ОСНОВЫ ОБРАБОТКИ ТЕКСТОВ

Лекция #7:
Синтаксический анализ

Лектор: м.н.с. ИСП РАН Майоров Владимир Дмитриевич

Синтаксис

- Предложение – это единица языка, которая представляет собой грамматически организованное соединение слов, обладающее смысловой законченностью.

- Грамматика – раздел лингвистики, который изучает закономерности построения правильных осмысленных речевых отрезков (словоформ, словосочетаний, предложений, текстов).
- Синтаксис — раздел лингвистики, изучающий и моделирующий правила, по которым образуются единицы, более крупные, чем слово, а именно словосочетания и предложения.

Синтаксические правила

- Существительное сочетается с прилагательным

быстрый + бег → быстрый бег

быстро + бег → не словосочетание

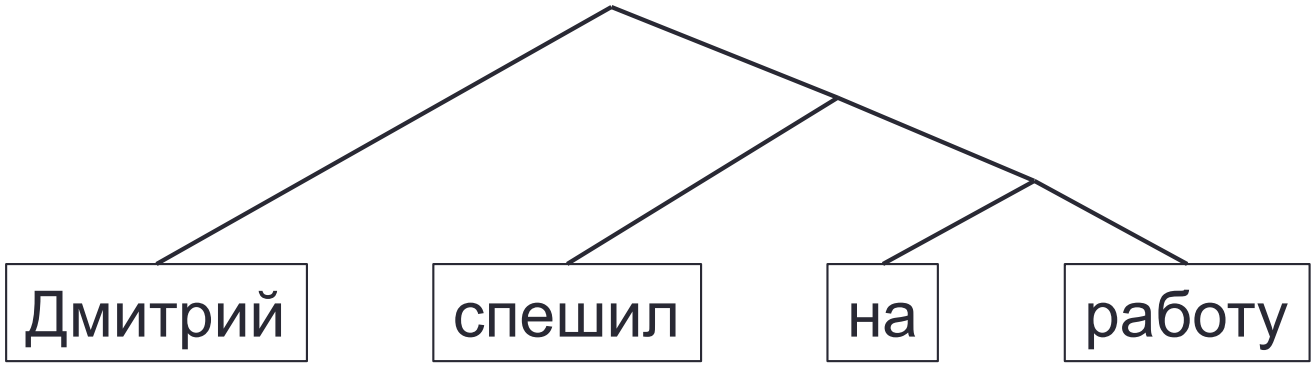
- Глагол сочетается с наречием

бегать + быстрый → не словосочетание

бегать + быстро → бегать быстро

Синтаксическая структура

- Представление предложения в виде вложенных составляющих.
- Составляющая (фраза, синтаксическая группа) – структурная единица предложения, составленная из более тесно связанных друг с другом составляющих меньшего размера.



Синтаксические правила

- Существительное управляет прилагательным

красная книга → красная книга

красный книга → не словосочетание

- Предлог управляет существительным

в книга → не словосочетание

в книге → в книге

Синтаксические правила

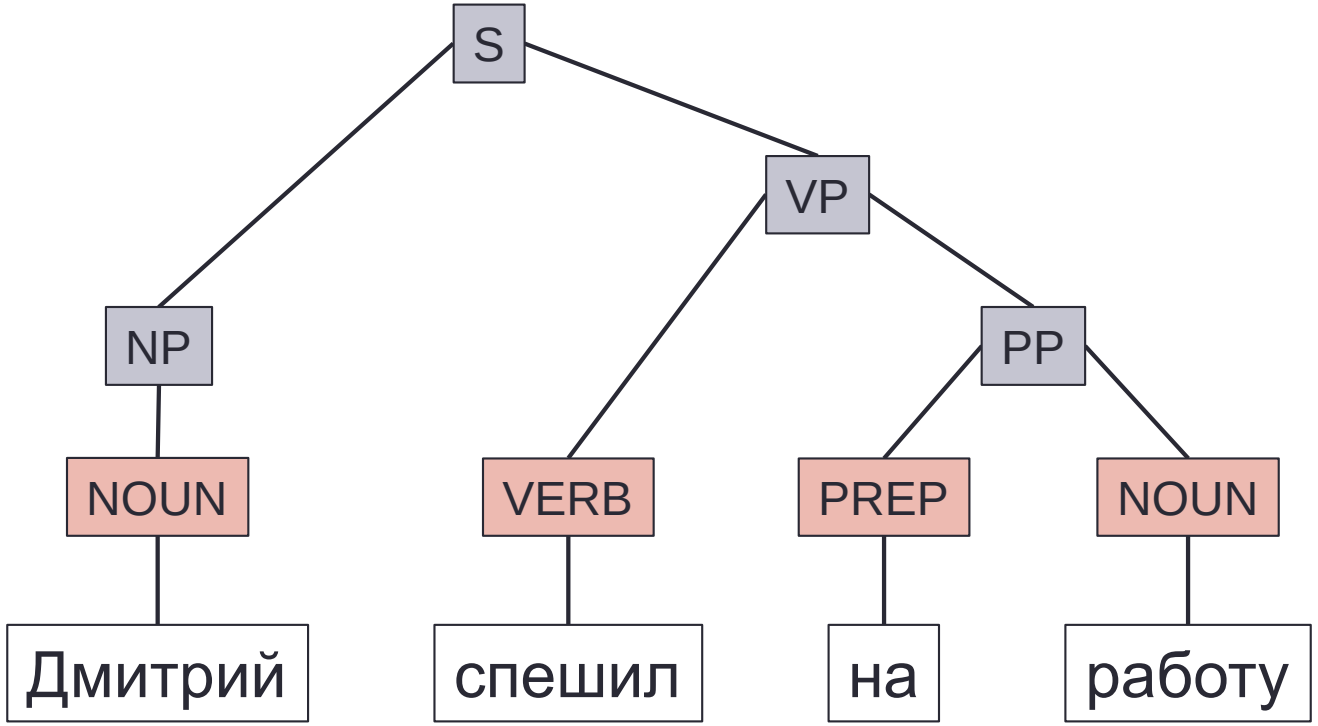
- В зависимости от главного слова в словосочетании выделяют
  - Именные группы (главное – существительное)
  - Группа прилагательного
  - Наречная группа
  - Предложная группа
  - Глагольная группа
  - Предложение

