2.7 语法解析（上）

林洲汉
上海交大电院

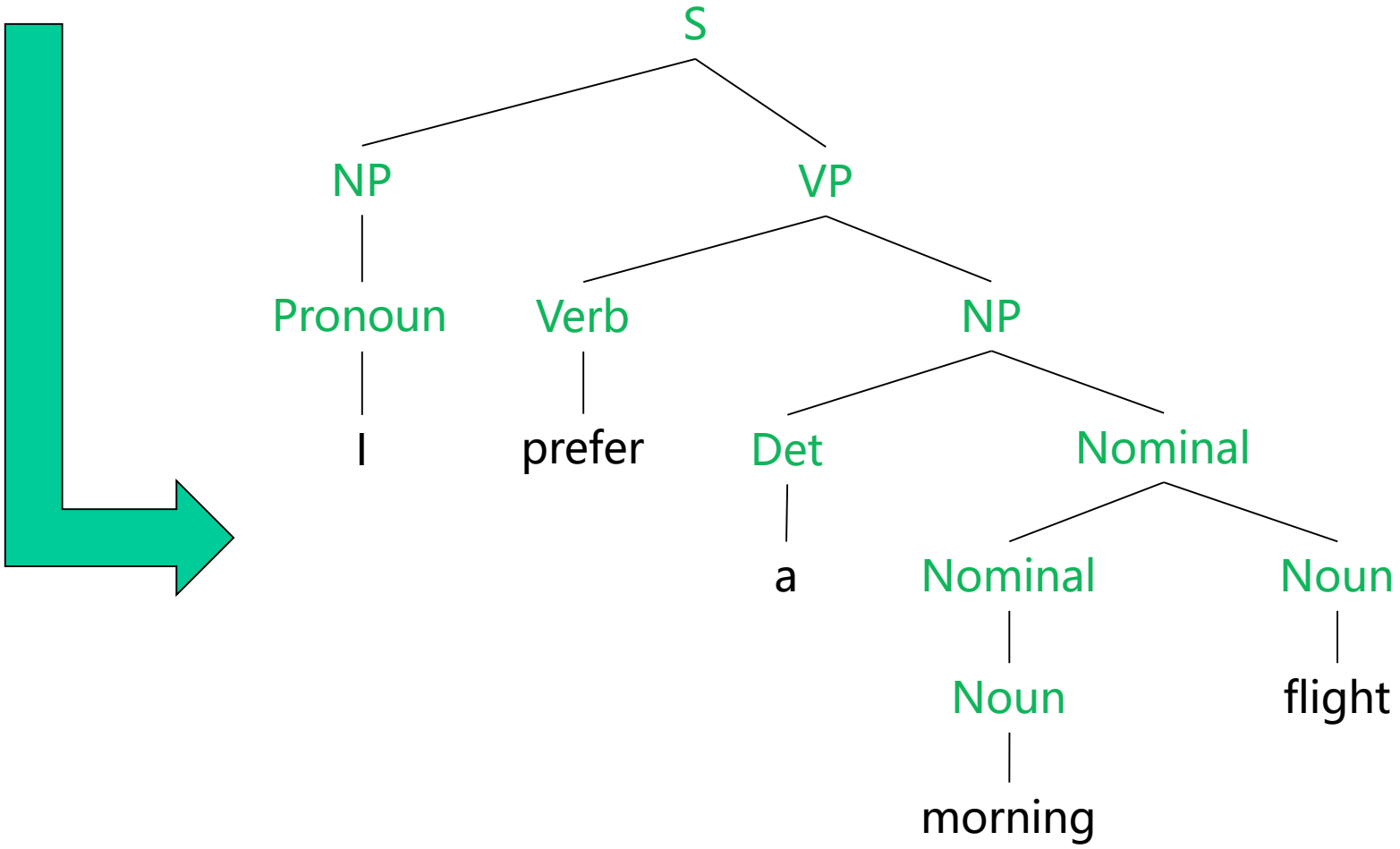
2023年秋季学期

目录

目录

构成式语法（constituency grammar）简介

I prefer a morning flight.



构成式语法（constituency grammar）简介

语法成分（constituent）

句子中的一组词，作为整体可以当做一个单独的语法单元

例如：名词性短语（NP），动词性短语（VP）……

语法规则（rules）

一组描述某个语法成分可以由什么组成的规则。

例如：

NP → Det Nominal（名词性短语可以由冠词加名词构成）
NP → ProperNoun（名词性短语可以专有名词构成）

词典（lexicon）

一组描述某个语法成分可以由什么词来构成的规则。

例如：

Det → a | an | the
Noun → flight | duck | paper

构成式语法（constituency grammar）简介

语法成分（constituent）

S(句子)
NP(名词性短语)
VP(动词性短语)
Pronoun(代词)
Proper-Noun(专有名词)
Det(冠词)
Nominal(名词性成分)
Verb(动词)
PP(介词短语)
Preposition(介词)

语法规则（rules）

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

词典（lexicon）

Noun	→ flights | breeze | trip | morning	9个元素
Verb	→ is | prefer | like | need | want | fly	
Adjective	→ cheapest | non-stop | first | latest | other | direct	
Pronoun	→ me | I | you | it	
Proper-Noun	→ Alaska | Baltimore | Los Angeles | Chicago | United | American	
Determiner	→ the | a | an | this | these | that	
Preposition	→ from | to | on | near	
Conjunction	→ and | or | but	

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition

S

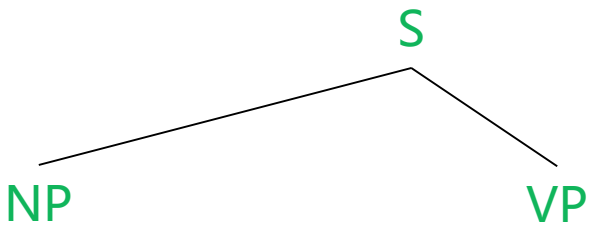
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun	→ flights | breeze | trip | morning
Verb	→ is | prefer | like | need | want | fly
Adjective	→ cheapest | non-stop | first | latest | other | direct
Pronoun	→ me | I | you | it
Proper-Noun	→ Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner	→ the | a | an | this | these | that
Preposition	→ from | to | on | near
Conjunction	→ and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



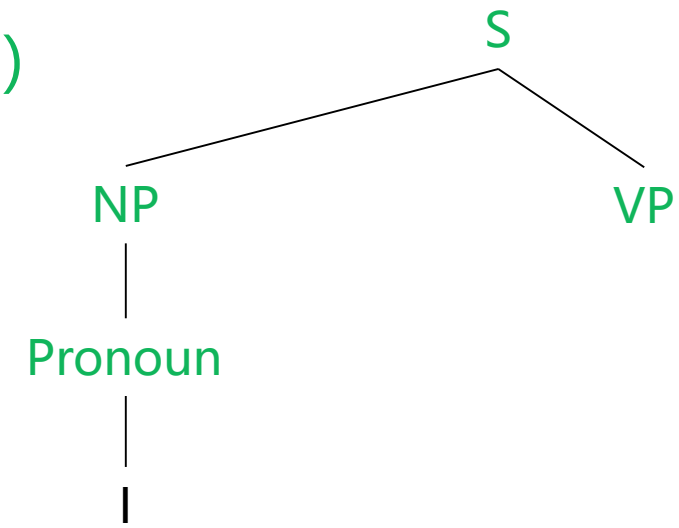
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun	→ flights | breeze | trip | morning
Verb	→ is | prefer | like | need | want | fly
Adjective	→ cheapest | non-stop | first | latest | other | direct
Pronoun	→ me | I | you | it
Proper-Noun	→ Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner	→ the | a | an | this | these | that
Preposition	→ from | to | on | near
Conjunction	→ and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



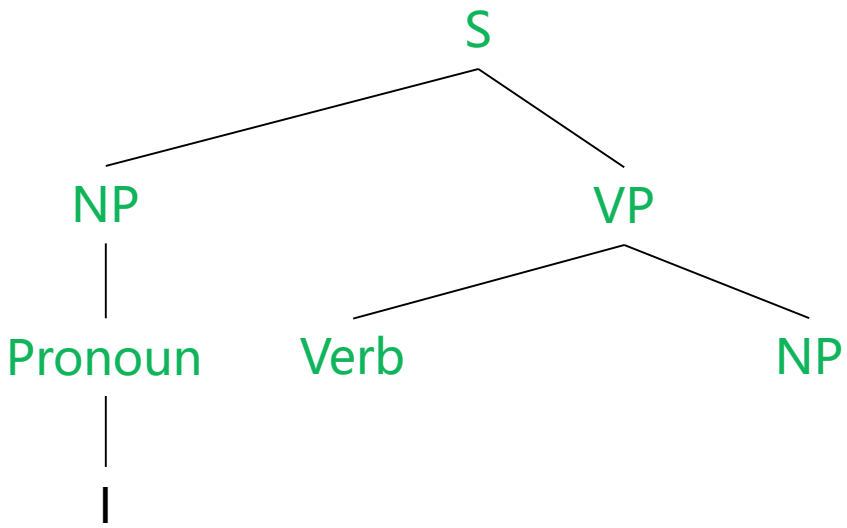
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun	→ flights | breeze | trip | morning
Verb	→ is | prefer | like | need | want | fly
Adjective	→ cheapest | non-stop | first | latest | other | direct
Pronoun	→ me | I | you | it
Proper-Noun	→ Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner	→ the | a | an | this | these | that
Preposition	→ from | to | on | near
Conjunction	→ and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



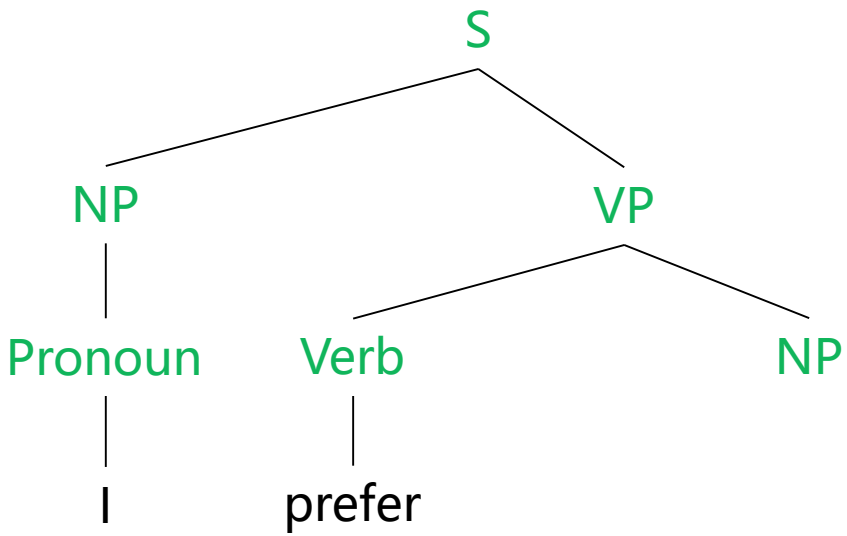
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun	→ flights | breeze | trip | morning
Verb	→ is | prefer | like | need | want | fly
Adjective	→ cheapest | non-stop | first | latest | other | direct
Pronoun	→ me | I | you | it
Proper-Noun	→ Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner	→ the | a | an | this | these | that
Preposition	→ from | to | on | near
Conjunction	→ and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



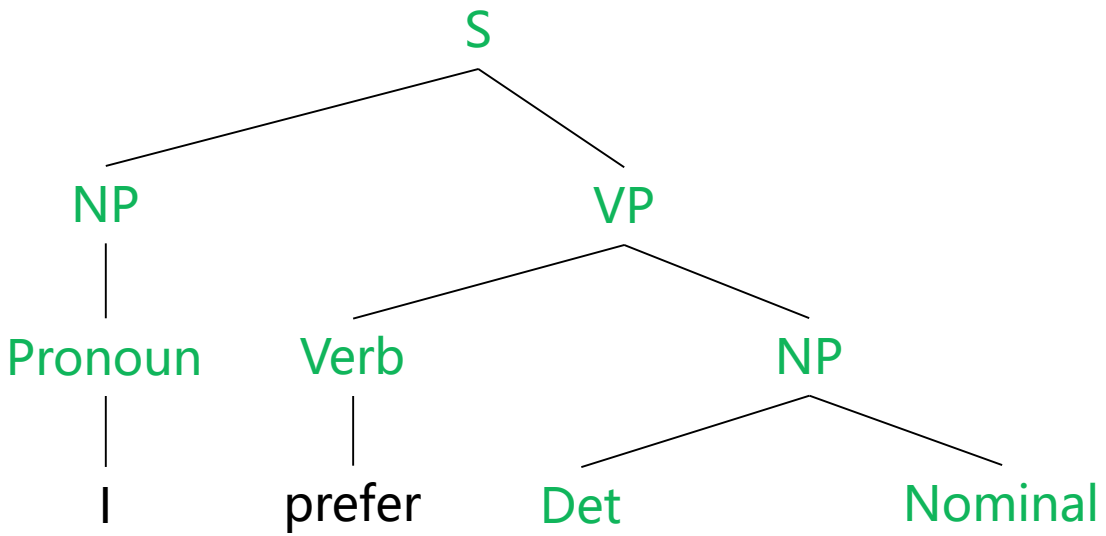
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



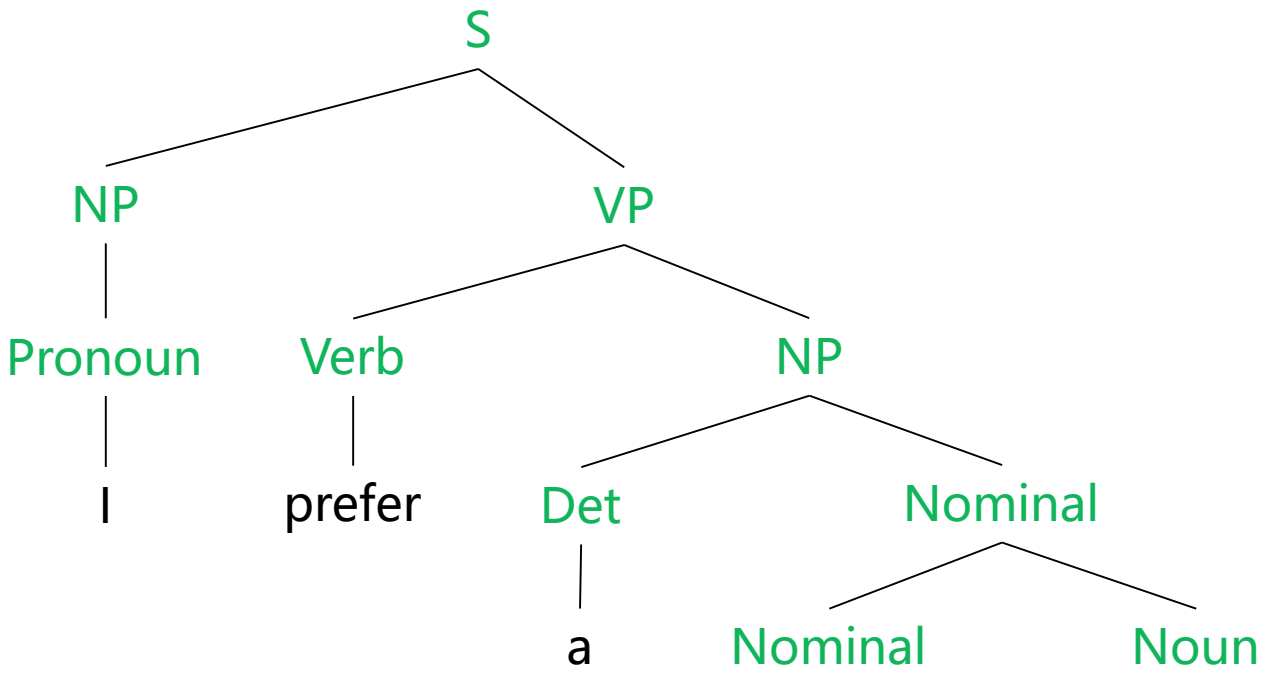
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun	→ flights | breeze | trip | morning
Verb	→ is | prefer | like | need | want | fly
Adjective	→ cheapest | non-stop | first | latest | other | direct
Pronoun	→ me | I | you | it
Proper-Noun	→ Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner	→ the | a | an | this | these | that
Preposition	→ from | to | on | near
Conjunction	→ and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



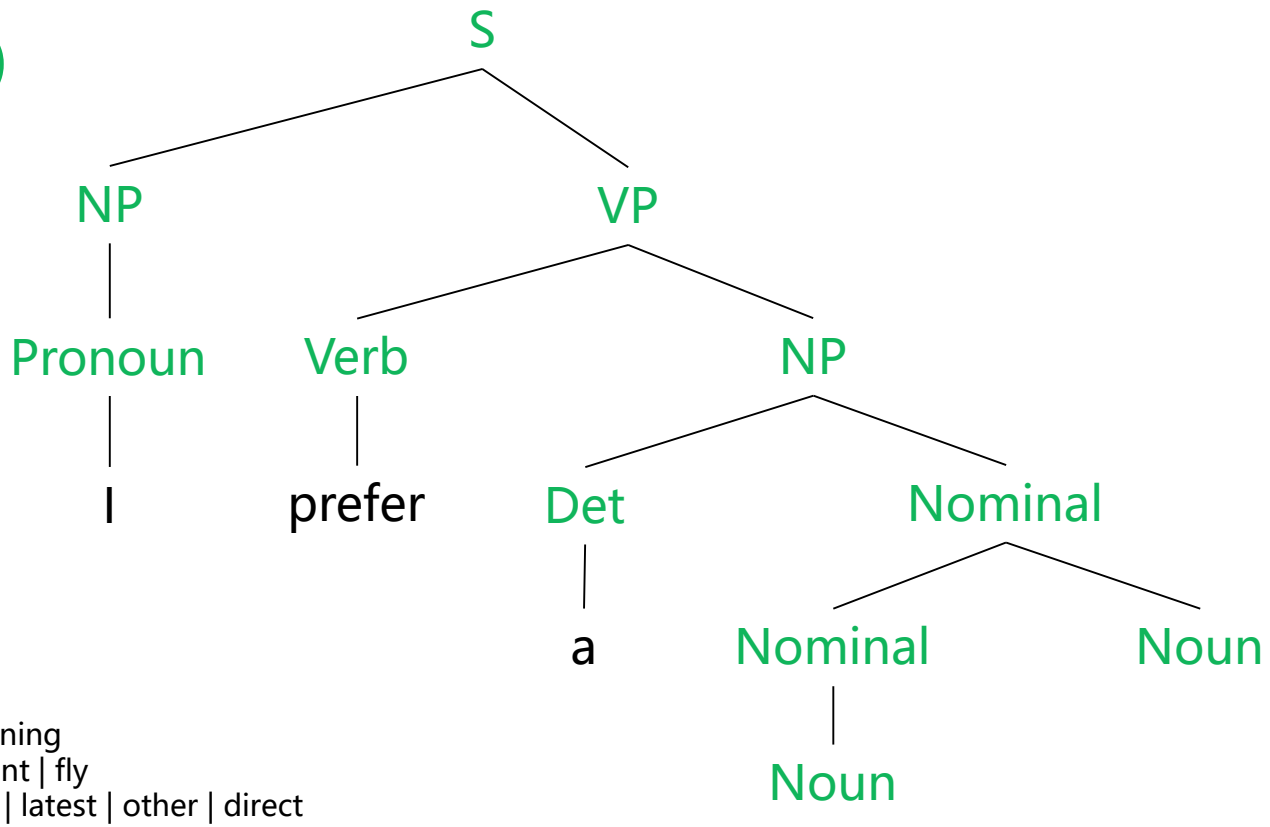
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles
| Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



