ 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?

# PARSING

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

## ○ Análise bottom-up

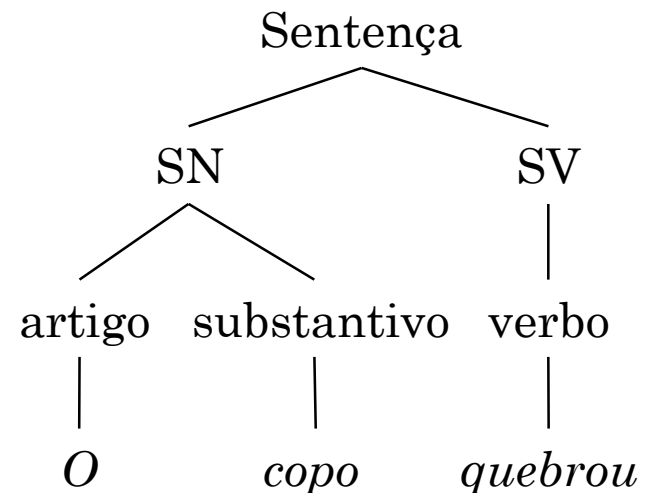
- Das folhas (palavras) para a raiz

## ○ 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 | ...



# 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

# AMBIGUIDADE

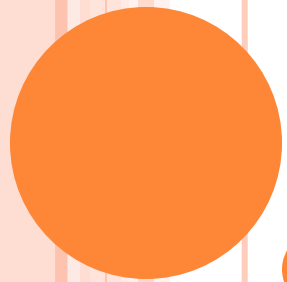
- Há vários tipos de ambiguidades que afetam o parsing
  - **Estrutural**
    - Ligação
      - *Ele viu a Torre Eiffel voando para Paris.*
        - viu ← voando para Paris
        - Torre Eiffel ← voando para Paris
    - Coordenação
      - Ele chamou amigos e amigas legais.
        - [amigos] e [amigas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

# 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

# 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?



# PARSING DE CONSTITUANTES

# MÉTODOS DE PARSING

## ○ Programação dinâmica

- Guarda em uma tabela (matriz) os constituintes já descobertos
  - Evita repetição de esforço
  - É possível recuperar todas as análises
- Método tradicional: CKY (1965)



# 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
        - $A \rightarrow a B \rightarrow A \rightarrow X B, X \rightarrow a$
        - $A \rightarrow B C D \rightarrow A \rightarrow X D, X \rightarrow B C$
      - União de produções
        - $A \rightarrow B C, B \rightarrow D, D \rightarrow E \rightarrow A \rightarrow E C$
      - Aplicação sistemática/recursiva dos procedimentos acima

# MÉTODOS DE PARSING

- Exemplo de gramática para o inglês

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

### Lexicon

*Det* → *that* | *this* | *a*  
*Noun* → *book* | *flight* | *meal* | *money*  
*Verb* → *book* | *include* | *prefer*  
*Pronoun* → *I* | *she* | *me*  
*Proper-Noun* → *Houston* | *NWA*  
*Aux* → *does*  
*Preposition* → *from* | *to* | *on* | *near* | *through*

### • Conversão da gramática para FNC

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

- Léxico não precisa ser convertido

# 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

# 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
      - *Book<sub>1</sub> the<sub>2</sub> flight<sub>3</sub> through<sub>4</sub> Houston<sub>5</sub>*

# 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

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

Lexicon	
<i>Det</i>	→ <i>that</i>   <i>this</i>   <i>a</i>
<i>Noun</i>	→ <i>book</i>   <i>flight</i>   <i>meal</i>   <i>money</i>
<i>Verb</i>	→ <i>book</i>   <i>include</i>   <i>prefer</i>
<i>Pronoun</i>	→ <i>I</i>   <i>she</i>   <i>me</i>
<i>Proper-Noun</i>	→ <i>Houston</i>   <i>NWA</i>
<i>Aux</i>	→ <i>does</i>
<i>Preposition</i>	→ <i>from</i>   <i>to</i>   <i>on</i>   <i>near</i>   <i>through</i>

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

Tabela vazia inicial

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

+ léxico

# MÉTODOS DE PARSING

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

Lexicon	
<i>Det</i>	→ <i>that</i>   <i>this</i>   <i>a</i>
<i>Noun</i>	→ <i>book</i>   <i>flight</i>   <i>meal</i>   <i>money</i>
<i>Verb</i>	→ <i>book</i>   <i>include</i>   <i>prefer</i>
<i>Pronoun</i>	→ <i>I</i>   <i>she</i>   <i>me</i>
<i>Proper-Noun</i>	→ <i>Houston</i>   <i>NWA</i>
<i>Aux</i>	→ <i>does</i>
<i>Preposition</i>	→ <i>from</i>   <i>to</i>   <i>on</i>   <i>near</i>   <i>through</i>

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

+ léxico

Começando pela diagonal principal  
→ etiquetas possíveis das palavras

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

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

- 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

	<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

- 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

	<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

## ○ 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

*Book the flight through Houston*

S, VP, Verb, Nominal, Noun	---	S, VP, X2	---	S, VP, X2
	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

# MÉTODOS DE PARSING

- 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

On-line, palavra a palavra

*Book the flight through Houston*

S, VP, Verb, 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	---			
	Det			

# MÉTODOS DE PARSING

- 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

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

	<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	---	
				Prep	

# 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	---	S, VP, X2
	Det	NP	---	NP
		Nominal, Noun	---	Nominal
			Prep	PP
				NP, proper noun

# EXERCÍCIO

- Pelo método CKY, analisar a sentença “O pai comprou o brinquedo.”