Синтаксические правила

- В зависимости от главного слова в словосочетании выделяют
  - Именные группы (главное – существительное)
  - Группа прилагательного
  - Наречная группа
  - Предложная группа
  - Глагольная группа
  - Предложение

Синтаксическая структура

- Каждой составляющей назначается тип в зависимости от главного слова



Формальная грамматика

- Способ описания формального языка.
- Формальная грамматика состоит из:
  - Σ – множество терминальных символов
  - N – множество нетерминальных символов
  - P – набор правил вывода $\alpha \longrightarrow \beta$, где
    - α – последовательность символов из $\Sigma \cup N$, хотя бы один из N
    - β – последовательность символов из $\Sigma \cup N$
  - S – начальный символ (из N)

Формальная грамматика

- P – набор правил вывода $\alpha \longrightarrow \beta$, где
  - α – последовательность символов из $\Sigma \cup N$, хотя бы один из N
  - β – последовательность символов из $\Sigma \cup N$

- Иерархия Хомского
  - тип 0. неограниченные грамматики
    любые правила
  - тип 1. контекстно-зависимые грамматики
    правила вида $c_1 A c_2 \longrightarrow c_1 \beta c_2$
  - тип 2. контекстно-свободные грамматики
    правила вида $A \longrightarrow \beta$
  - тип 3. регулярные грамматики
    правила вида $A \longrightarrow a$, $A \longrightarrow aB$ или $A \longrightarrow \varepsilon$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

S

S

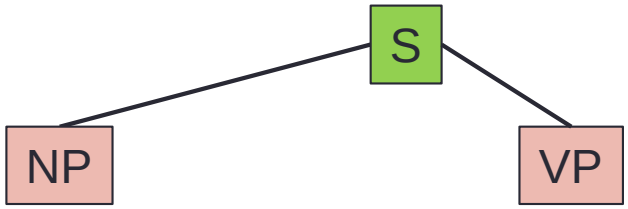
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$S \xrightarrow{1} NP\ VP$



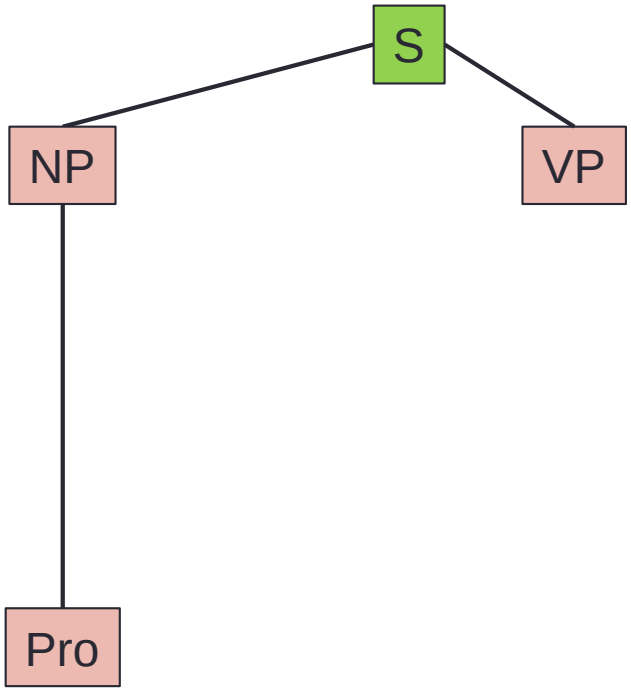
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$NP\ VP \overset{4}{\rightarrow} Pro\ VP$