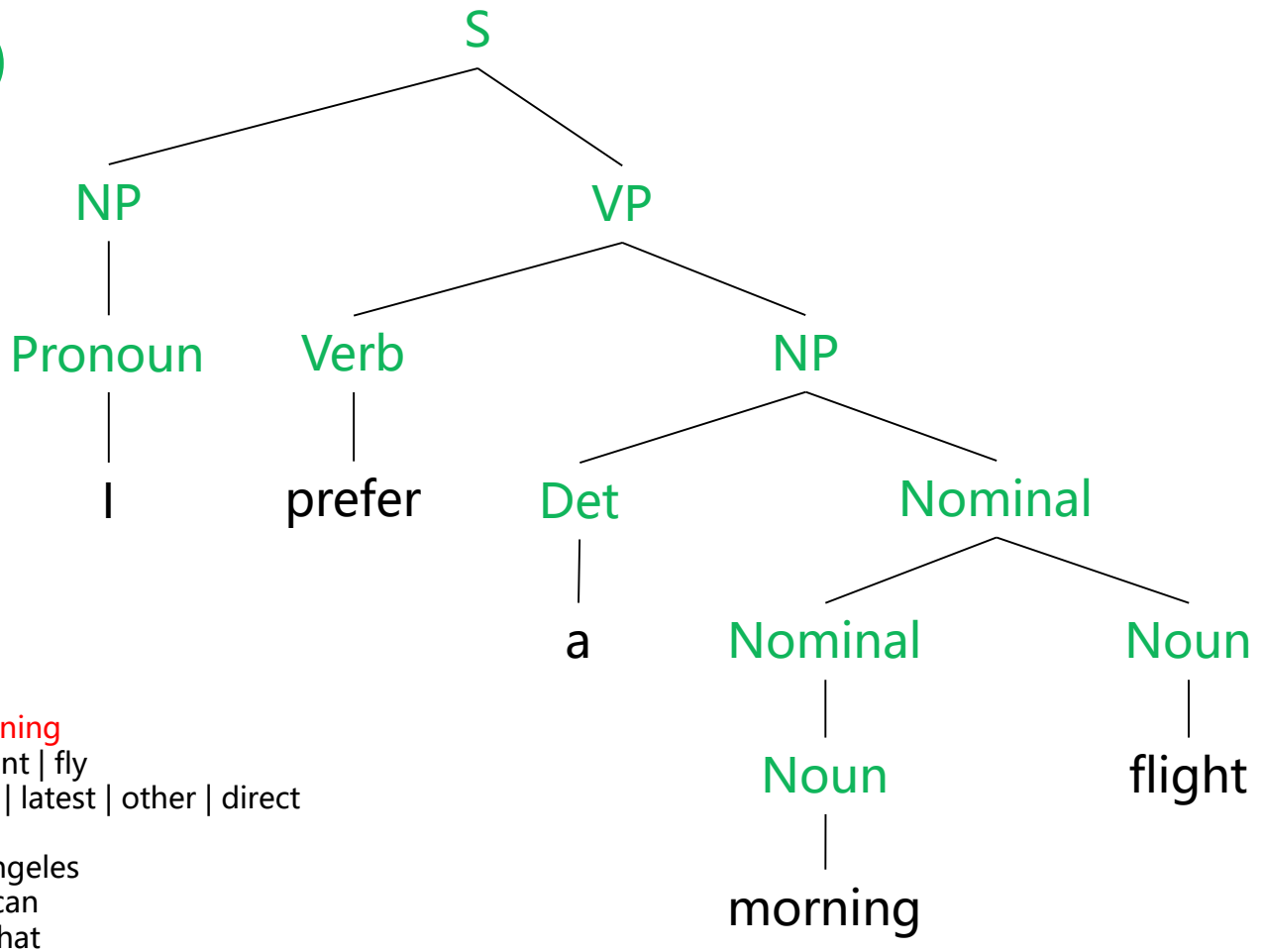
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles
| Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



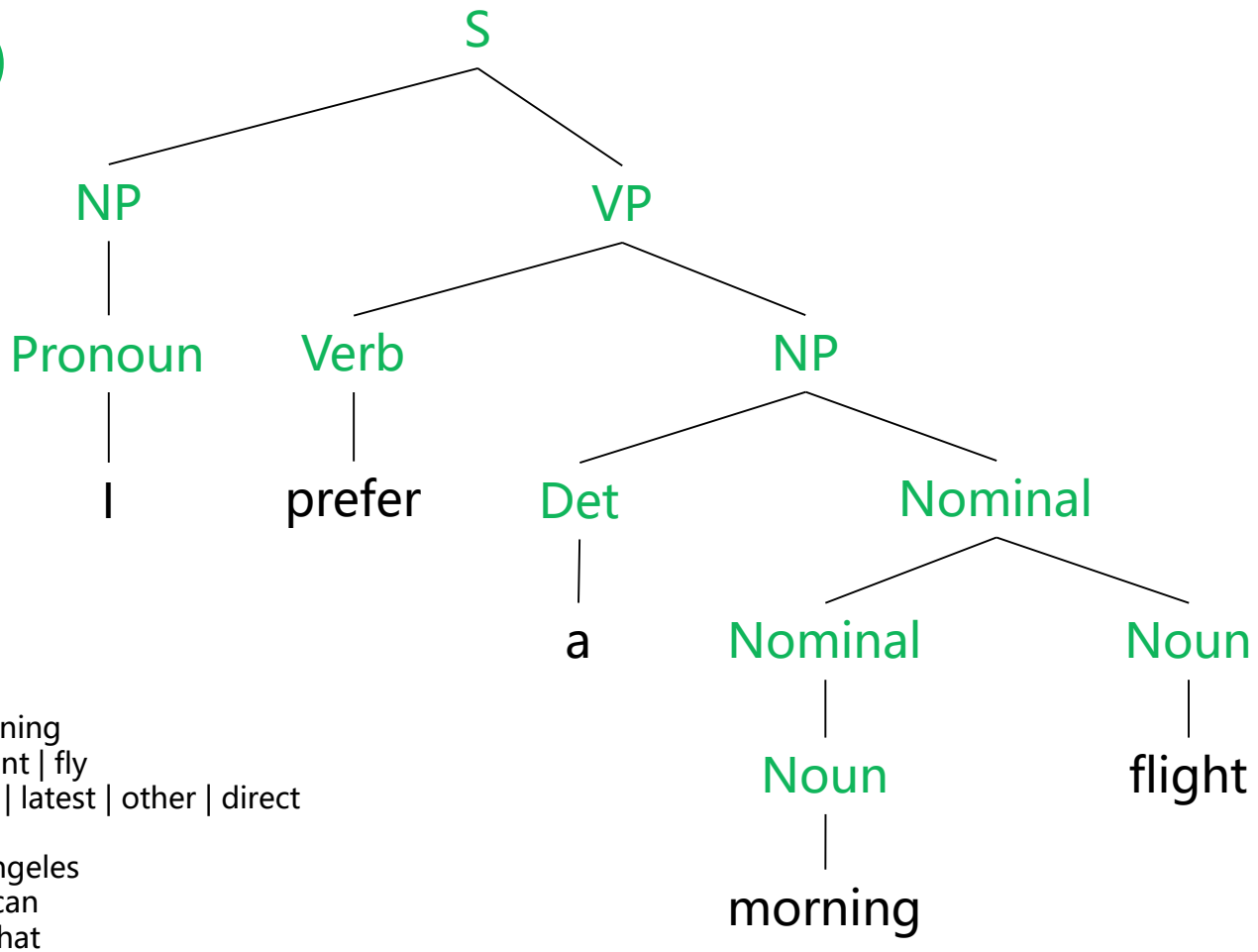
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

构成式语法（constituency grammar）简介

- 语法成分（constituent）
- 语法规则（rules）
- 词典（lexicon）

S, NP, VP, Pronoun, Proper-Noun,
Det, Nominal, Verb, PP, Preposition



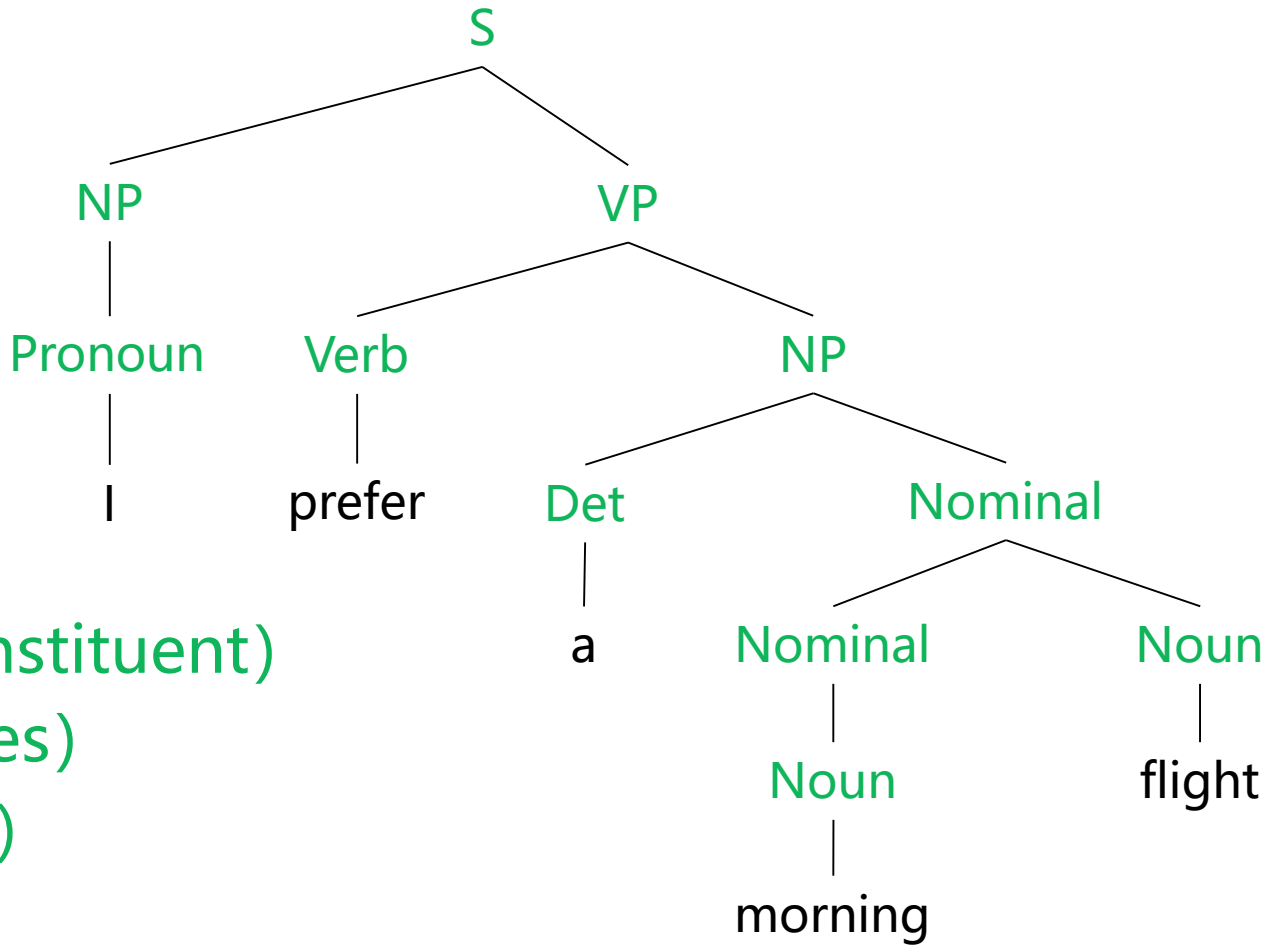
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

构成式语法（constituency grammar）简介

给定语法成分、语法规则和词典，我们即可由某一根节点S出发，生成语句。

- 这样的三要素的集合称为语法（grammar）；
- 能够由某一语法生成的语句，称为合乎语法的，否则称为不合乎语法的。



语法成分（constituent）
语法规则（rules）
词典（lexicon）

目录

构成式语法（constituency grammar）简介

语法成分（constituent）

S(句子)	Proper-Noun(专有名词)	Verb(动词)
NP(名词性短语)	Det(冠词)	PP(介词短语)
VP(动词性短语)	Nominal(名词性成分)	Preposition(介词)
Pronoun(代词)		

N

语法规则（rules）

S → NP VP	PP → Preposition NP
NP → Pronoun	VP → Verb
NP → Proper-Noun	VP → Verb NP
NP → Det Nominal	VP → Verb NP PP
Nominal → Nominal Noun	VP → Verb PP
Nominal → Noun	

词典（lexicon）

Noun	→ flights | breeze | trip | morning
Verb	→ is | prefer | like | need | want | fly
Adjective	→ cheapest | non-stop | first | latest | other | direct
Pronoun	→ me | I | you | it
Proper-Noun	→ Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner	→ the | a | an | this | these | that
Preposition	→ from | to | on | near
Conjunction	→ and | or | but

Σ

R

构成式语法（constituency grammar）简介

N 一组定义好的语法成分（constituent），并且只能作为语法树中的非叶子节点（non-terminal symbols）

S(句子)　Proper-Noun(专有名词)　Verb(动词)
NP(名词性短语)　Det(冠词)　PP(介词短语)
VP(动词性短语)　Nominal(名词性成分)　Preposition(介词)
Pronoun(代词)

Σ 词汇表，只能作为语法树中的叶子节点（terminal symbols）

flights | breeze | trip | morning | is | prefer | like | need | want | fly | cheapest | non-stop | first | latest | other | direct | Pronoun | me | I | you | it | Alaska | Baltimore | Los Angeles | Chicago | United | American | the | a | an | this | these | that | Preposition | from | to | on | near | and | or | but

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

S 语法所规定的，每个句子的根节点

CSG与CFG

上下文相关语法
(Context-sensitive Grammar，CSG)

语法树的生成与上下文相关，即：语法规则中包含“→”号左边不是单个的元素

VP NP → VP Nominal
NP NP → NP Det Nominal
prefer Pronoun PP → *prefer* Pronoun to Verb Noun
prefer NP→ *prefer* Proper-Noun

一种约束

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个元素

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

目录

构成式语法（constituency grammar）简介

N 一组定义好的语法成分（constituent），并且只能作为语法树中的非叶子节点（non-terminal symbols）

S(句子)　Proper-Noun(专有名词)　Verb(动词)
NP(名词性短语)　Det(冠词)　PP(介词短语)
VP(动词性短语)　Nominal(名词性成分)　Preposition(介词)
Pronoun(代词)

Σ 词汇表，只能作为语法树中的叶子节点（terminal symbols）

flights | breeze | trip | morning | is | prefer | like | need | want | fly | cheapest | non-stop | first | latest | other | direct | Pronoun | me | I | you | it | Alaska | Baltimore | Los Angeles | Chicago | United | American | the | a | an | this | these | that | Preposition | from | to | on | near | and | or | but

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

S 语法所规定的，每个句子的根节点

从Treebanks中构建语法

从语言学家标注的语料集中去搜罗！

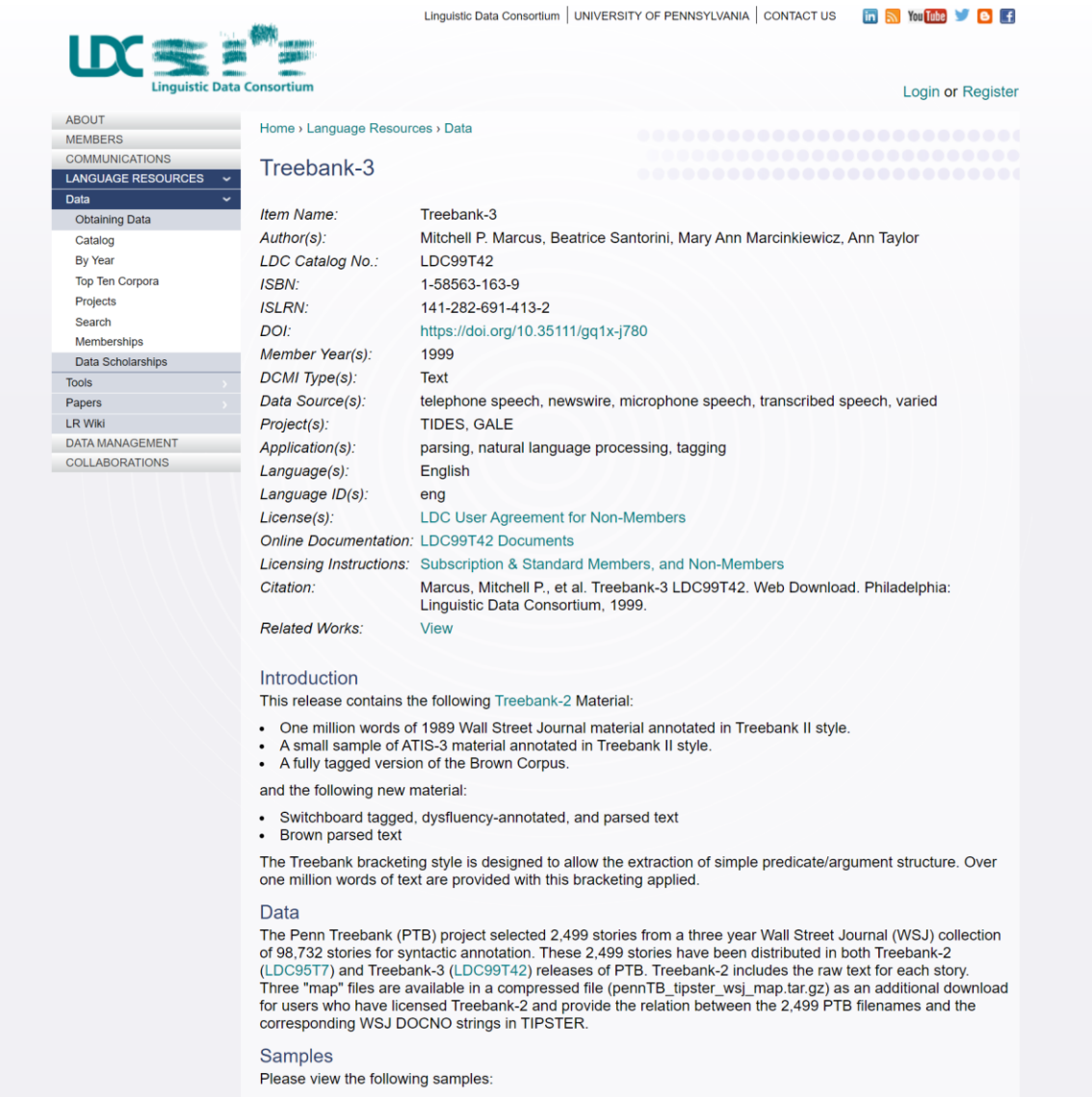
R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun

PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
→ is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

从Treebanks中构建语法：Penn Treebank语料集



Linguistic Data Consortium | UNIVERSITY OF PENNSYLVANIA | CONTACT US

Linguistic Data Consortium

Login or Register

ABOUT
MEMBERS
COMMUNICATIONS
LANGUAGE RESOURCES
Data
Obtaining Data
Catalog
By Year
Top Ten Corpora
Projects
Search
Memberships
Data Scholarships
Tools
Papers
LR Wiki
DATA MANAGEMENT
COLLABORATIONS

Home › Language Resources › Data

Treebank-3

Item Name:	Treebank-3
Author(s):	Mitchell P. Marcus, Beatrice Santorini, Mary Ann Marcinkiewicz, Ann Taylor
LDC Catalog No.:	LDC99T42
ISBN:	1-58563-163-9
ISLRN:	141-282-691-413-2
DOI:	https://doi.org/10.35111/gq1x-j780
Member Year(s):	1999
DCMI Type(s):	Text
Data Source(s):	telephone speech, newswire, microphone speech, transcribed speech, varied
Project(s):	TIDES, GALE
Application(s):	parsing, natural language processing, tagging
Language(s):	English
Language ID(s):	eng
License(s):	LDC User Agreement for Non-Members
Online Documentation:	LDC99T42 Documents
Licensing Instructions:	Subscription & Standard Members, and Non-Members
Citation:	Marcus, Mitchell P., et al. Treebank-3 LDC99T42. Web Download. Philadelphia: Linguistic Data Consortium, 1999.
Related Works:	View

Introduction

This release contains the following Treebank-2 Material:

- One million words of 1989 Wall Street Journal material annotated in Treebank II style.
- A small sample of ATIS-3 material annotated in Treebank II style.
- A fully tagged version of the Brown Corpus.

and the following new material:

- Switchboard tagged, dysfluency-annotated, and parsed text
- Brown parsed text

The Treebank bracketing style is designed to allow the extraction of simple predicate/argument structure. Over one million words of text are provided with this bracketing applied.