## ○ 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 → pai | casa | brinquedo | copo | abacaxi | ...
- verbo → correu | comprou | faça | quebrou | deu | ...
- preposição → de | para | em | ...

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

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

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

# PARSING PARCIAL

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

[O vôo]<sub>SN</sub> de [São Paulo]<sub>SN</sub> chegou

# 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



# PARSING PARCIAL

## ○ Abordagens

- **Aprendizado de máquina**

- Classificação sequencial, da esquerda para a direita
- Exige treinamento: portanto, cópua 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

# PARSING PARCIAL

- **Abordagens**

- **Aprendizado de máquina**

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

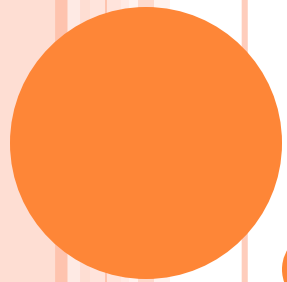
# PARSING PARCIAL

## ○ Esquema de anotação

- Esquema **IOB** para marcação de cópús (também usado para outros fins)
  - 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



# PARSING PROBABILÍSTICO

# ESTATÍSTICA

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

# EXEMPLO DE GLCP

## ○ REGRAS

- Sentença → SN SV [0.80]
- Sentença → SV [0.20]
- SN → pronome [0.50]
- SN → substantivo [0.15]
- SN → artigo substantivo [0.35]
- SV → verbo [0.40]
- SV → verbo SN [0.40]
- SV → verbo SN SP [0.20]
- SP → preposição SN [1.00]

## ○ LÉXICO

- artigo → o [0.20] | a [0.20] | os [0.15] | ...
- Etc.

# GLCP

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

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

# GLCP

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



# GLCP

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

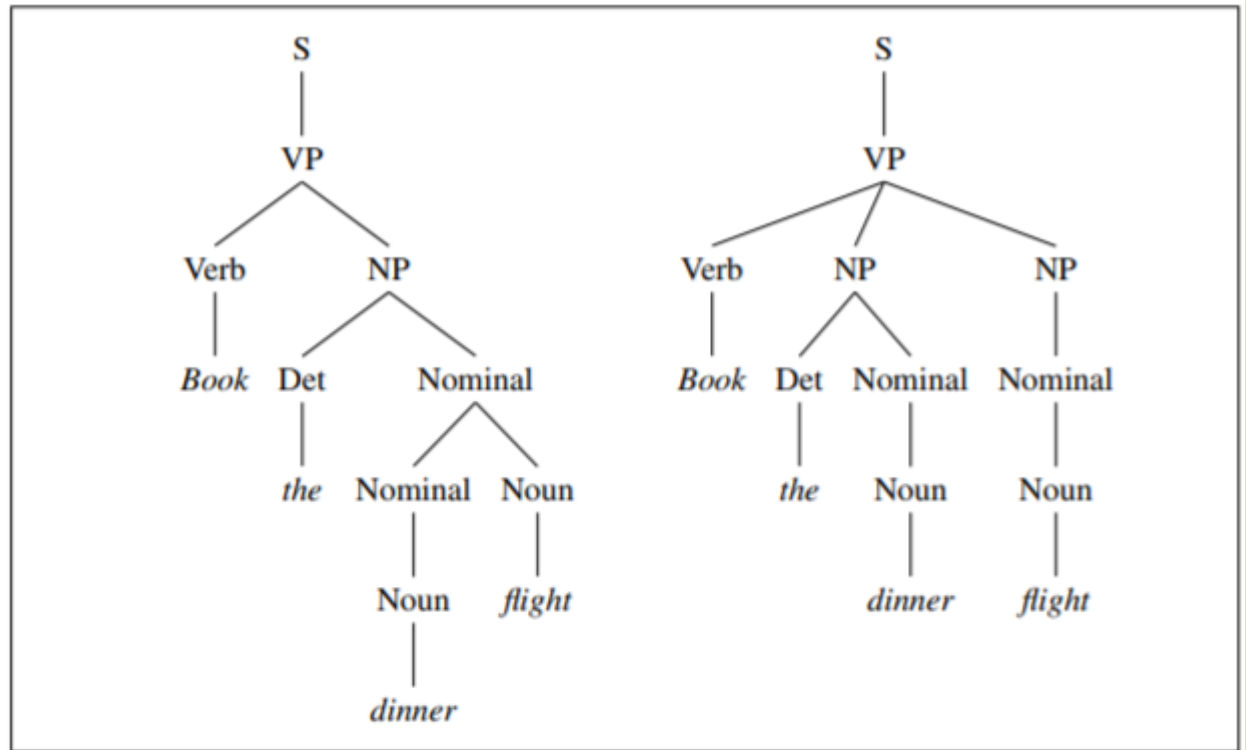
$$P(\textit{sentença}, \textit{árvore}) = \prod_{i=1}^n P(LDR_i | LER_i)$$

# GLCP

## EXEMPLO

Qual a correta?