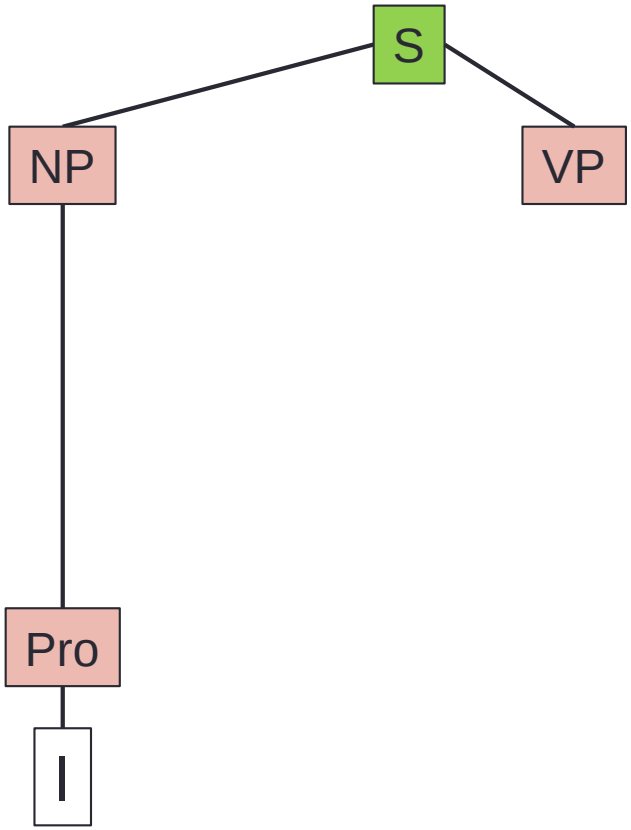
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $\underline{NP \rightarrow Pro}$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$Pro\ VP \overset{12}{\rightarrow} I\ VP$



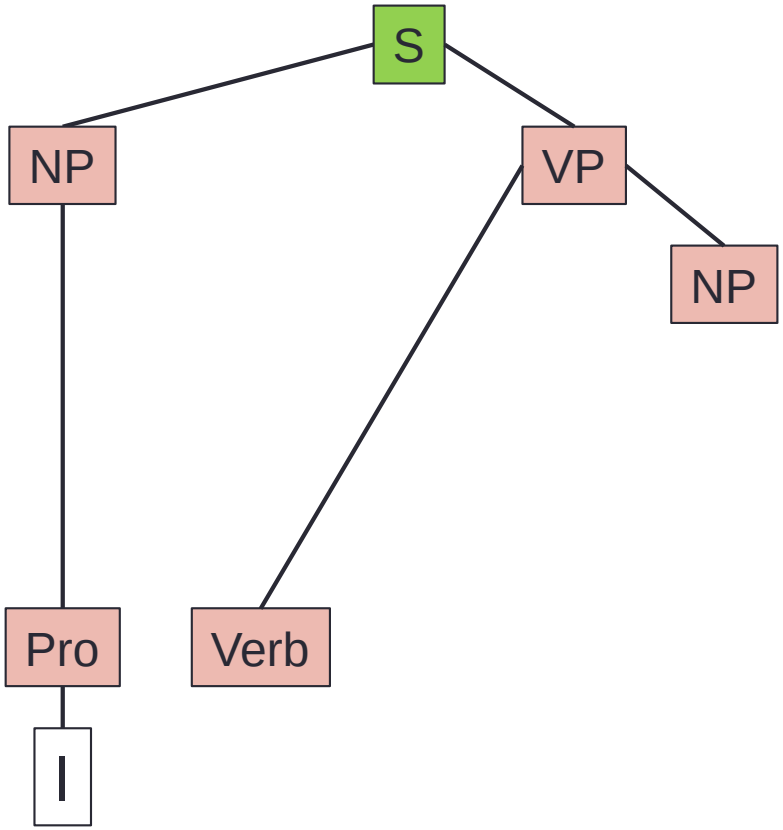
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $\underline{Pro \rightarrow I}$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$I\ VP \overset{2}{\rightarrow} I\ Verb\ NP$



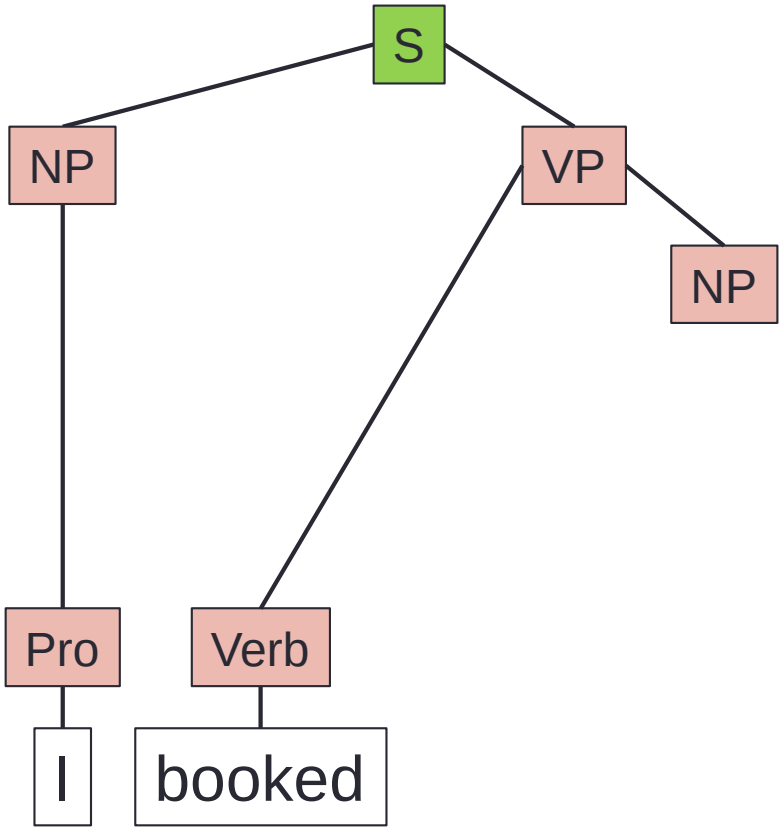
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$I\ Verb\ NP \overset{9}{\rightarrow} I\ booked\ NP$