Data

The Penn Treebank (PTB) project selected 2,499 stories from a three year Wall Street Journal (WSJ) collection of 98,732 stories for syntactic annotation. These 2,499 stories have been distributed in both Treebank-2 (LDC95T7) and Treebank-3 (LDC99T42) releases of PTB. Treebank-2 includes the raw text for each story. Three "map" files are available in a compressed file (pennTB_tipster_wsj_map.tar.gz) as an additional download for users who have licensed Treebank-2 and provide the relation between the 2,499 PTB filenames and the corresponding WSJ DOCNO strings in TIPSTER.

Samples

Please view the following samples:

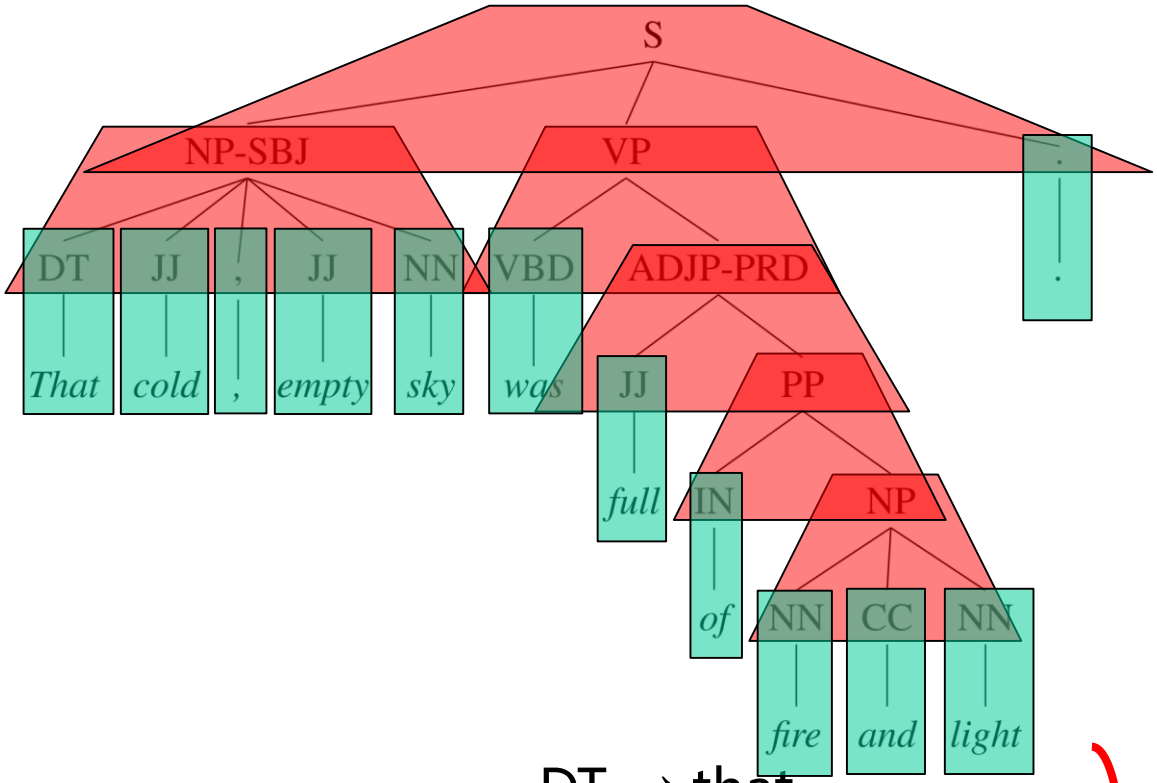
Penn treebank是宾夕法尼亚大学上世纪末收集整理的语法树数据集。

包含了从华盛顿邮报、Newswire等媒体中摘选的数万个英文语句，共约一百万个单词。并由语言学家们完成对其的语法树标注。

目前最新的treebank-3版本更新于1999年，但在语言模型、语法解析领域，仍被广泛使用。

从Treebanks中构建语法：Penn Treebank语料集

```
((S
    (NP-SBJ (DT That)
      (JJ cold) (, ,)
      (JJ empty) (NN sky) )
    (VP (VBD was)
      (ADJP-PRD (JJ full)
        (PP (IN of)
          (NP (NN fire)
            (CC and)
            (NN light) ))))
    (. .) ))
```



R

NP → NN CC NN
PP → IN NP
ADJP → PRD
PRD → JJ PP

VP → VBD ADJP
SBJ → DT JJ , JJ NN
NP → SBJ
S → NP VP .

DT → that
JJ → cold | empty | full
NN → sky | fire | light
IN → of
CC → and
VBD → was
. → .
, → ,

从Treebanks中构建语法：Penn Treebank语料集

整个语料集可以收集到17500个语法规则（不包含词典部分，即上面例子中的绿色框线）。

其中，光NP的生成法则就有数千个

如何处理长尾的分布方式

```
NP → DT JJ NN
NP → DT JJ NNS
NP → DT JJ NN NN
NP → DT JJ JJ NN
NP → DT JJ CD NNS
NP → RB DT JJ NN NN
NP → RB DT JJ JJ NNS
NP → DT JJ JJ NNP NNS
NP → DT NNP NNP NNP NNP JJ NN
NP → DT JJ NNP CC JJ JJ NN NNS
NP → RB DT JJS NN NN SBAR
NP → DT VBG JJ NNP NNP CC NNP
NP → DT JJ NNS , NNS CC NN NNS NN
NP → DT JJ JJ VBG NN NNP NNP FW NNP
NP → NP JJ , JJ `` SBAR '' NNS
```

……

目录

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

整个语料集可以收集到17500个语法规则（不包含词典部分，即上面例子中的绿色框线）。

其中，光NP的生成法则就有数千个。

事实上，某个语法成分的展开方式与具体用词息息相关。在这17500条规则中，有很大一部分是只出现了少数几次的。

```
NP → DT JJ NN
NP → DT JJ NNS
NP → DT JJ NN NN
NP → DT JJ JJ NN
NP → DT JJ CD NNS
NP → RB DT JJ NN NN
NP → RB DT JJ JJ NNS
NP → DT JJ JJ NNP NNS
NP → DT NNP NNP NNP NNP JJ NN
NP → DT JJ NNP CC JJ JJ NN NNS
NP → RB DT JJS NN NN SBAR
NP → DT VBG JJ NNP NNP CC NNP
NP → DT JJ NNS , NNS CC NN NNS NN
NP → DT JJ JJ VBG NN NNP NNP FW NNP
NP → NP JJ , JJ `` SBAR '' NNS
```

.....

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

整个语料集可以收集到17500个语法规则（不包含词典部分，即上面例子中的绿色框线）。

其中，光NP的生成法则就有数千个。

事实上，某个语法成分的展开方式与具体用词息息相关。在这17500条规则中，有很大一部分是只出现了少数几次的。

如上的构成式语法以语法成分为中心，认为语法成分的展开是有规律可循的。

但是既然PTB上的结果告诉我们，
这样的规律也并没有那么好找，所以……

不妨反过来定义语法：以词汇为中心，对每个词定义他可以怎么用。

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

词汇化语法（lexicalized grammar）：更多依赖具体词汇，认为合法的语法结构与词汇具体是什么相关。

help sb do sth
ask sb to do sth
pass sth to sb
....

词汇化语法
(lexicalized grammar)

- Lexical-Functional Grammar (LFG)
  (Bresnan, 1982)
- Head-Driven Phrase Structure Grammar (HPSG)
  (Pollard and Sag, 1994)
- Tree-Adjoining Grammar (TAG)
  (Joshi,1985)
- Combinatory Categorial Grammar (CCG)
  (Steedman 1989, Steedman 2000)

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

词汇化语法（lexicalized grammar）：更多依赖具体词汇，认为合法的语法结构与词汇具体是什么相关。

help sb do sth
ask sb to do sth
pass sth to sb
....

词汇化语法
(lexicalized grammar)

- Lexical-Functional Grammar (LFG)
  (Bresnan, 1982)
- Head-Driven Phrase Structure Grammar (HPSG)
  (Pollard and Sag, 1994)
- Tree-Adjoining Grammar (TAG)
  (Joshi,1985)
- **Combinatory Categorial Grammar (CCG)**
  (Steedman 1989, Steedman 2000)

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

组合范畴语法
(combinatory categorial grammar, CCG)

核心思想：词汇的类别（category）变得复杂；通过复杂的类别指导每个词该怎么用。

类别（category）

1）基础类别（atomic elements）

S(句子)	Proper-Noun(专有名词)	Verb(动词)
NP(名词性短语)	Det(冠词)	PP(介词短语)
VP(动词性短语)	Nominal(名词性成分)	Preposition(介词)
Pronoun(代词)	……	

2）由基础类别构成的函数，以及函数的函数

(PP/NP): 在该词的右边放上一个NP成分，那么该词与其右边的成分即可共同构成一个PP。

(S\Pronoun)：在该词的左边放上一个代词，那么该词与其左边的成分即可共同构成一个句子。

(S\NP)/NP：在该词的左边放上一个NP，然后在右边放上一个NP，那么该词与两边的成分共同构成一个句子。

to	you	I	say	Gov.	serve	people
(PP/NP)	NP	Pronoun	(S\Pronoun)	NP	(S\NP)/NP	NP
PP		S		S		

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

组合范畴语法
(combinatory categorial grammar, CCG)

核心思想：词汇的类别（category）变得复杂；通过复杂的类别指导每个词该怎么用。

词典（lexicon）

每个词的类别直接指定为以上形式的类别。这类复杂的类别模式已经暗含了词的使用方法

Flight:	N
Say:	Verb；(S\Pronoun)
Cancel:	(S\NP)/NP
……	

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

组合范畴语法
(combinatory categorial grammar, CCG)

核心思想：词汇的类别（category）变得复杂；通过复杂的类别指导每个词该怎么用。

组合规则（rules）

1）前向匹配与后向匹配(forward function application, backward function application)

$$X/Y\ Y \Rightarrow X$$
$$Y\ X\backslash Y \Rightarrow X$$

2）并联(conjunction)

$$X\ CONJ\ X \Rightarrow X$$

to	you
(PP/NP)	NP
PP	

Geneva	and	Paris
NP	CONJ	NP
NP		

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

组合范畴语法
(combinatory categorial grammar, CCG)

核心思想：词汇的类别（category）变得复杂；通过复杂的类别指导每个词该怎么用。

组合规则（rules）

1）前向匹配与后向匹配(forward function application, backward function application)

$$X/Y\ Y \Rightarrow X$$
$$Y\ X\backslash Y \Rightarrow X$$

2）并联(conjunction)

$$X\ CONJ\ X \Rightarrow X$$

3）前向串联与后向串联(forward/backward composition)

$$X/Y\ Y/Z \Rightarrow X/Z$$
$$Y\backslash Z\ X/Y \Rightarrow X\backslash Z$$

\

Gov.	serve
S/(S\NP)	(S\NP)/NP
S/NP	

词汇化语法（lexicalized grammar）与组合范畴语法（CCG）

组合范畴语法
(combinatory categorial grammar, CCG)

核心思想：词汇的类别（category）变得复杂；通过复杂的类别指导每个词该怎么用。

组合规则（rules）

1）前向匹配与后向匹配(forward function application, backward function application)

$$X/Y\ Y \Rightarrow X$$
$$Y\ X\backslash Y \Rightarrow X$$

2）并联(conjunction)

$$X\ CONJ\ X \Rightarrow X$$

3）前向串联与后向串联(forward/backward composition)

$$X/Y\ Y/Z \Rightarrow X/Z$$
$$Y\backslash Z\ X/Y \Rightarrow X\backslash Z$$

打错了

4）类别升迁(type raising)

$$X \Rightarrow T/(T\backslash X)$$
$$X \Rightarrow T\backslash(T/X)$$

Gov.

$$\frac{NP}{S/(S\backslash NP)}$$

目录

语法间的等同关系，乔姆斯基范式（CNF）

强等价：两个语法能够生成一模一样的合法语句集合，且对于同一个句子，语法结构也相同。

弱等价：两个语法能够生成一模一样的合法语句集合，但是对于同一个句子，语法结构不一定相同

CSG与CFG

上下文相关语法
(Context-sensitive Grammar，CSG)

语法树的生成与上下文相关，即：语法规则中包含“→”号左边不是单个的N中元素情况

VP NP → VP Nominal
NP NP → NP Det Nominal
prefer Pronoun PP → *prefer* Pronoun to Verb Noun
prefer NP→ *prefer* Proper-Noun

约束R中规则的形式

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

CSG与CFG

上下文相关语法
(Context-sensitive Grammar，CSG)

语法树的生成与上下文相关，即：语法规则中包含“→”号左边不是单个的N中元素情况

VP NP → VP Nominal
NP NP → NP Det Nominal
prefer Pronoun PP → *prefer* Pronoun to Verb Noun
prefer NP→ *prefer* Proper-Noun

约束R中规则的形式

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

乔姆斯基范式（CNF）

乔姆斯基范式
(Chomsky Normal Form， CNF)

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

约束R中规则的形式

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

乔姆斯基范式（CNF）

乔姆斯基范式
(Chomsky Normal Form， CNF）

在CFG要求的基础上：

每条语法规则只能从一个元素生成两个元素，且不得包含单词

- √ S → NP VP
- √ NP → Det Nominal
- × NP → Pronoun
- × VP → Verb NP PP

每条词典规则只能从一个元素生成一个单词作为元素

- √ Noun → flights | breeze | trip | morning
- × Noun → New York

约束R中规则的形式

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

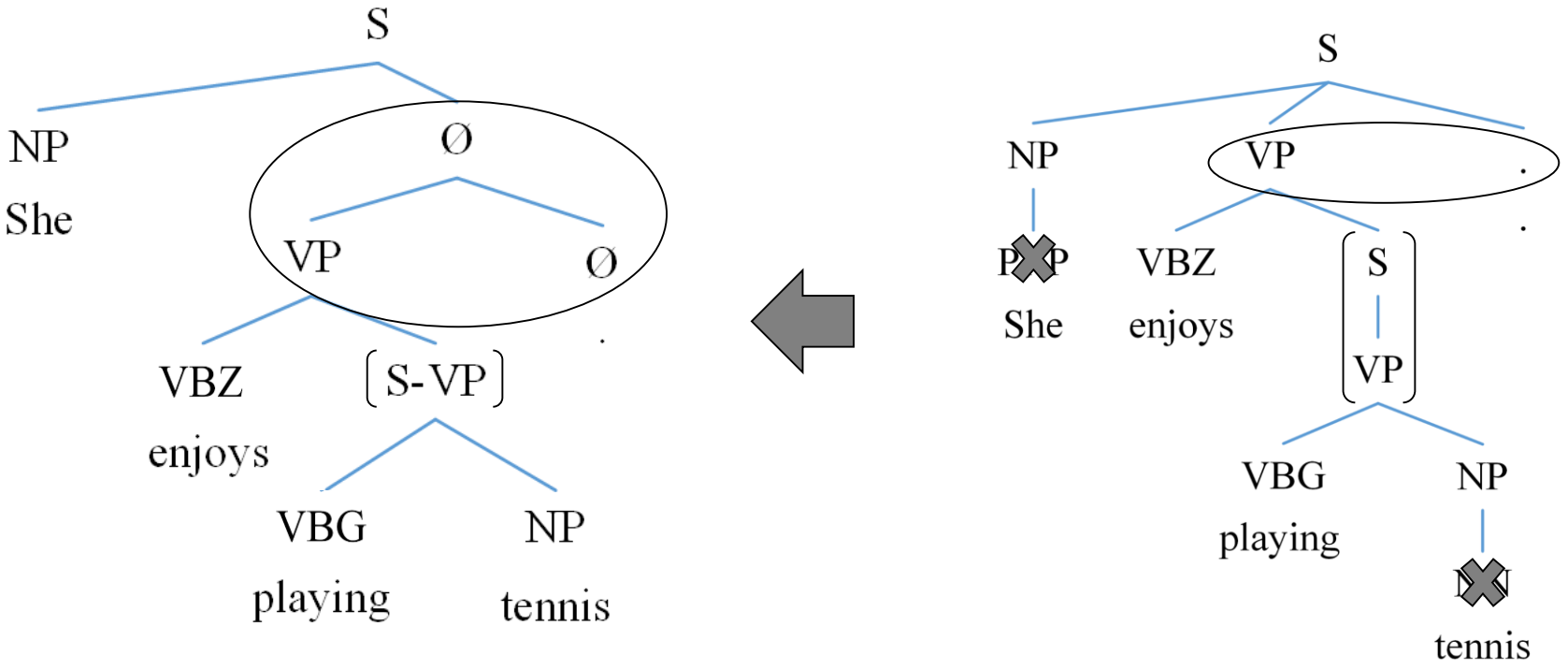
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

乔姆斯基范式（CNF）

乔姆斯基范式
（Chomsky Normal Form， CNF ）

在CFG要求的基础上：

1. 每条语法规则只能从一个元素生成两个元素，且不得包含单词
2. 每条词典规则只能从一个元素生成一个单词作为元素



目录

构成式语法的语法解析算法：CKY

待解决的问题：

给定一个定义好的CNF语法，对某一句话做自动语法分析

The input string is generated by grammar

R 符合CNF的语法规则，即$A \longrightarrow B\ C$形式的语法规则，及$A \longrightarrow < word >$形式的词典规则。