O que significam?



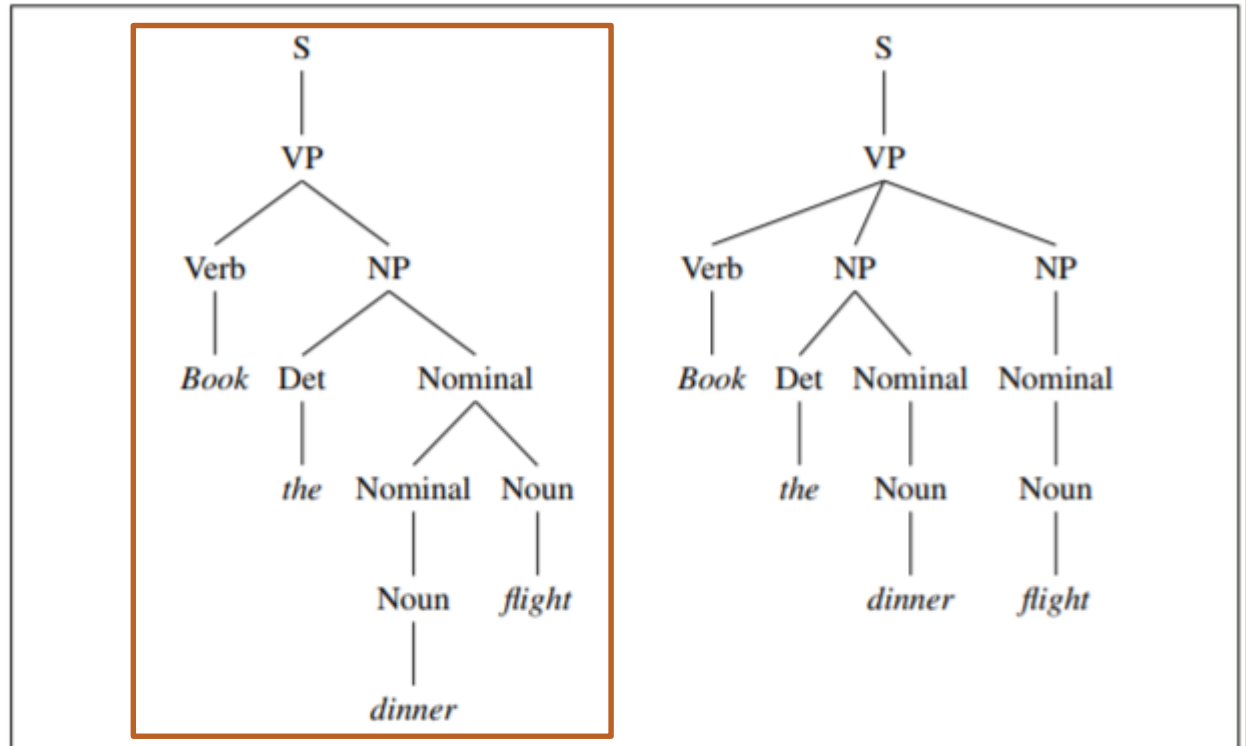
Rules			Rules		
	Rules	P		Rules	P
S	→ VP	.05	S	→ VP	.05
VP	→ Verb NP	.20	VP	→ Verb NP NP	.10
NP	→ Det Nominal	.20	NP	→ Det Nominal	.20
Nominal	→ Nominal Noun	.20	NP	→ Nominal	.15
Nominal	→ Noun	.75	Nominal	→ Noun	.75
Verb	→ book	.30	Nominal	→ Noun	.75
Det	→ the	.60	Verb	→ book	.30
Noun	→ dinner	.10	Det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

# GLCP

## EXEMPLO

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

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



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

# PARSING PROBABILÍSTICO

- É simples estender o método CKY para considerar probabilidades
  - Podem-se guardar todas ou somente as melhores análises

# PARSING PROBABILÍSTICO

- É simples estender o método CKY para considerar probabilidades
  - Podem-se guardar todas ou somente as melhores análises

## Trecho de uma gramática

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

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

# PARSING PROBABILÍSTICO

- Lindo! Mas...

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

# PARSING PROBABILÍSTICO

- Aprendizado de probabilidades
  - Quando há um treebank

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

- Exemplo hipotético

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

# PARSING PROBABILÍSTICO

- Aprendizado de probabilidades
  - E se não há um treebank?



# TAREFAS

- Leitura obrigatória da semana
  - *Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection*. In the Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020).
    - No e-Disciplinas
- Provinha 12 disponível à tarde no e-Disciplinas