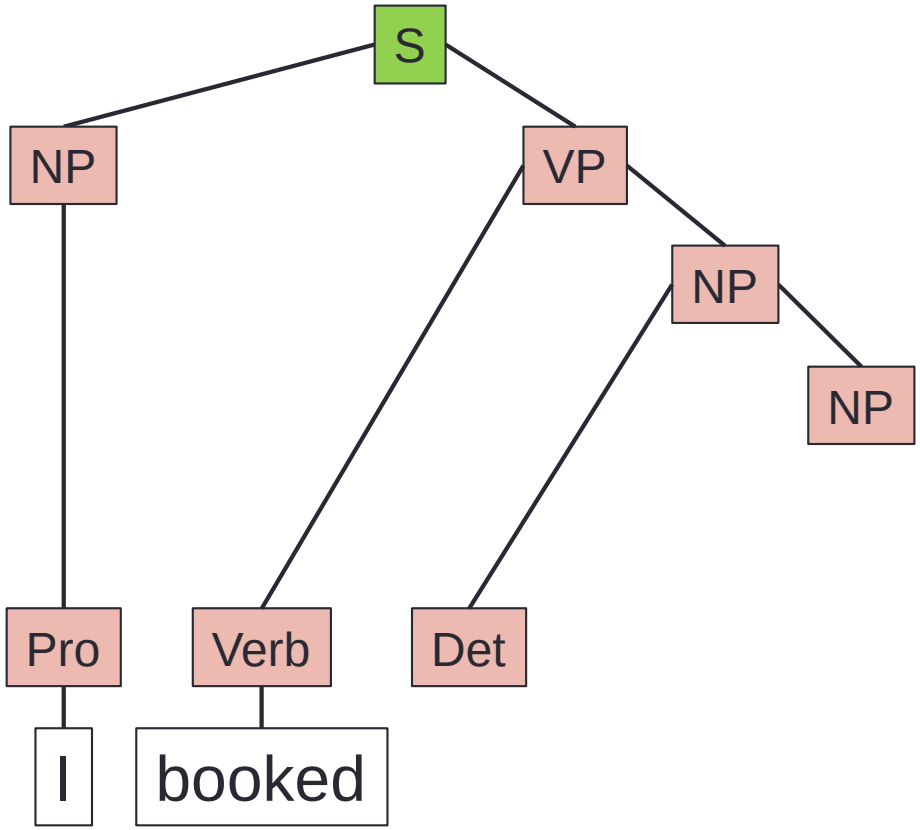
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $\underline{Verb \rightarrow booked}$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$I\ booked\ NP \overset{5}{\rightarrow} I\ booked\ Det\ NP$



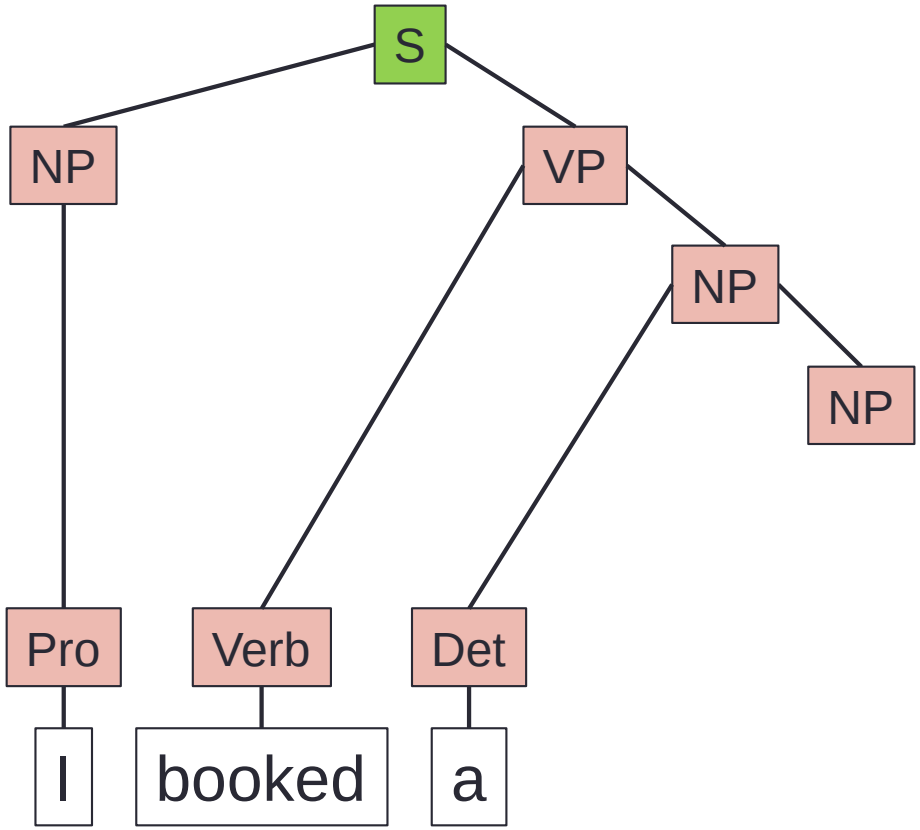
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $\underline{NP \rightarrow Det\ NP}$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$I\ booked\ Det\ NP \overset{13}{\rightarrow} I\ booked\ a\ NP$