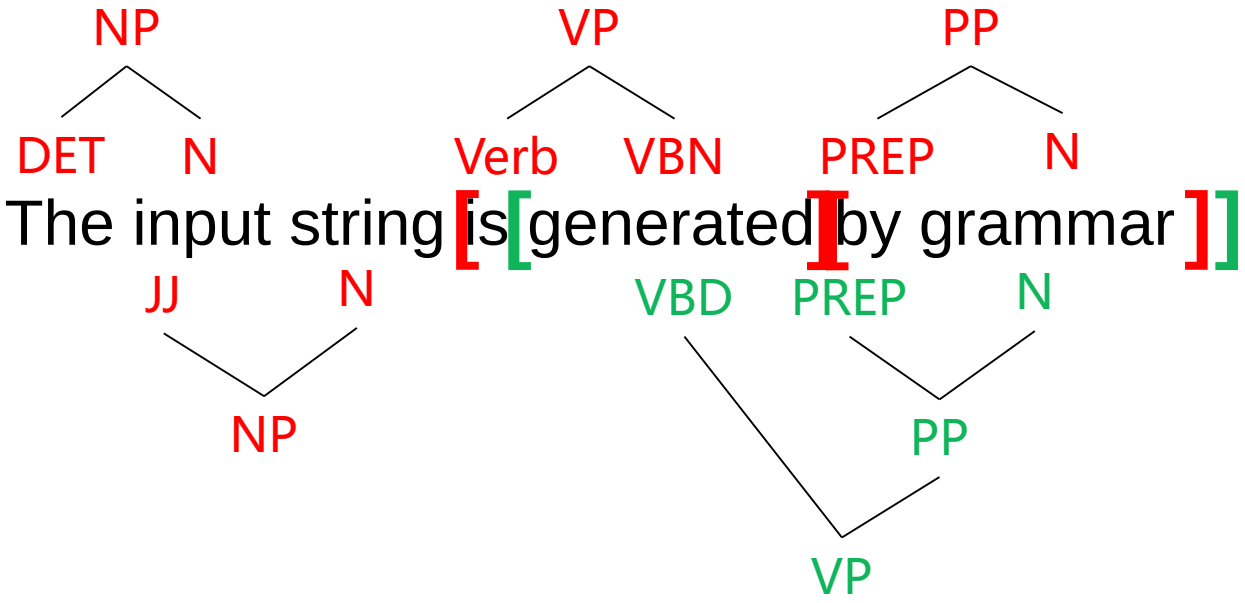
S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

构成式语法的语法解析算法：CKY

待解决的问题：

给定一个定义好的CNF语法，对某一句话做自动语法分析



根据语法规则，找到两三个词的局部语法树相对简单；但是：

如何从所有可能的局部语法树中

——这些局部语法树之间既可能相互重叠，又可能相互矛盾——

组合出一棵能够完整构建起整个句子的全局语法树？

构成式语法的语法解析算法：CKY

例子：

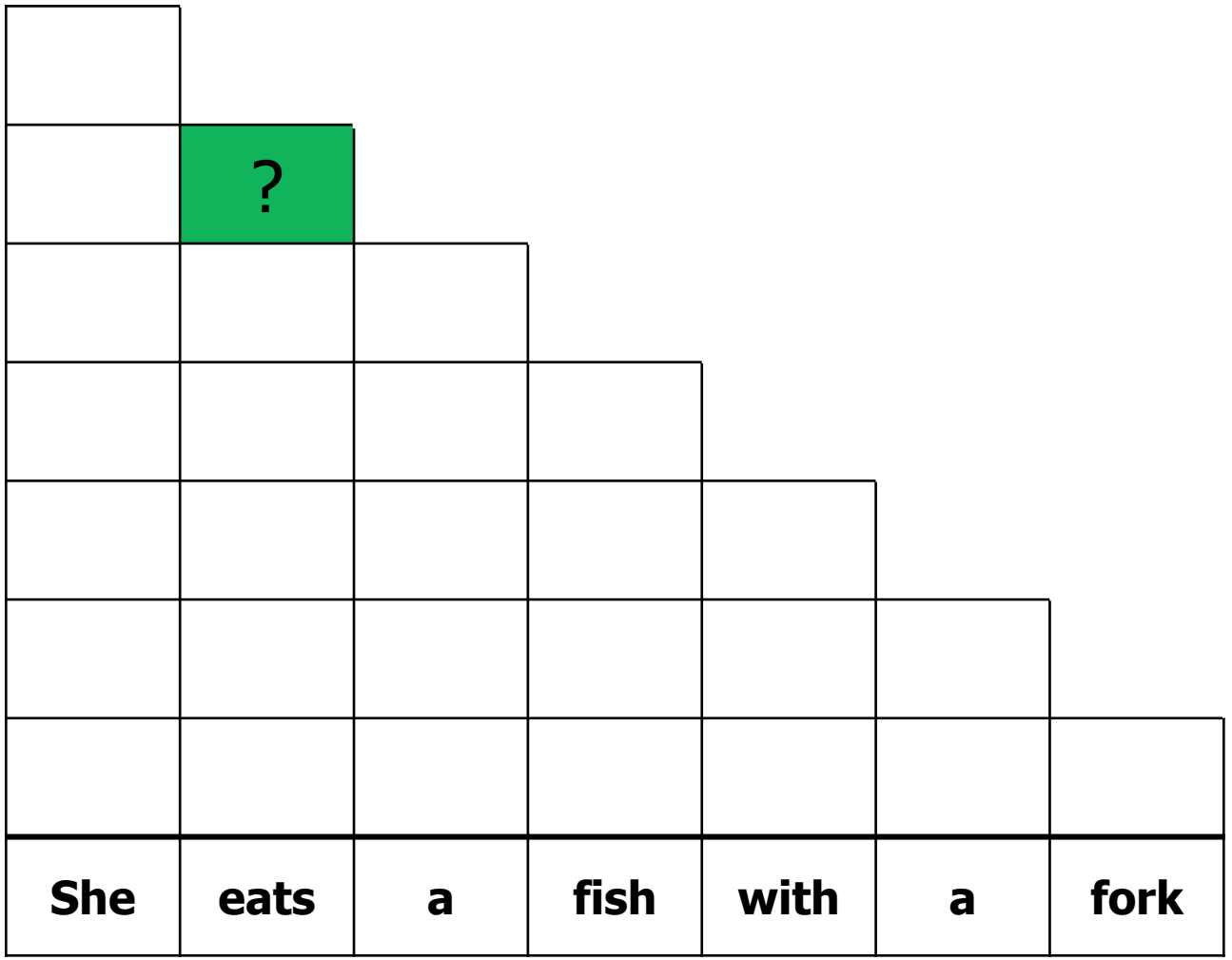
$$
\begin{aligned}
\mathrm{S} &\longrightarrow \mathrm{NP\ VP} \\
\mathrm{VP} &\longrightarrow \mathrm{VP\ PP} \\
\mathrm{VP} &\longrightarrow \mathrm{V\ NP} \\
\mathrm{PP} &\longrightarrow \mathrm{P\ NP} \\
\mathrm{NP} &\longrightarrow \mathrm{Det\ N}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{VP} &\longrightarrow \mathrm{eats} \\
\mathrm{NP} &\longrightarrow \mathrm{she} \\
\mathrm{V} &\longrightarrow \mathrm{eats} \\
\mathrm{P} &\longrightarrow \mathrm{with} \\
\mathrm{N} &\longrightarrow \mathrm{fish} \\
\mathrm{N} &\longrightarrow \mathrm{fork} \\
\mathrm{Det} &\longrightarrow \mathrm{a}
\end{aligned}
$$