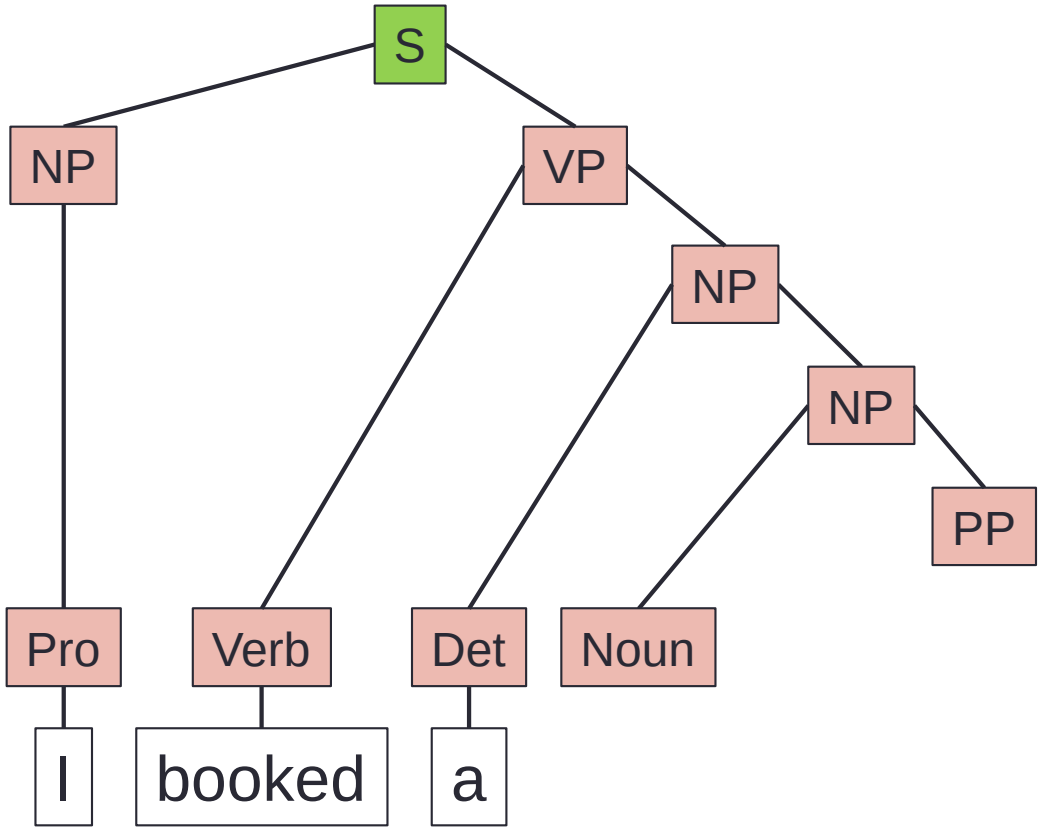
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $\underline{Det \rightarrow a}$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$I\ booked\ a\ NP \overset{6}{\rightarrow} I\ booked\ a\ Noun\ PP$



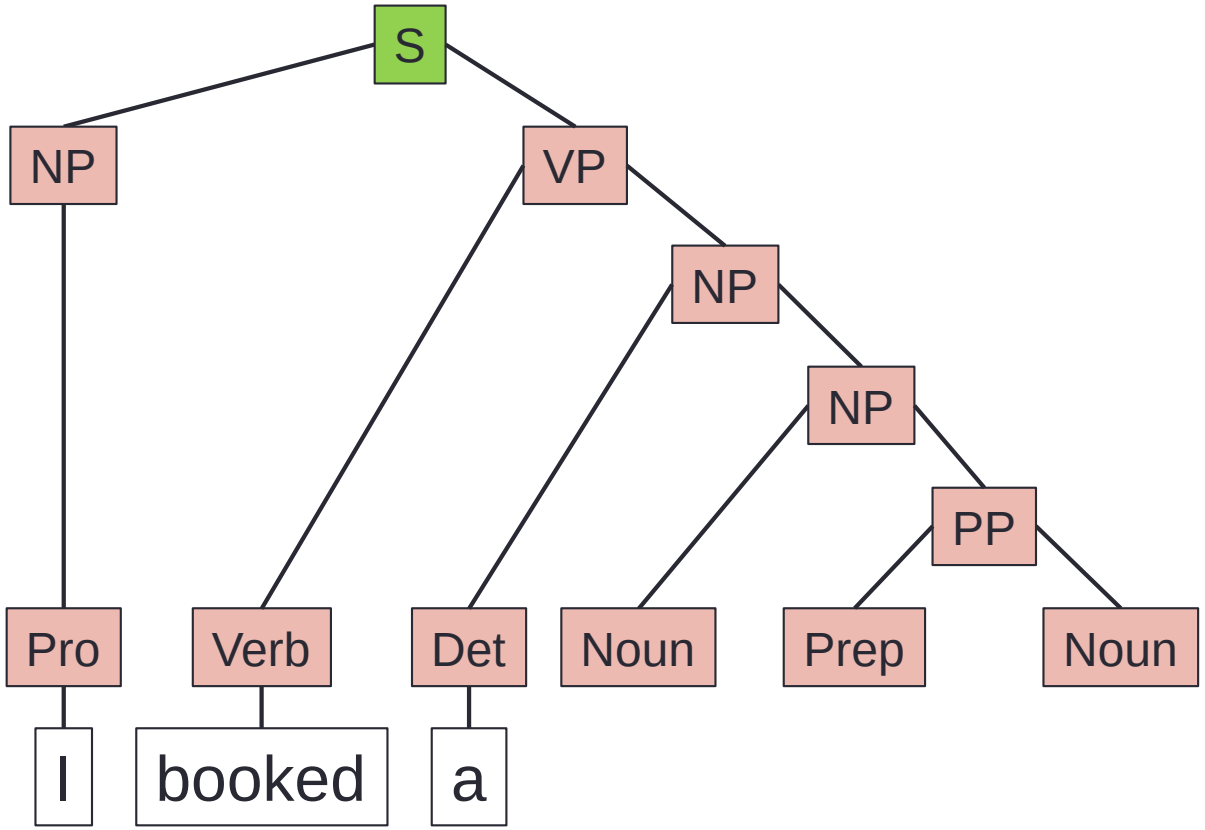
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $\underline{NP \rightarrow Noun\ PP}$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$I\ booked\ a\ Noun\ PP \overset{8}{\rightarrow} I\ booked\ a\ Noun\ Prep\ Noun$



1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $\underline{PP \rightarrow Prep\ Noun}$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