She eats a fish with a fork

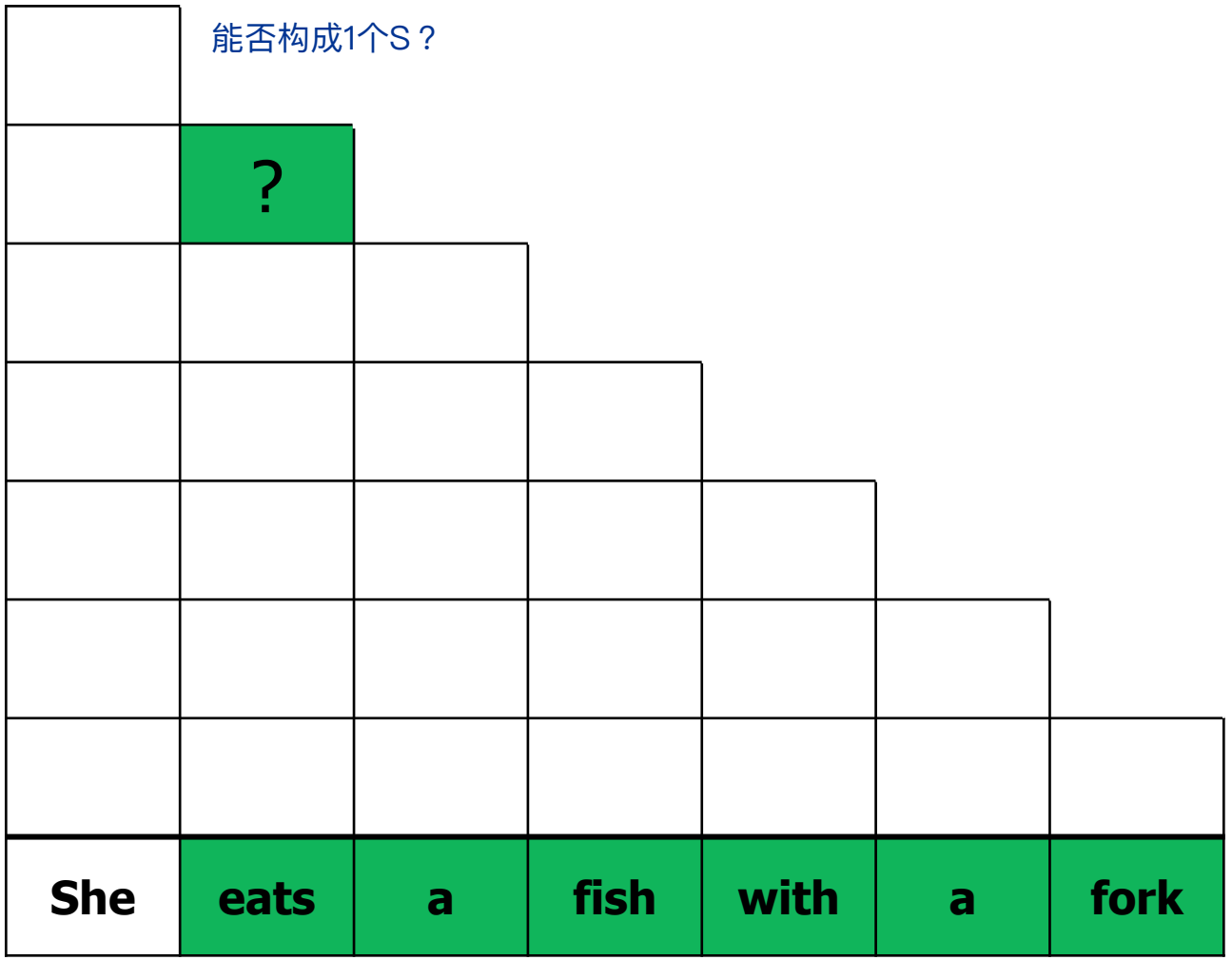
构成式语法的语法解析算法：CKY

例子：



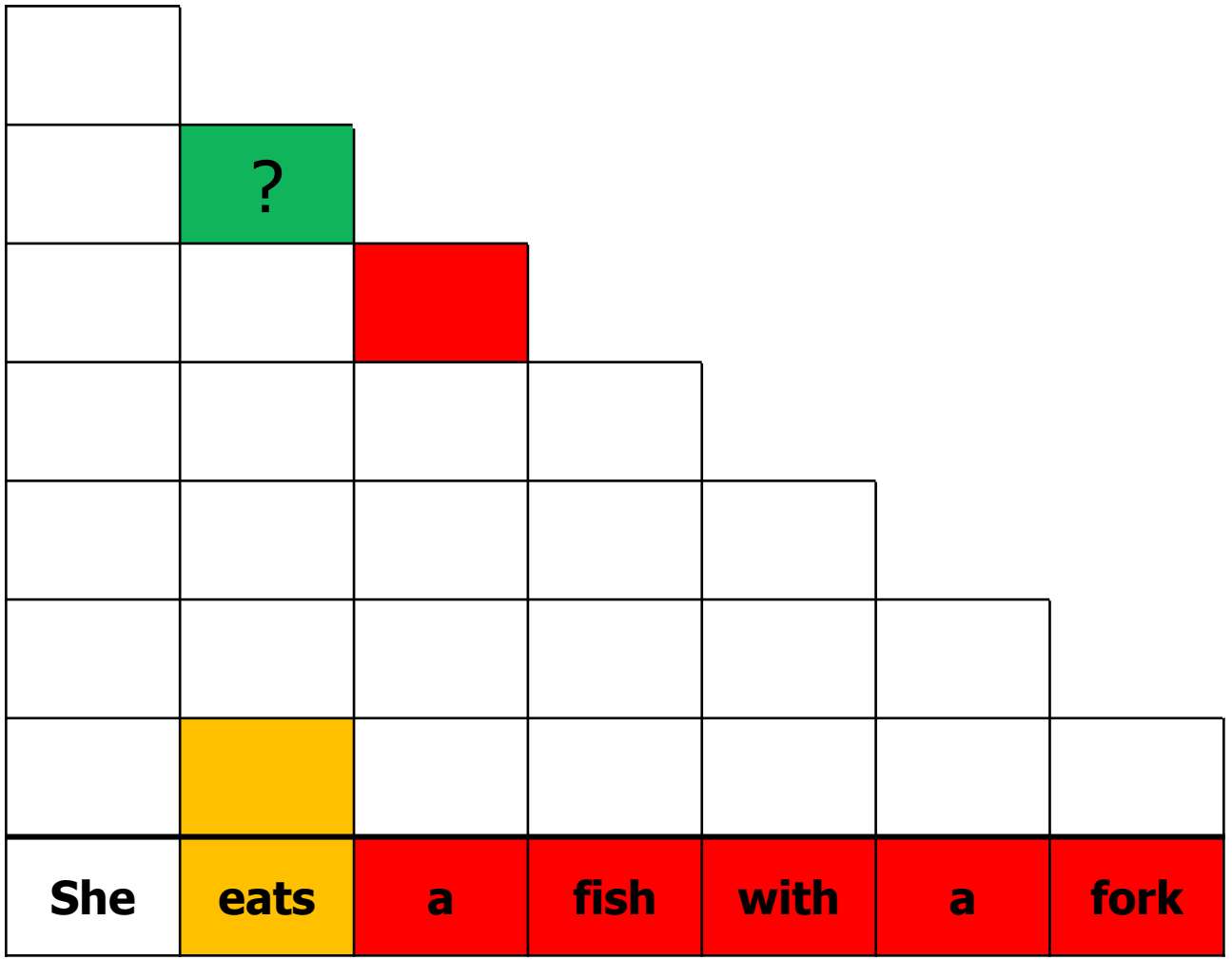
构成式语法的语法解析算法：CKY

例子：



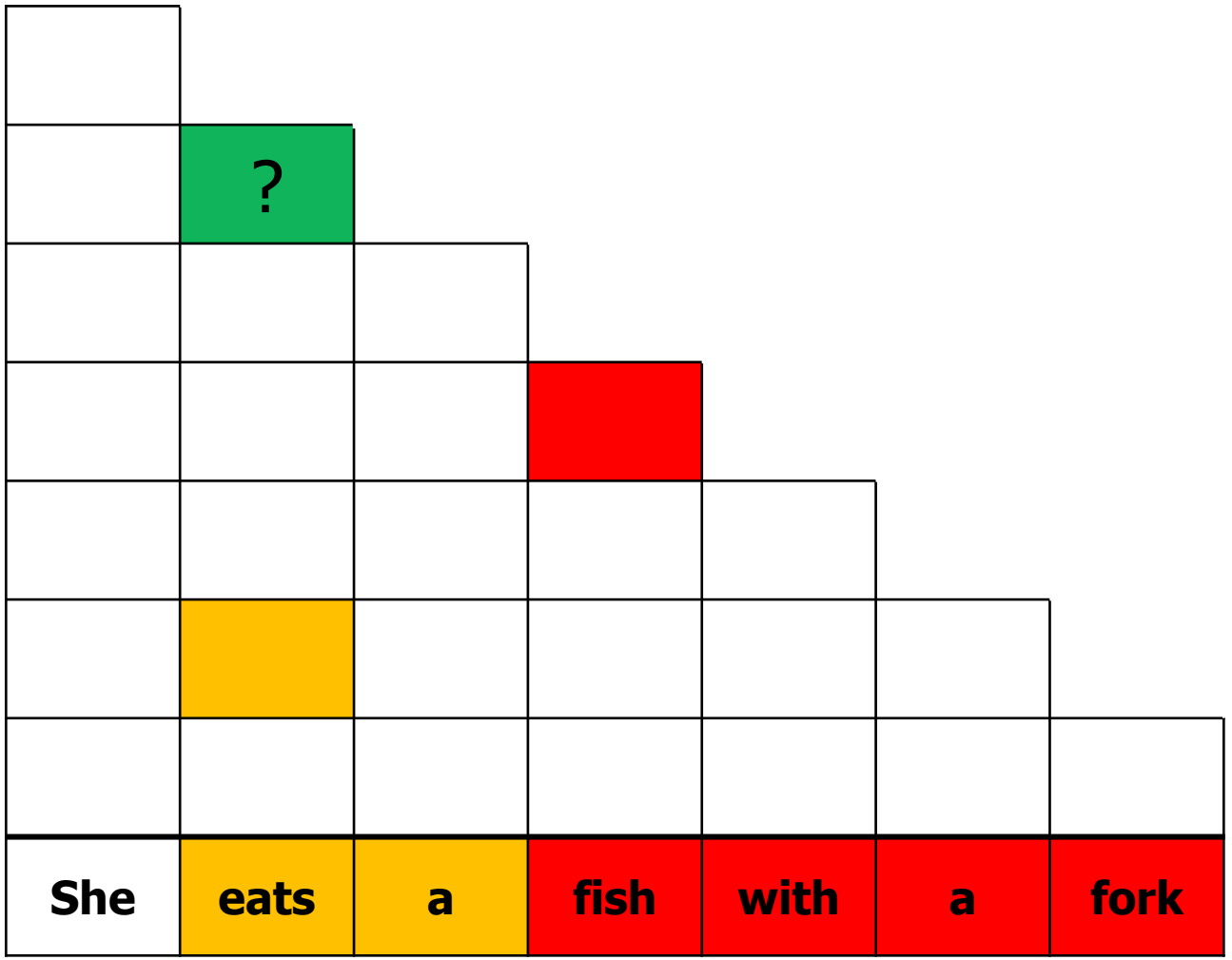
构成式语法的语法解析算法：CKY

例子：



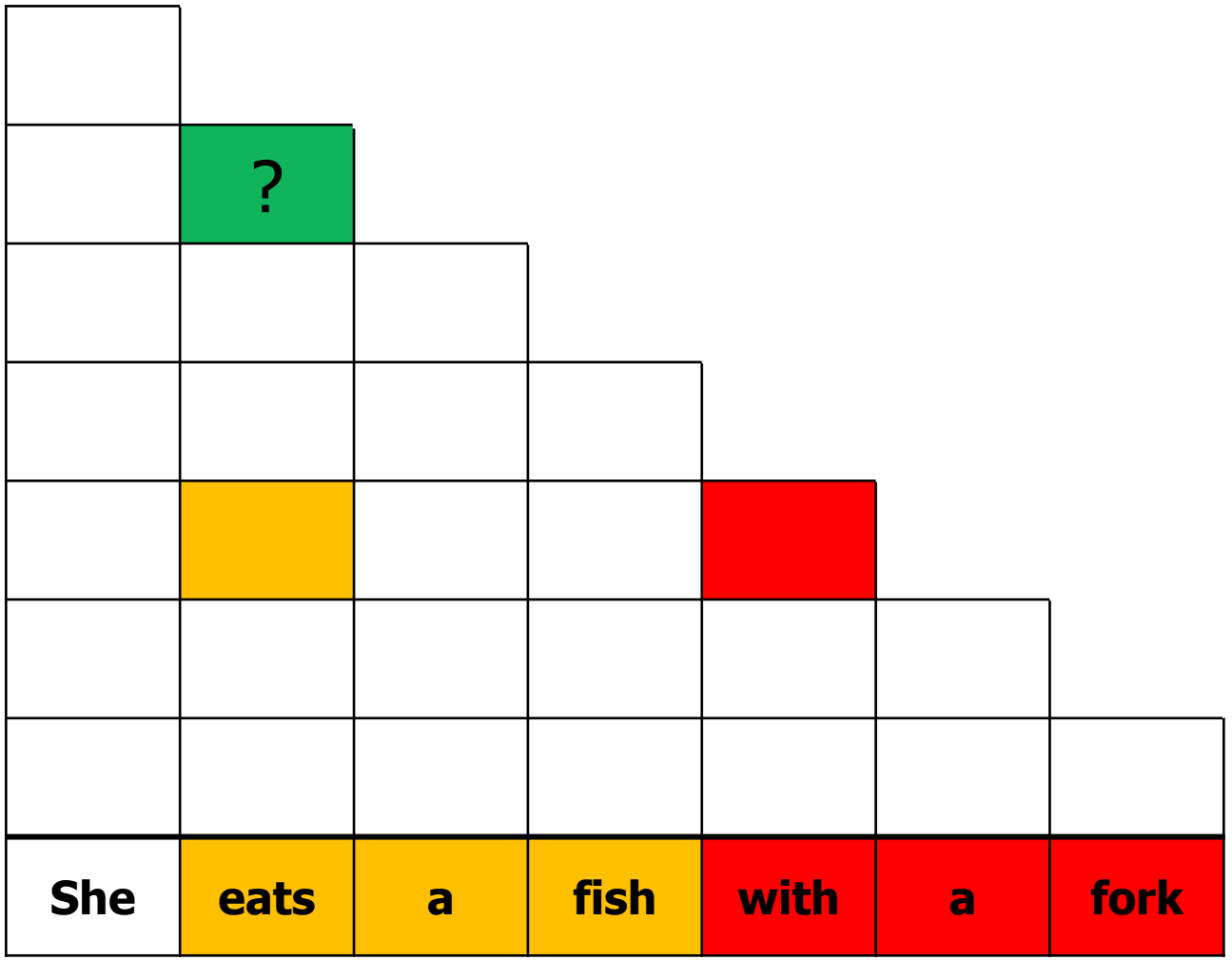
构成式语法的语法解析算法：CKY

例子：



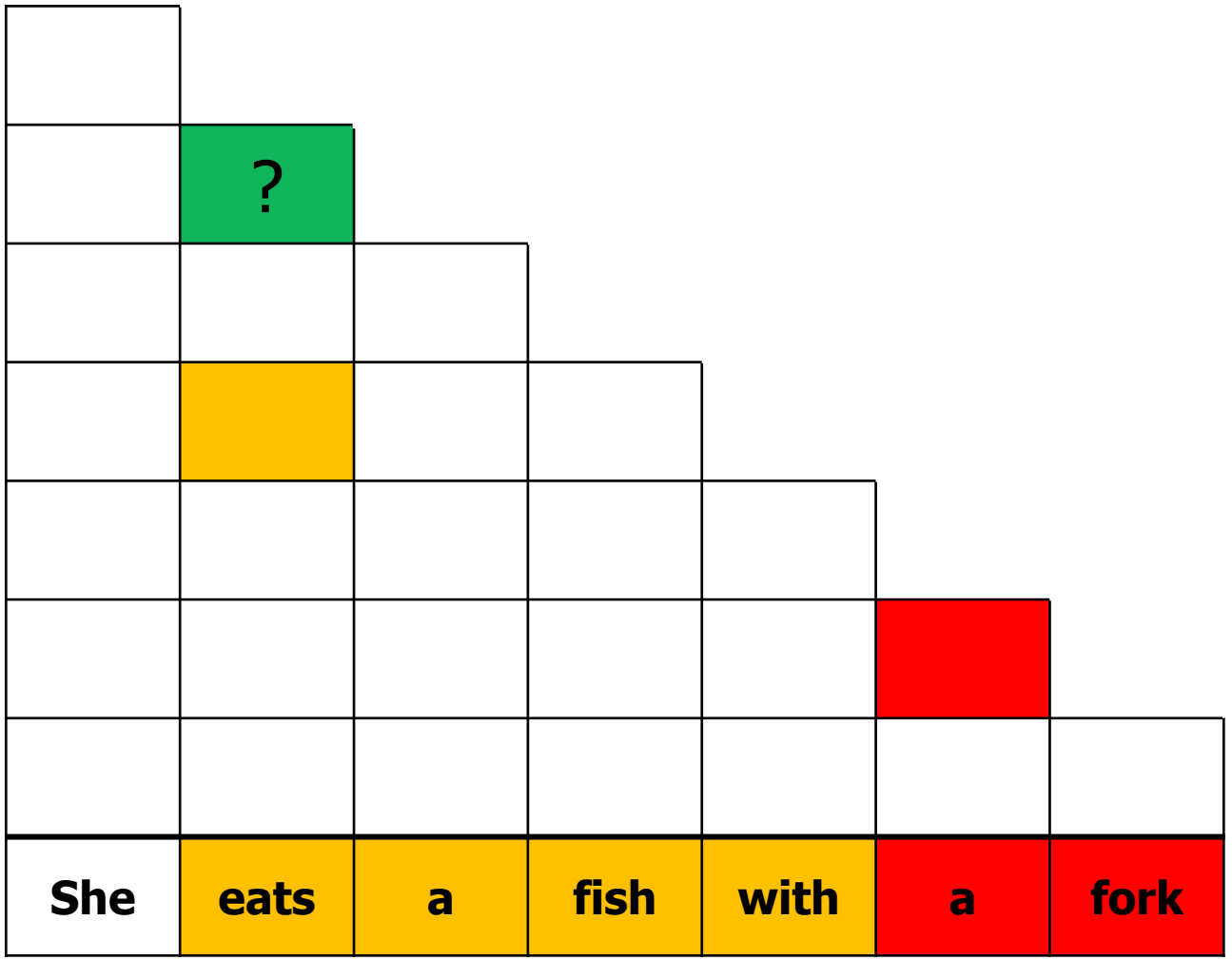
构成式语法的语法解析算法：CKY

例子：



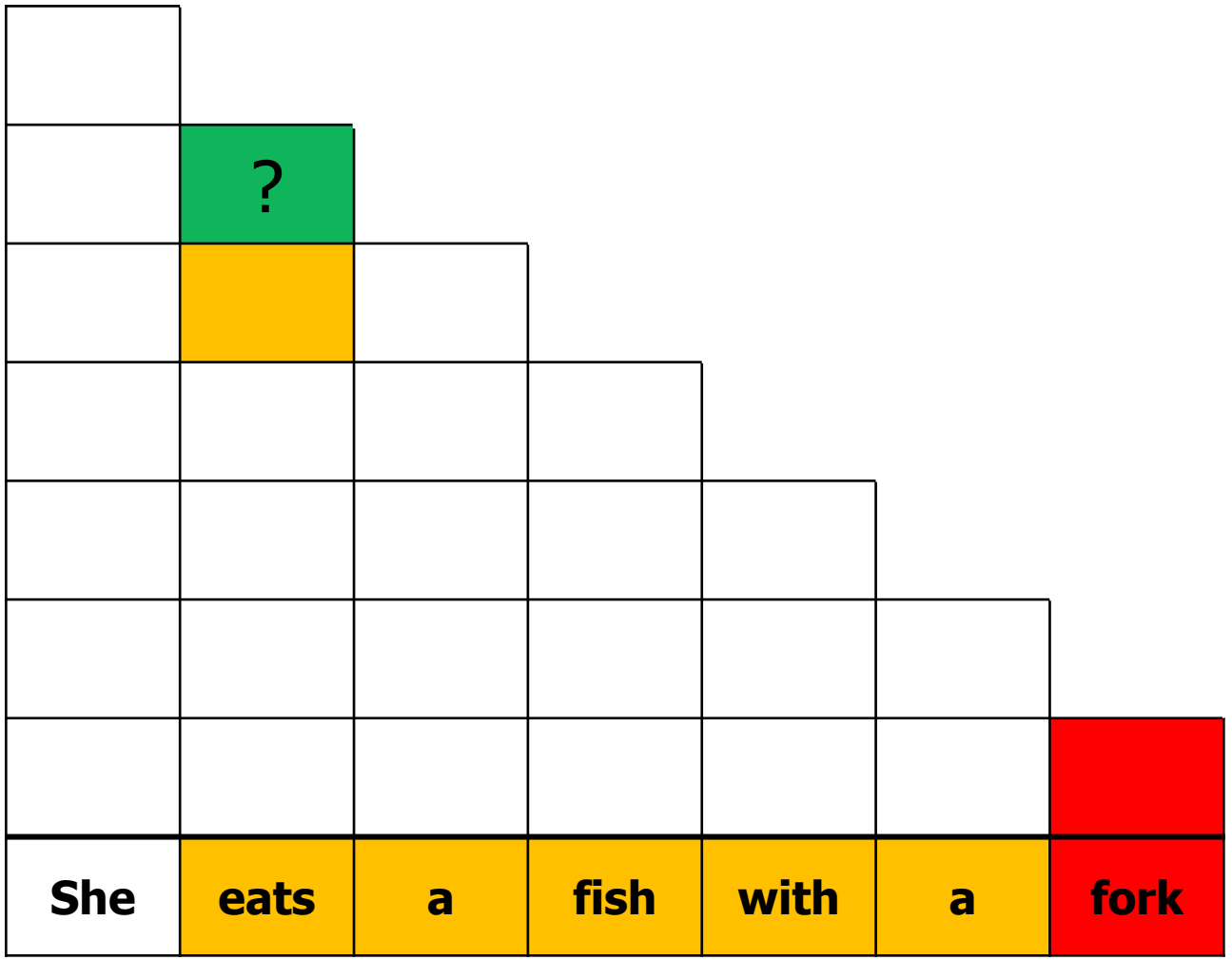
构成式语法的语法解析算法：CKY

例子：



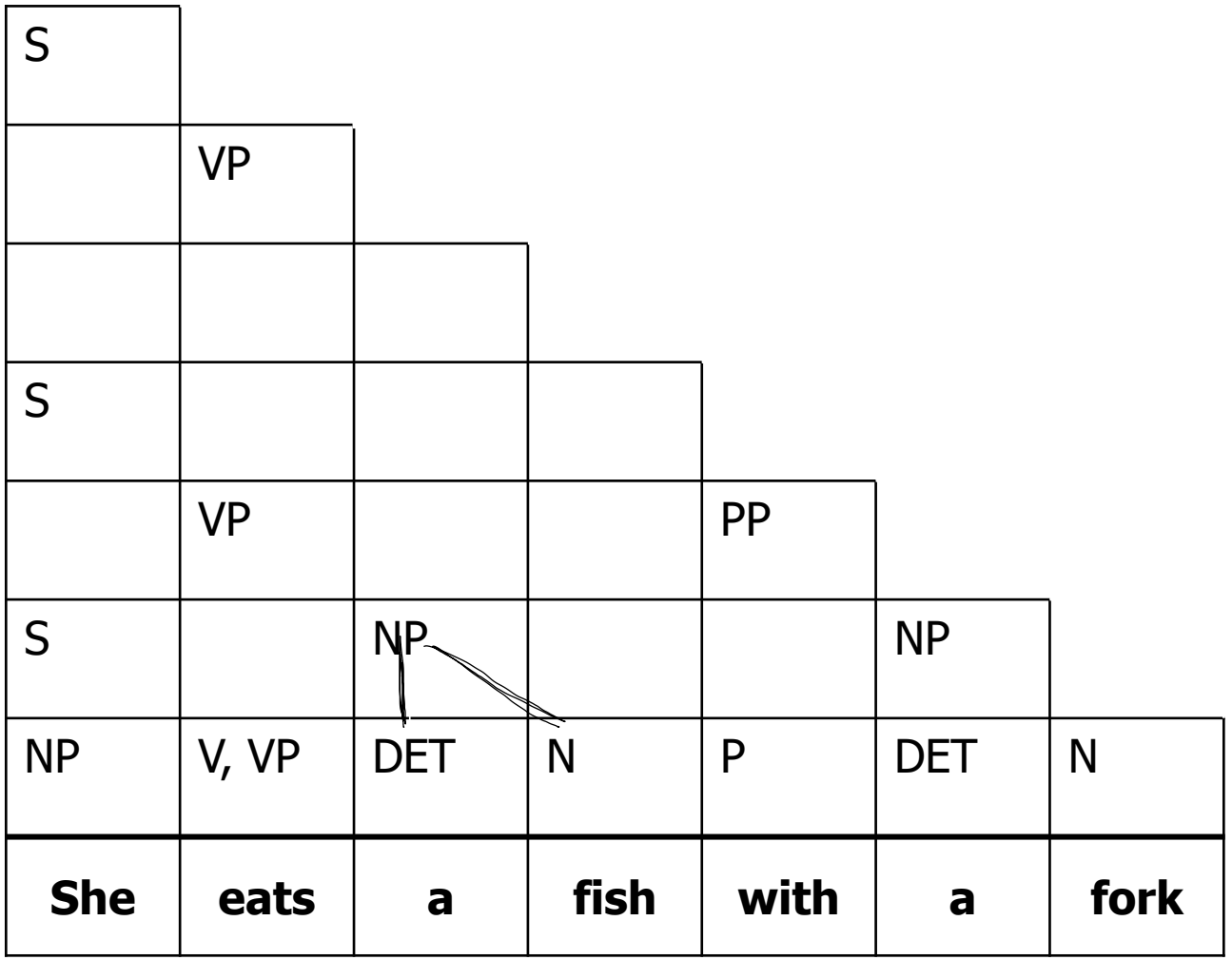
构成式语法的语法解析算法：CKY

例子：



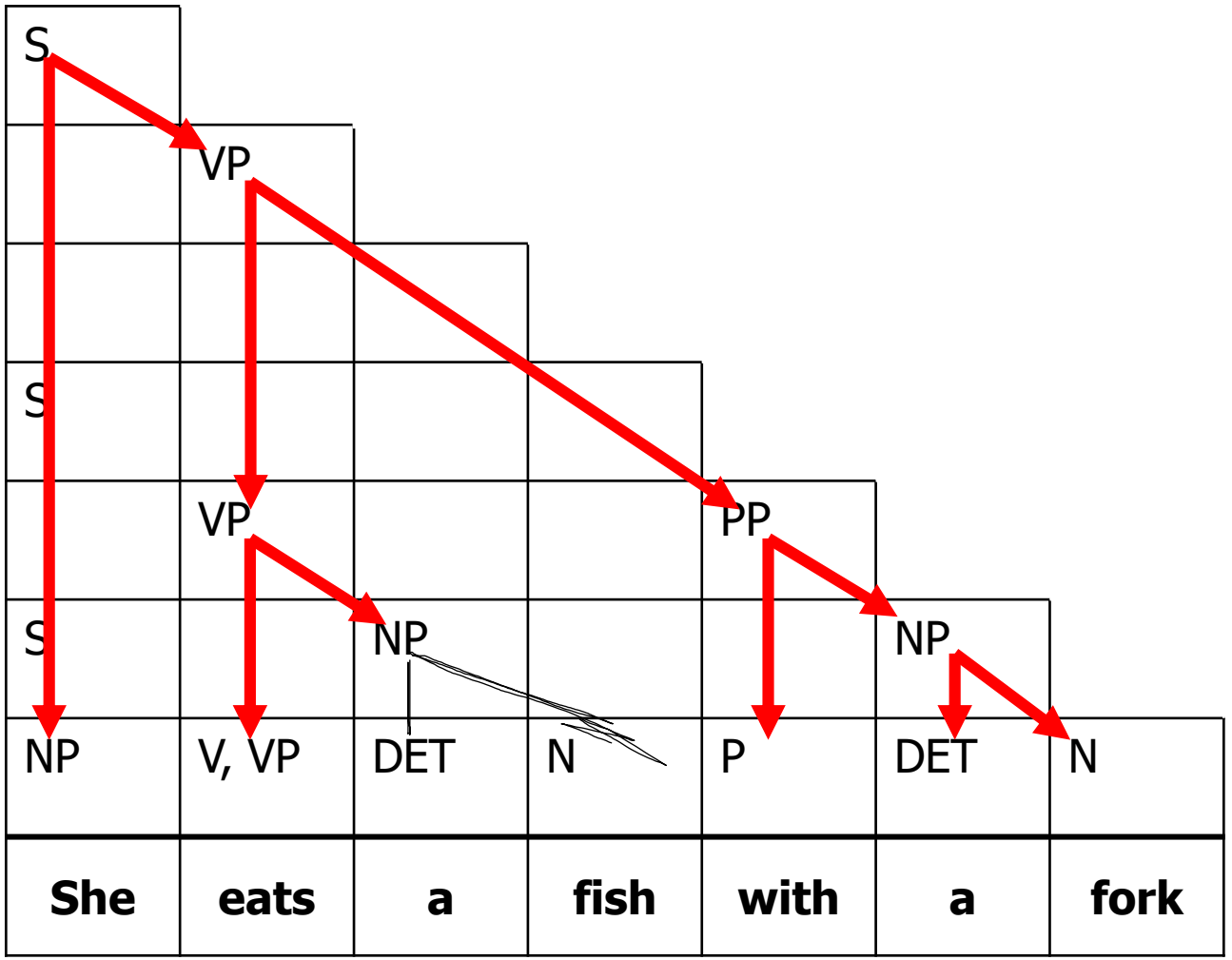
构成式语法的语法解析算法：CKY

例子：



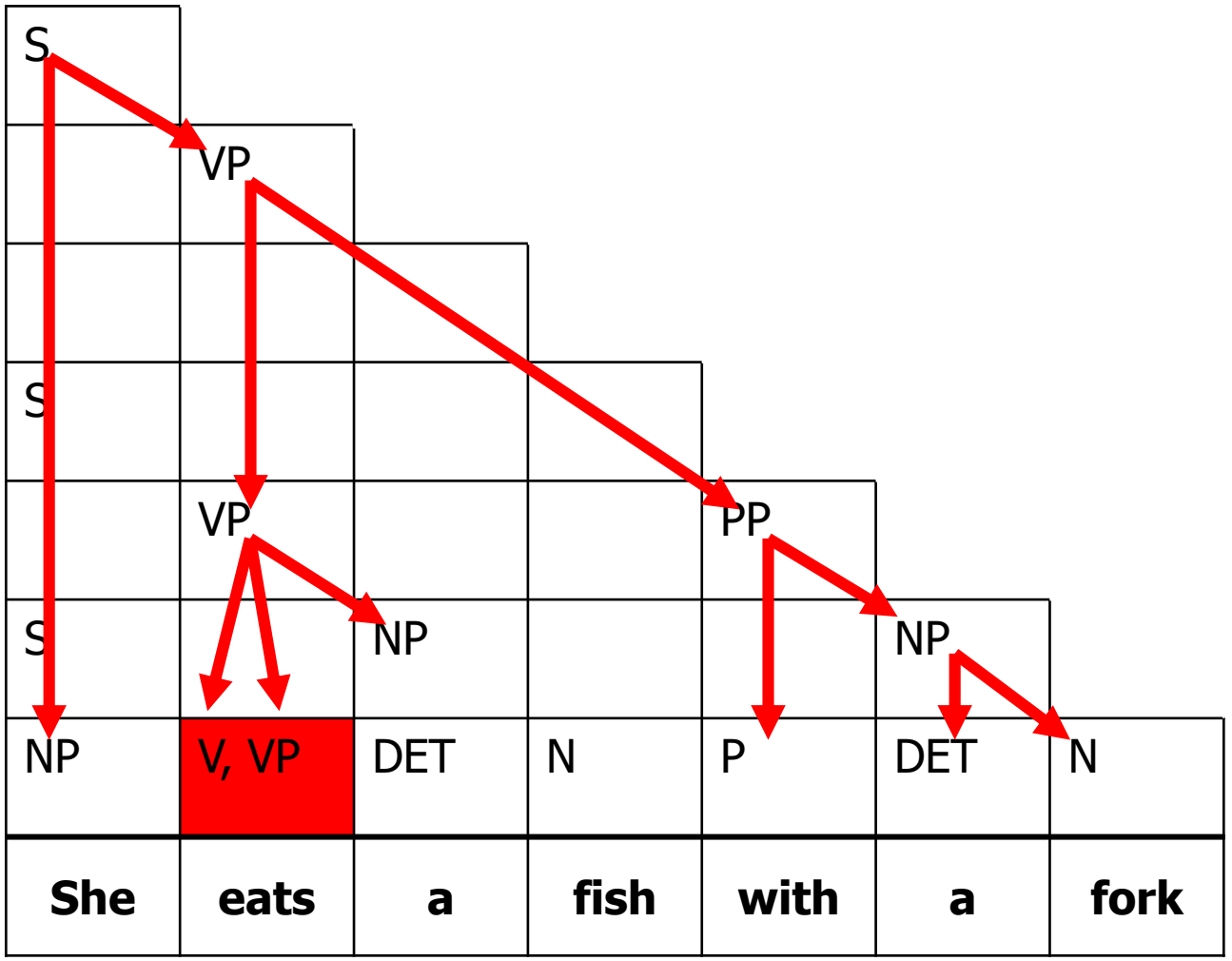
构成式语法的语法解析算法：CKY

例子：



构成式语法的语法解析算法：CKY

例子：



不知道是由eats的哪个语法成分表述的，只能说是合乎规范的。

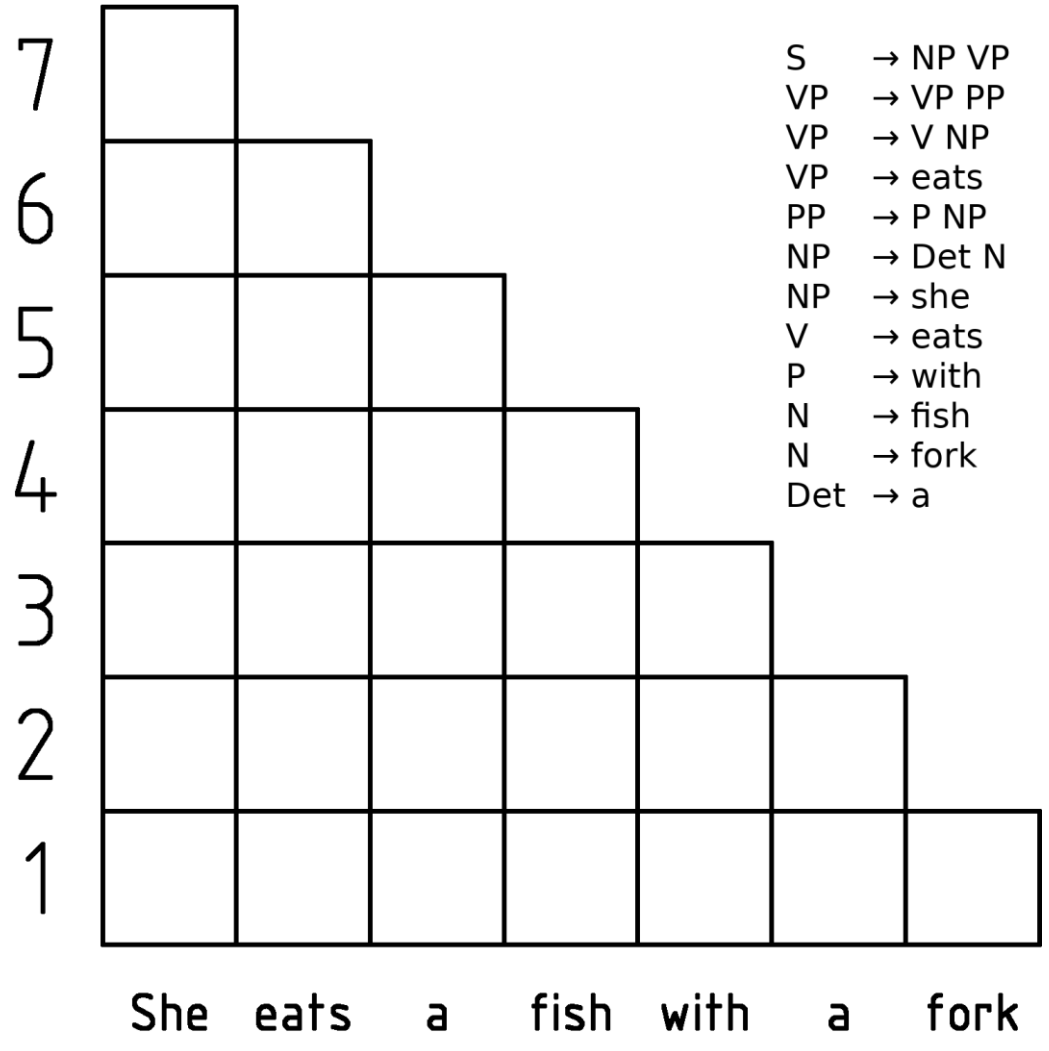
构成式语法的语法解析算法：CKY

CKY算法，全称Cocke–Younger–Kasami算法，由三位计算机科学家于1970年提出。

John Cocke
1987 Turing Award

Tadao Kasami
嵩忠雄

Daniel H. Younger



目录

CSG与CFG

上下文相关语法
(Context-sensitive Grammar，CSG)

语法树的生成与上下文相关，即：语法规则中包含“→”号左边不是单个的N中元素情况

VP NP → VP Nominal
NP NP → NP Det Nominal
prefer Pronoun PP → *prefer* Pronoun to Verb Noun
prefer NP→ *prefer* Proper-Noun

约束R中规则的形式

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

CSG与CFG

上下文相关语法
(Context-sensitive Grammar，CSG)

语法树的生成与上下文相关，即：语法规则中包含“→”号左边不是单个的N中元素情况

VP NP → VP Nominal
NP NP → NP Det Nominal
prefer Pronoun PP → *prefer* Pronoun to Verb Noun
prefer NP→ *prefer* Proper-Noun

将R中规则赋予概率

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

CFG与PCFG

概率化的上下文无关语法
（Probablistic CFG，PCFG）

多种合法解析的时候，哪种更好

将R中规则赋予概率

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

R 语法规则，即$\alpha \rightarrow \beta$形式的规则。α与β均可代表由$N \cup \Sigma$中的元素构成的序列

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Nominal Noun
Nominal → Noun
PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP

Noun → flights | breeze | trip | morning
Verb → is | prefer | like | need | want | fly
Adjective → cheapest | non-stop | first | latest | other | direct
Pronoun → me | I | you | it
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American
Determiner → the | a | an | this | these | that
Preposition → from | to | on | near
Conjunction → and | or | but

概率化的上下文无关语法（PCFG）

概率化的上下文无关语法
（Probablistic CFG，PCFG）

在CFG要求的基础上：

对每一条生成规则赋予概率p，表示给定规则左边的constituent之后，这个constituent依照该条语法规则生成子树的条件概率。

即对于由A生成的规则：

$$A \longrightarrow \beta\ [p]$$

有

$$p = \mathrm{P}(\beta|A)$$

S → NP VP [1]
NP → Pronoun [0.3]
NP → Proper-Noun [0.3]
NP → Det Nominal [0.4]

PP → Preposition NP [1]
VP → Verb [0.2]
VP → Verb NP [0.2]
VP → Verb NP PP [0.6]

将R中规则赋予概率

上下文无关语法
(Context-free Grammar，CFG)

语法树的生成与上下文无关，即：每一条语法规则中，“→”号左边均为单个的N中元素情况

S → NP VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal

PP → Preposition NP
VP → Verb
VP → Verb NP
VP → Verb NP PP

概率化的上下文无关语法（PCFG）

概率化的上下文无关语法
（Probablistic CFG，PCFG）

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

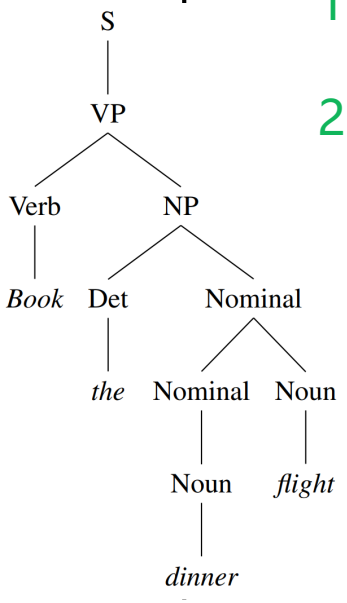
Rules			P
S	→	VP	.05
VP	→	Verb NP NP	.10
NP	→	Det Nominal	.20
NP	→	Nominal	.15
Nominal	→	Noun	.75
Nominal	→	Noun	.75
Verb	→	book	.30
Det	→	the	.60
Noun	→	dinner	.10
Noun	→	flight	.40

1. 判断给定的语法树是否合法。
2. 给定句子，推断出他合法的语法树。
3. 给定句子，在所有合法的语法树中，给出最有可能的语法结构。
4. 像语言模型一样计算给定句子的概率。

上下文无关语法
（Context-free Grammar，CFG）

Rules			Rules		
S	→	VP	S	→	VP
VP	→	Verb NP	VP	→	Verb NP NP
NP	→	Det Nominal	NP	→	Det Nominal
Nominal	→	Nominal Noun	NP	→	Nominal
Nominal	→	Noun	Nominal	→	Noun
			Nominal	→	Noun
Verb	→	book	Verb	→	book
Det	→	the	Det	→	the
Noun	→	dinner	Noun	→	dinner
Noun	→	flight	Noun	→	flight

1. 判断给定的语法树是否合法。
2. 给定句子，推断出他合法的语法树。

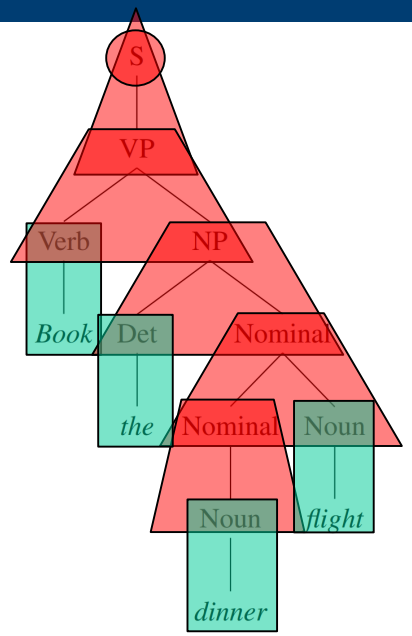


目录

概率化的上下文无关语法（PCFG）

概率化的上下文无关语法
（Probablistic CFG，PCFG）

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



给定左边语法规则，计算右边句子(V)及其对应的语法树(T)出现的概率。

$$P(T,V) = P(S)P(S \to VP|S)P(VP \to Verb\ NP|VP)P(Nominal \to Det\ Nominal|NP)$$
$$P(Nominal \to Nominal\ Noun|Nominal)$$
$$P(Nominal \to Noun|Nominal)$$
$$P(Verb \to Book|Verb)P(Det \to The|Det)P(Noun \to dinner|Noun)$$
$$P(Noun \to flight|Noun)$$

所以把P（S）扔掉

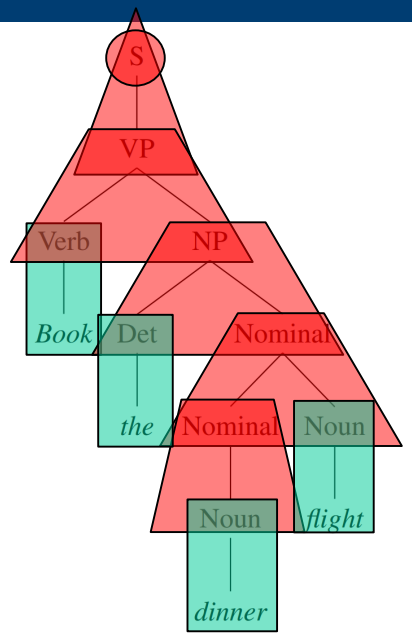
$$P(S) = 1$$

概率化的上下文无关语法（PCFG）

概率化的上下文无关语法
（Probablistic CFG，PCFG）

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

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



给定左边语法规则，计算右边句子(V)及其对应的语法树(T)出现的概率。

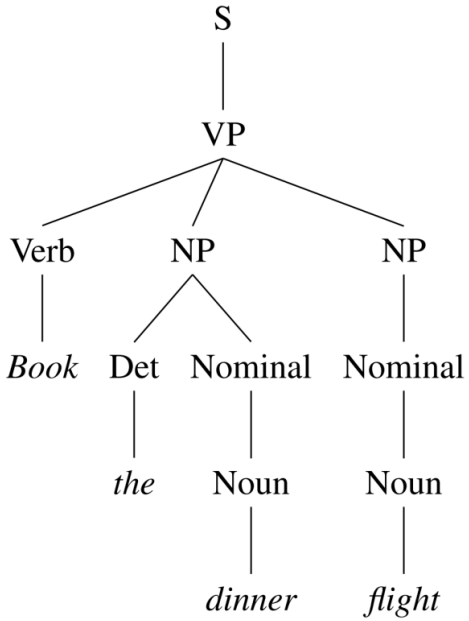
$$P(T,V) = P(S \to VP|S)P(VP \to Verb\ NP|VP)P(Nominal \to Det\ Nominal|NP)$$
$$P(Nominal \to Nominal\ Noun|Nominal)$$
$$P(Nominal \to Noun|Nominal)$$
$$P(Verb \to Book|Verb)P(Det \to The|Det)P(Noun \to dinner|Noun)$$
$$P(Noun \to flight|Noun)$$

$$P(T,V) = 0.05 \times 0.2 \times 0.2 \times 0.2 \times 0.75 \times 0.3 \times 0.6 \times 0.1 \times 0.4 = 2.2 \times 10^{-6}$$

概率化的上下文无关语法（PCFG）

概率化的上下文无关语法
（Probablistic CFG，PCFG）

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



给定左边语法规则，计算右边句子(V)及其对应的语法树(T)出现的概率。

$$P(T,V) =$$

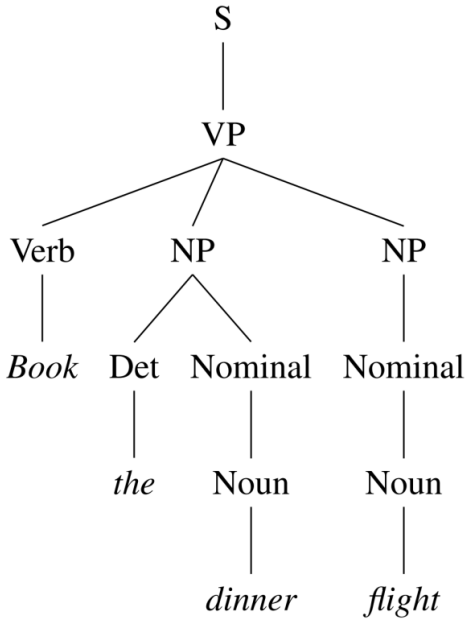
$$=$$

目录

概率化的上下文无关语法（PCFG）

概率化的上下文无关语法
（Probablistic CFG，PCFG）

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



给定一个句子，推断出其概率最大的语法树(T)即是这样一个优化问题：

$$\hat{T}(V) = \underset{T\ s.t.\ y(T)=V}{\operatorname{argmax}} P(T|V)$$

我们可以通过Probablistic CKY算法来求解这个argmax()的结果

目录

PCFG的语法解析：Probablistic CKY

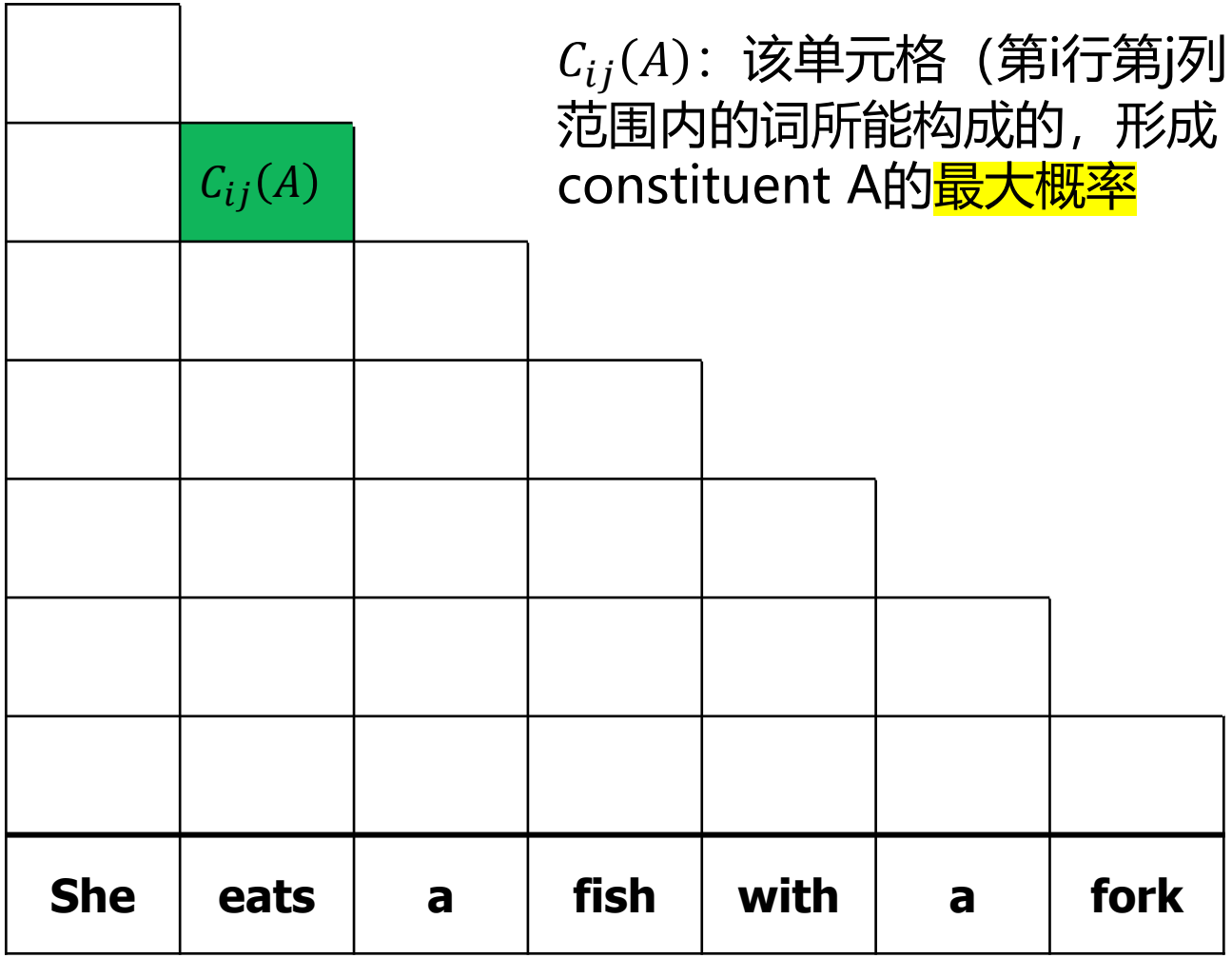
例子：

$$
\begin{aligned}
S &\longrightarrow NP\ VP \\
VP &\longrightarrow VP\ PP \\
VP &\longrightarrow V\ NP \\
PP &\longrightarrow P\ NP \\
NP &\longrightarrow Det\ N
\end{aligned}
$$