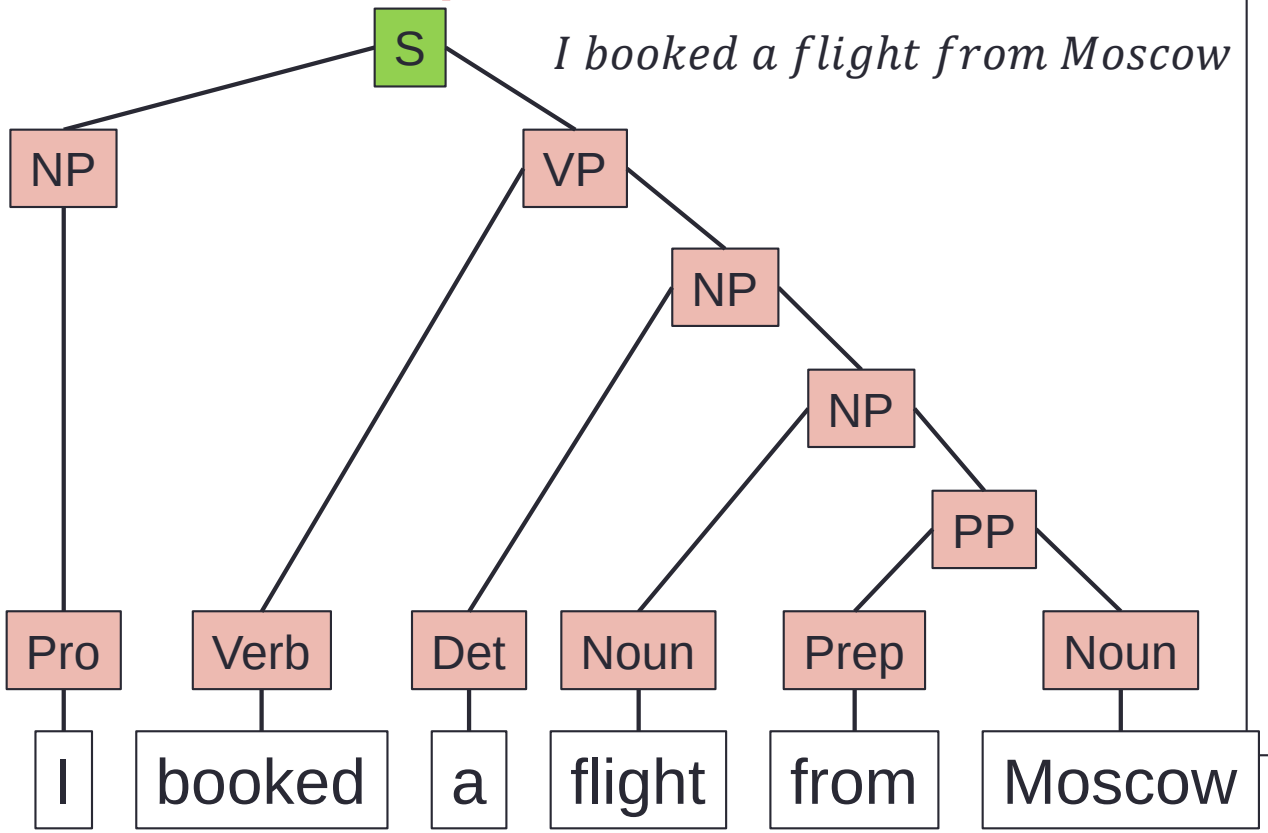
Формальная грамматика

$\Sigma = \{booked, flight, Moscow, I, a, from\}$

$N = \{S, VP, NP, PP, Verb, Noun, Pro, Det, Prep\}$

$I\ booked\ a\ Noun\ Prep\ Noun \overset{10}{\rightarrow} \overset{15}{\rightarrow} \overset{11}{\rightarrow}$

$I\ booked\ a\ flight\ from\ Moscow$



1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Построение грамматики

```
(S
        (NP
                (Pro I))
        (VP
                (Verb booked)
                (NP
                        (Det a)
                        (NP
                                (Noun flight)
                                (PP
                                        (Prep from)
                                        (Noun Moscow))))))
```

Построение грамматики

```
(S                                                    1.  S → NP VP
        (NP
                (Pro I))
        (VP
                (Verb booked)
                (NP
                        (Det a)
                        (NP
                                (Noun flight)
                                (PP
                                        (Prep from)
                                        (Noun Moscow))))))
```

Построение грамматики

```
(S
        (NP
                (Pro I))
        (VP
                (Verb booked)
                (NP
                        (Det a)
                        (NP
                                (Noun flight)
                                (PP
                                        (Prep from)
                                        (Noun Moscow))))))
```

1. $S \rightarrow NP\ VP$
2. $NP \rightarrow Pro$

Построение грамматики

```
(S
        (NP
                (Pro I))
        (VP
                (Verb booked)
                (NP
                        (Det a)
                        (NP
                                (Noun flight)
                                (PP
                                        (Prep from)
                                        (Noun Moscow))))))
```

1. $S \rightarrow NP\ VP$
2. $NP \rightarrow Pro$
3. $Pro \rightarrow I$

Построение грамматики

```
(S
        (NP
                (Pro I))
        (VP
                (Verb booked)
                (NP
                        (Det a)
                        (NP
                                (Noun flight)
                                (PP
                                        (Prep from)
                                        (Noun Moscow))))))
```

1. $S \rightarrow NP\ VP$
2. $NP \rightarrow Pro$
3. $Pro \rightarrow I$
4. $VP \rightarrow Verb\ NP$

Разбор предложения

- На входе: последовательность символов из Σ (слова)
- На выходе:
  - Синтаксическое дерево
  - Последовательность правил для генерации входной строки

- Методы построения дерева:
  - Сверху-вниз

    Из начального символа S, вывести входную строку
  - Снизу-вверх

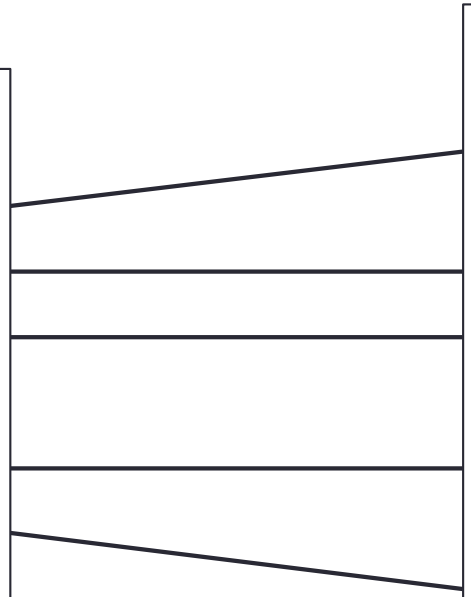
    Из входной строки вывести начальный символ S

Алгоритм CYK

- Для работы необходима КС грамматика в нормальной форме Хомского (CNF)
  - Все правила должны иметь вид: $A \longrightarrow BC$ или $A \longrightarrow \beta$
  - Любая КС грамматика может быть преобразована в CNF (кроме правила $S \longrightarrow \varepsilon$)
- Алгоритм позволяет определить выводимо ли предложение в заданной КС грамматике
- При небольшой модификации алгоритм строит все возможные разборы предложения