$$
\begin{aligned}
VP &\longrightarrow \text{eats} \\
NP &\longrightarrow \text{she} \\
V &\longrightarrow \text{eats} \\
P &\longrightarrow \text{with} \\
N &\longrightarrow \text{fish} \\
N &\longrightarrow \text{fork} \\
Det &\longrightarrow \text{a}
\end{aligned}
$$

She eats a fish with a fork

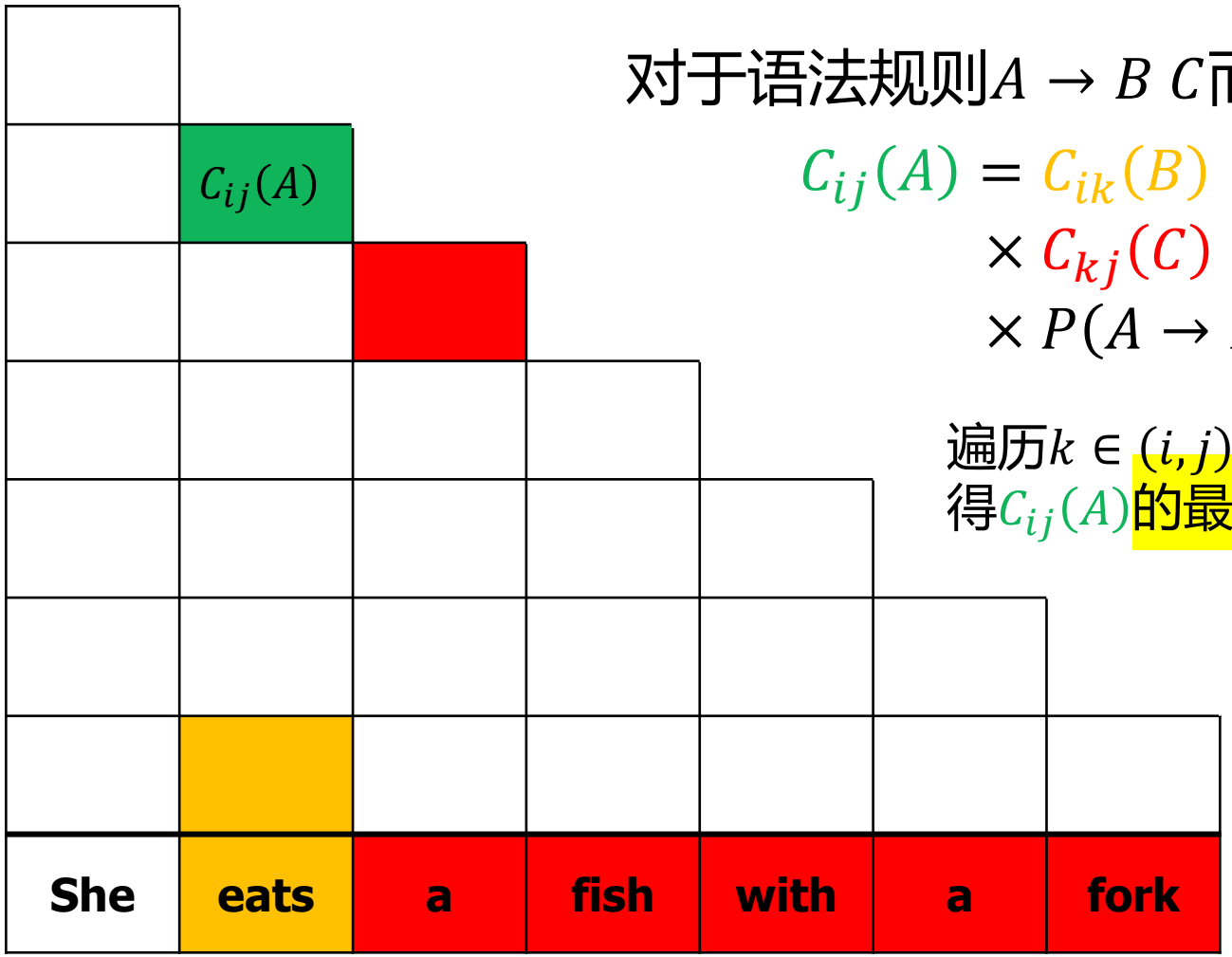
PCFG的语法解析：Probablistic CKY



$C_{ij}(A)$：该单元格（第i行第j列）所对应范围内的词所能构成的，形成constituent A的最大概率

PCFG的语法解析：Probablistic CKY

例子：



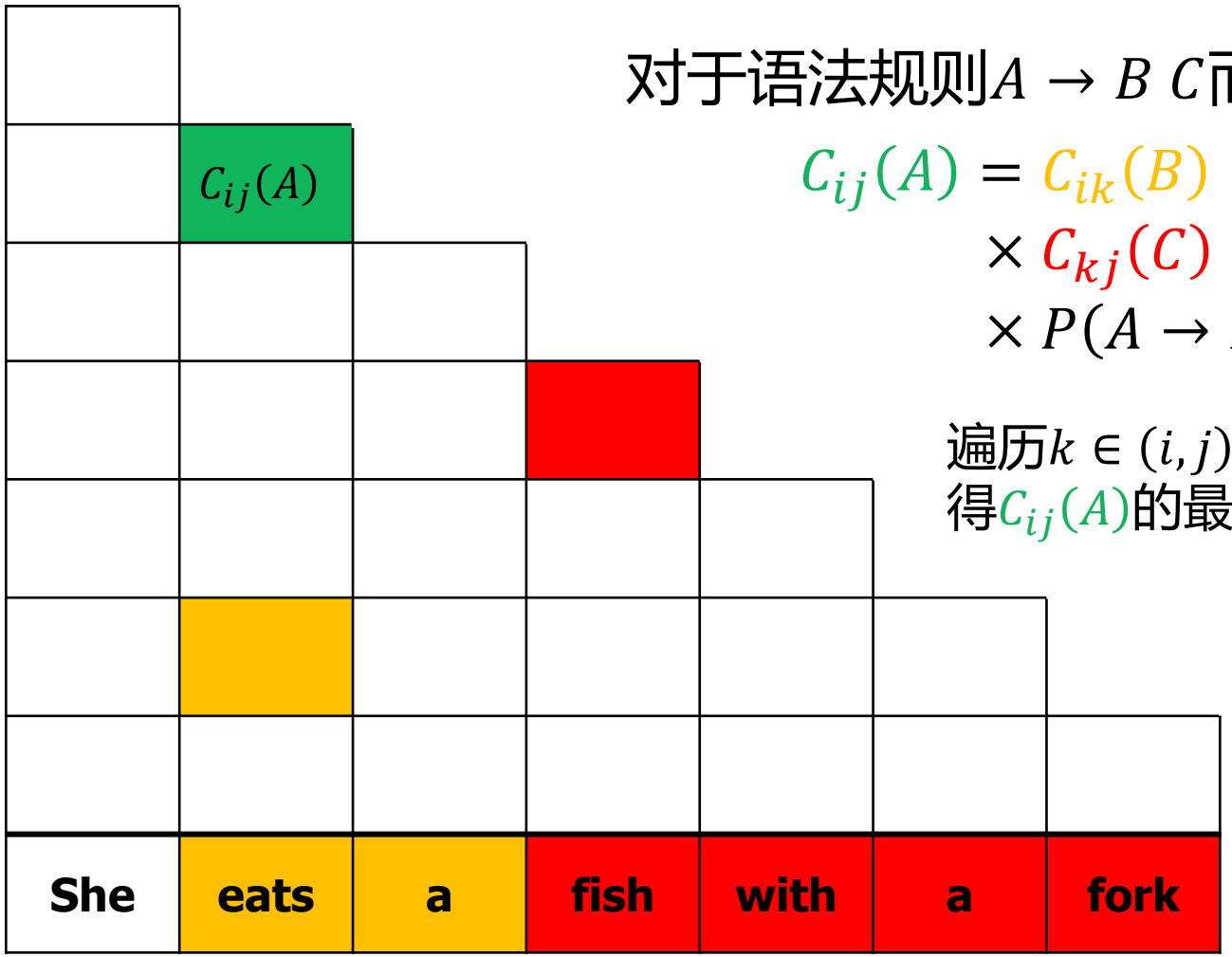
对于语法规则$A \rightarrow B\ C$而言：

$$C_{ij}(A) = C_{ik}(B) \times C_{kj}(C) \times P(A \rightarrow B\ C|A)$$

遍历$k \in (i,j)$即可求得$C_{ij}(A)$的最大值

PCFG的语法解析：Probablistic CKY

例子：



概率的算法

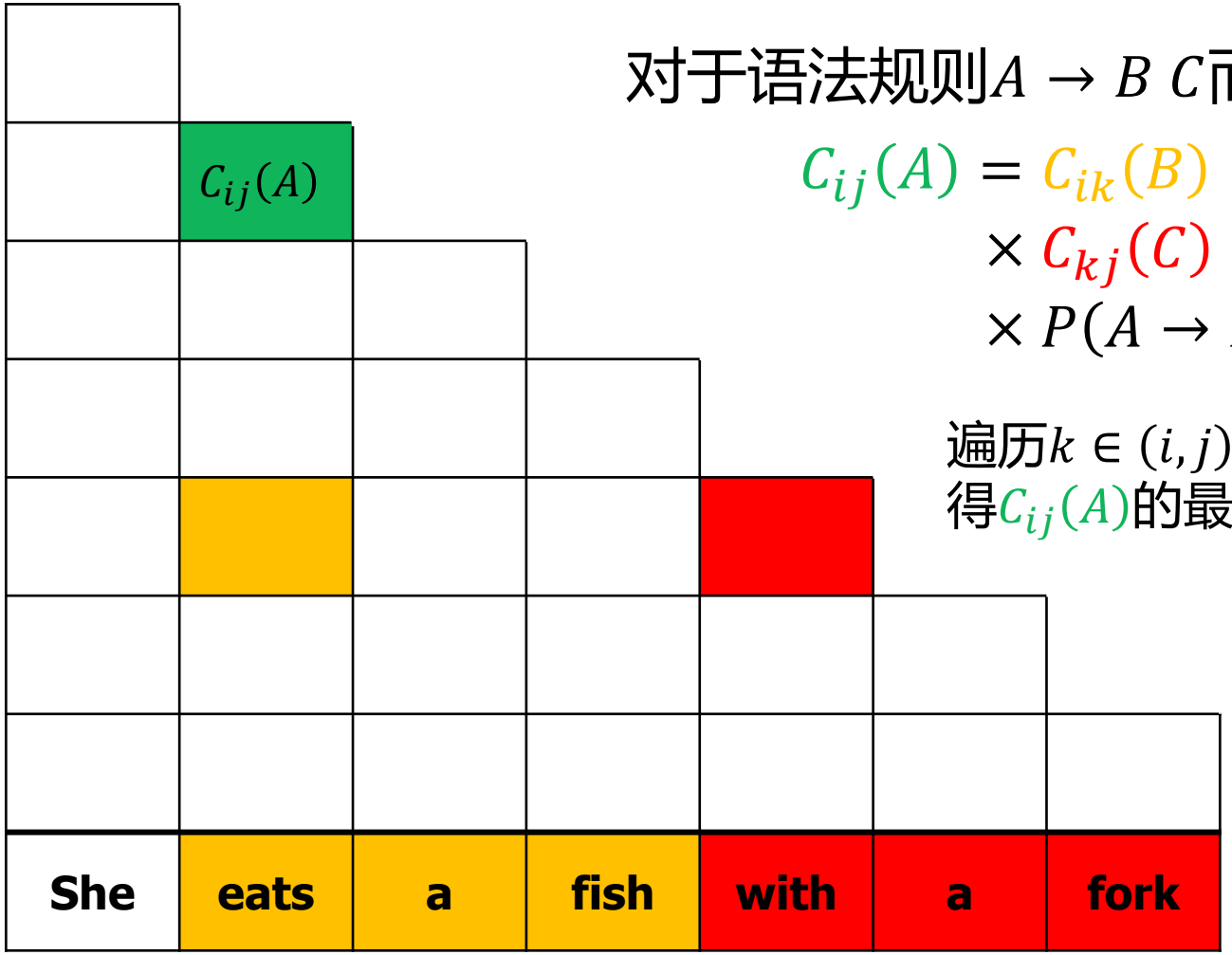
对于语法规则$A \rightarrow B\ C$而言：

$$C_{ij}(A) = C_{ik}(B) \times C_{kj}(C) \times P(A \rightarrow B\ C|A)$$

遍历$k \in (i,j)$即可求得$C_{ij}(A)$的最大值

PCFG的语法解析：Probablistic CKY

例子：



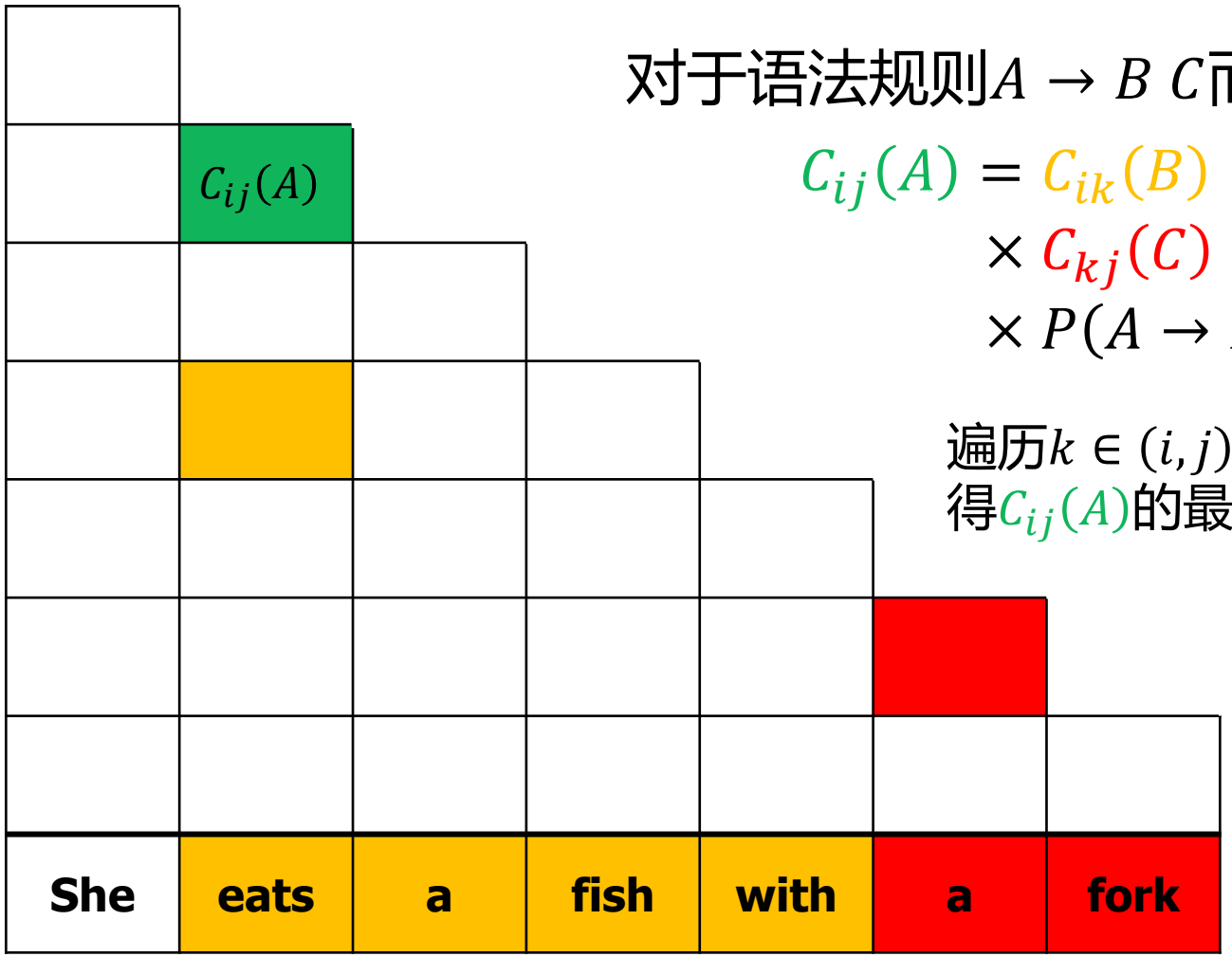
对于语法规则$A \rightarrow B\ C$而言：

$$C_{ij}(A) = C_{ik}(B) \times C_{kj}(C) \times P(A \rightarrow B\ C|A)$$

遍历$k \in (i,j)$即可求得$C_{ij}(A)$的最大值

PCFG的语法解析：Probablistic CKY

例子：



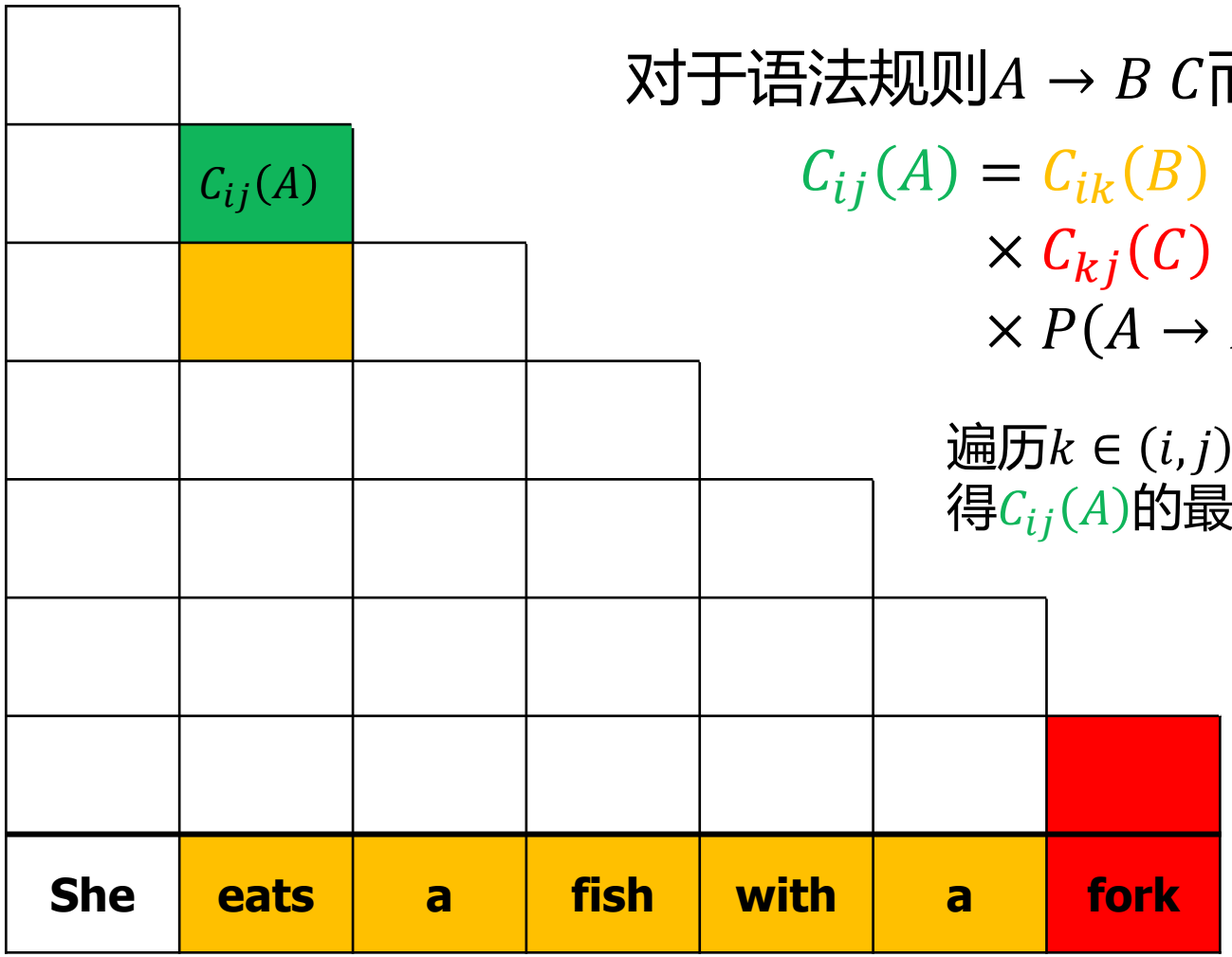
对于语法规则$A \rightarrow B\ C$而言：

$$C_{ij}(A) = C_{ik}(B) \times C_{kj}(C) \times P(A \rightarrow B\ C|A)$$

遍历$k \in (i,j)$即可求得$C_{ij}(A)$的最大值

PCFG的语法解析：Probablistic CKY

例子：



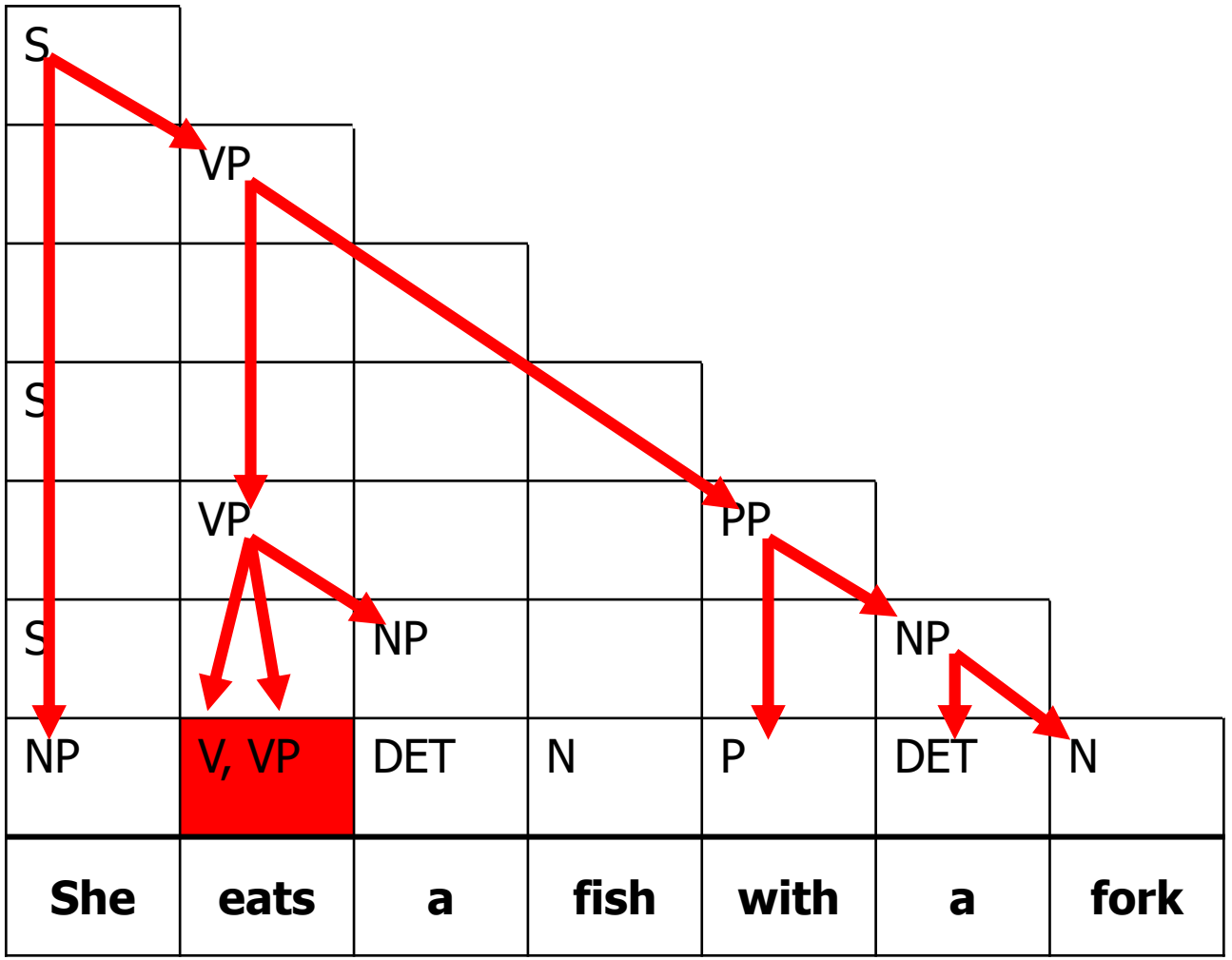
对于语法规则$A \rightarrow B\ C$而言：

$$C_{ij}(A) = C_{ik}(B) \times C_{kj}(C) \times P(A \rightarrow B\ C|A)$$

遍历$k \in (i,j)$即可求得$C_{ij}(A)$的最大值

构成式语法的语法解析算法：CKY

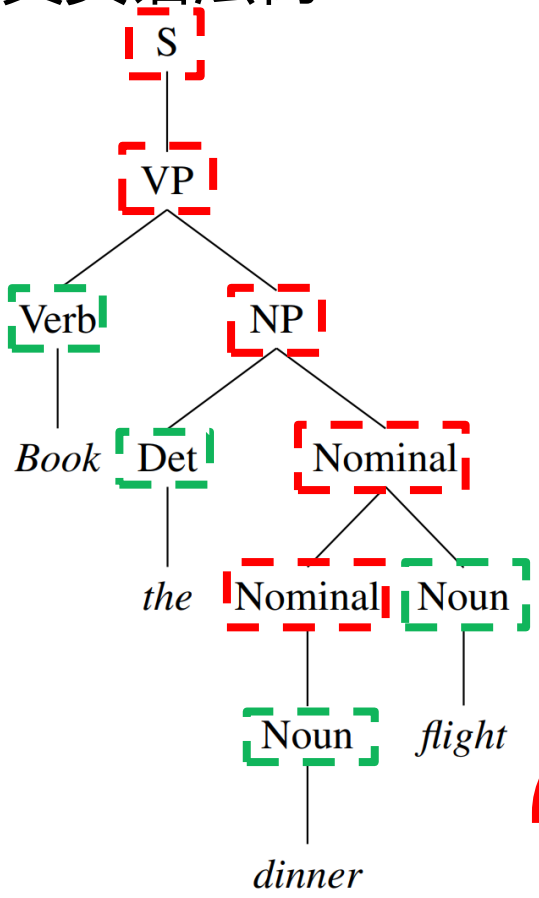
例子：



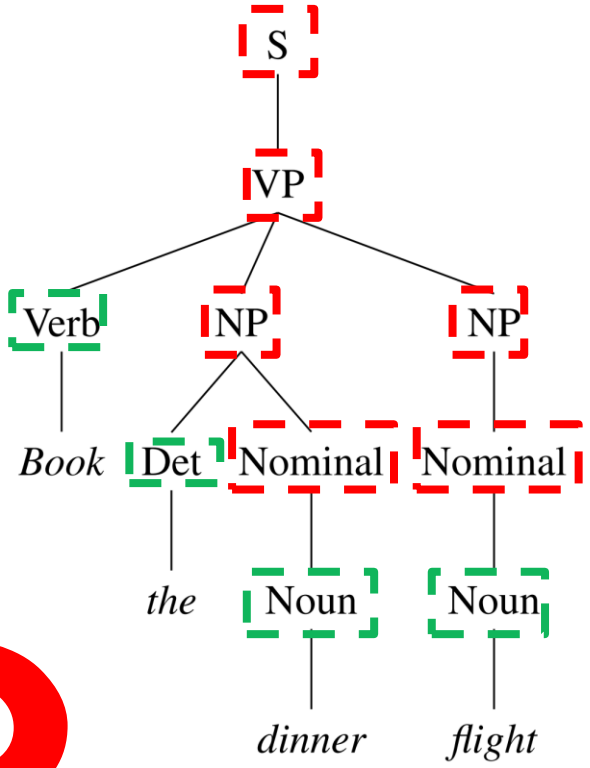
目录

评价指标

真实语法树



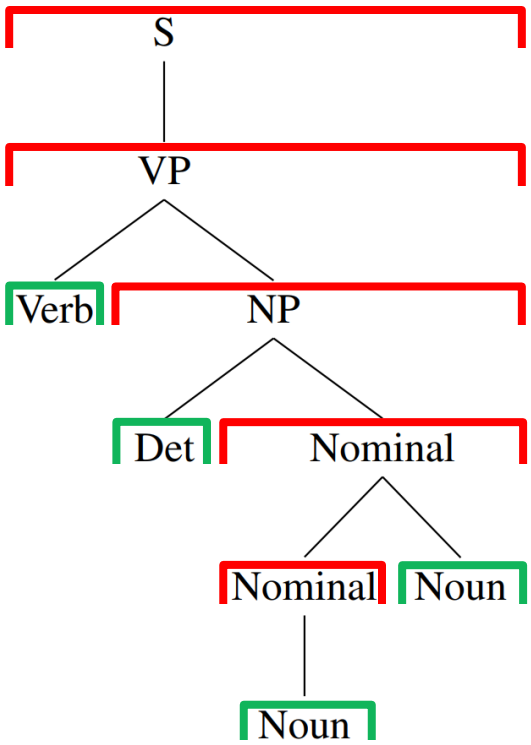
模型预测的语法树



评价指标

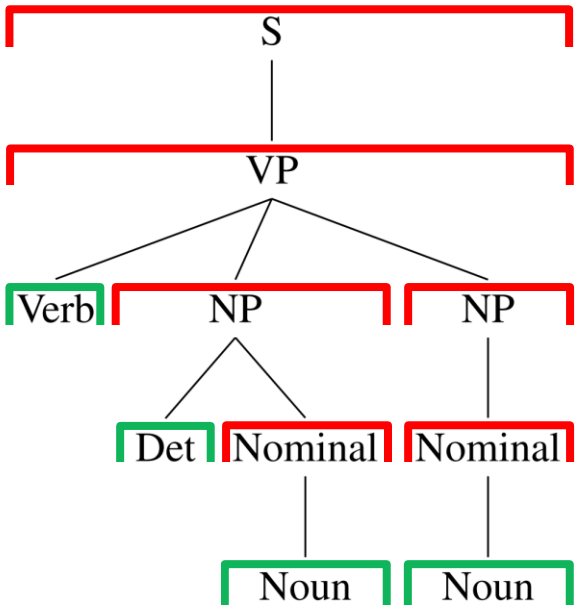
是否用了过多的constituent？

真实语法树



Book the dinner flight

模型预测的语法树



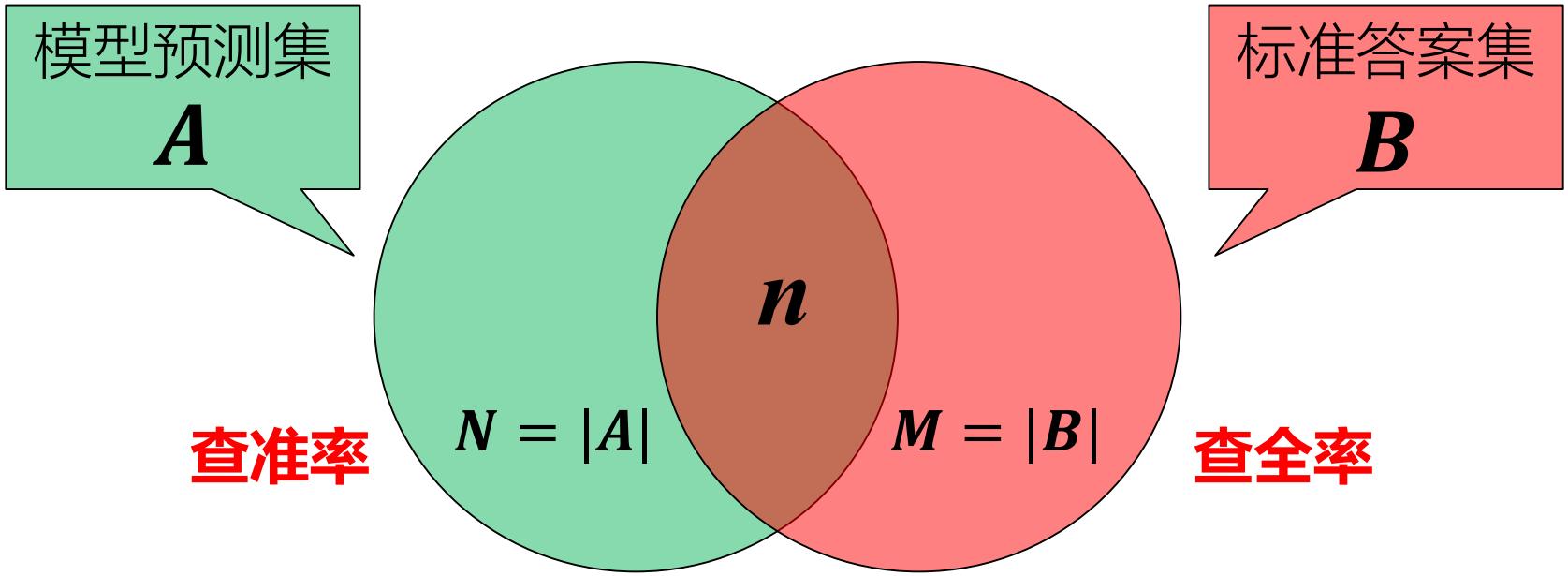
Book the dinner flight

考虑每一个constituent的涵盖范围。对于某个constituent而言，若能在另一棵语法树中找到一个涵盖范围与标签均相同的constituent，则视为一个匹配；否则视为不匹配。

→ ∀CNF，若两棵树中若所有的constituent都匹配，则必然语法树完全相同。

中文分词的评价指标：F-measure

假设系统输出N个结果，其中，正确的结果为n个，标准答案的个数为M个



$$F_1 = \frac{2PR}{P+R} \times 100\%$$

$$P = \frac{n}{N} \times 100\%$$

$$R = \frac{n}{M} \times 100\%$$

评价指标

使用 {模型的预测的constituents的集合} 与 {真实语法树中的constituents} 两个集合的重合程度，即F_1分数，来量化评价语法解析的质量。

?

- Labeled F_1：考虑constituents涵盖范围以及标签。仅当两方面都正确时，才视作匹配
- Unlabeled F_1：只考虑constituents涵盖范围，即便标签预测错误，也视作是匹配

目录

常用工具

在线语法解析demo：

https://parser.kitaev.io/（Berkeley Neural Parser）
https://www.link.cs.cmu.edu/link/submit-sentence-4.html
https://corenlp.run/（Stanford Parser）

语法解析评价指标的标准实现：

Evalb (https://nlp.cs.nyu.edu/evalb/)

repo和模型：

Model	F1 score	Paper / Source
Label Attention Layer + HPSG + XLNet (Mrini et al., 2020)	96.38	Rethinking Self-Attention: Towards Interpretability for Neural Parsing
Attach-Juxtapose Parser + XLNet (Yang and Deng, 2020)	96.34	Strongly Incremental Constituency Parsing with Graph Neural Networks
Head-Driven Phrase Structure Grammar Parsing (Joint) + XLNet (Zhou and Zhao, 2019)	96.33	Head-Driven Phrase Structure Grammar Parsing on Penn Treebank
Head-Driven Phrase Structure Grammar Parsing (Joint) + BERT (Zhou and Zhao, 2019)	95.84	Head-Driven Phrase Structure Grammar Parsing on Penn Treebank
CRF Parser + BERT (Zhang et al., 2020)	95.69	Fast and Accurate Neural CRF Constituency Parsing