Алгоритм CYK

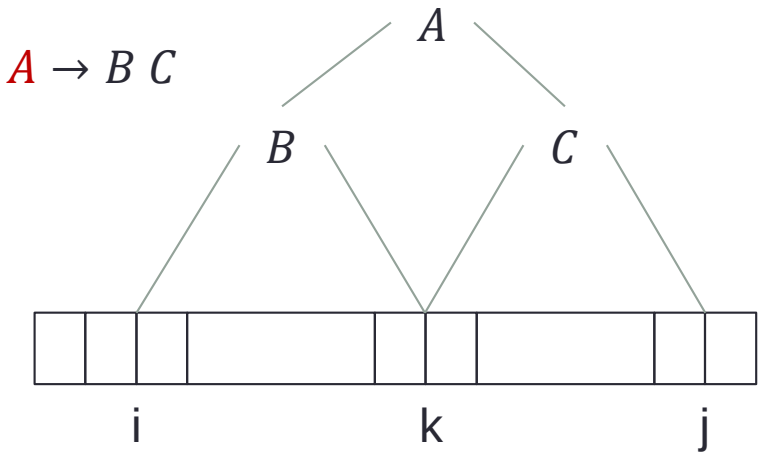
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ NP\ PP$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Verb \rightarrow booked$
10. $Noun \rightarrow flight$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$



1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK

- Идея алгоритма заключается в рассмотрении всех подстрок исходной строки
  - Для каждой подстроки определить, можно ли вывести ее в заданной грамматике, зная какие подстроки меньшего размера можно вывести в этой грамматике

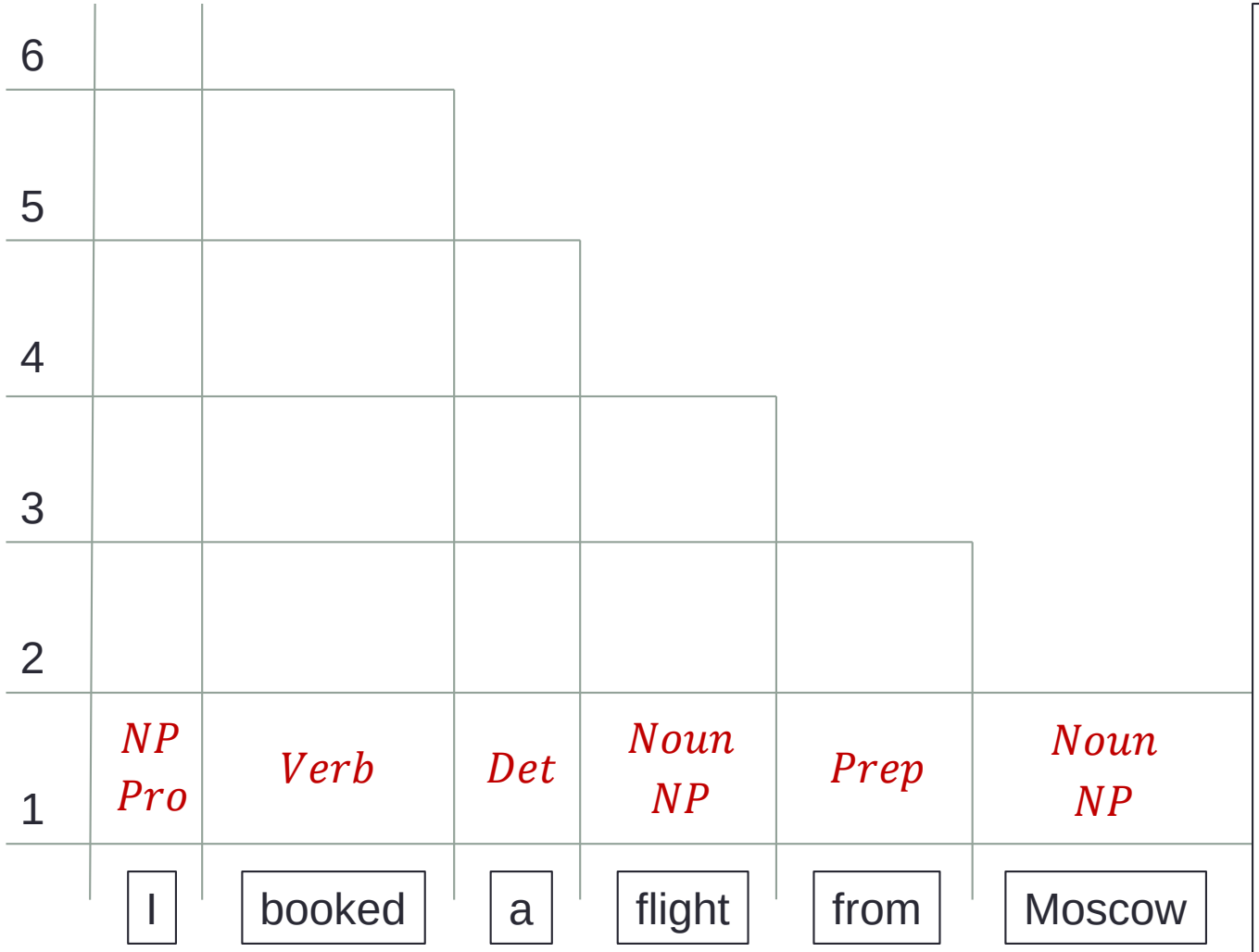


Алгоритм CYK

- Дана грамматика в CNF, N нетерминальных символов
- Дана строка w длины n
- Зададим трехмерный массив $d \in B^{N \times n \times n}$,
  $d[A][i][j] = 1$, если из A можно вывести строку $w[i..j]$

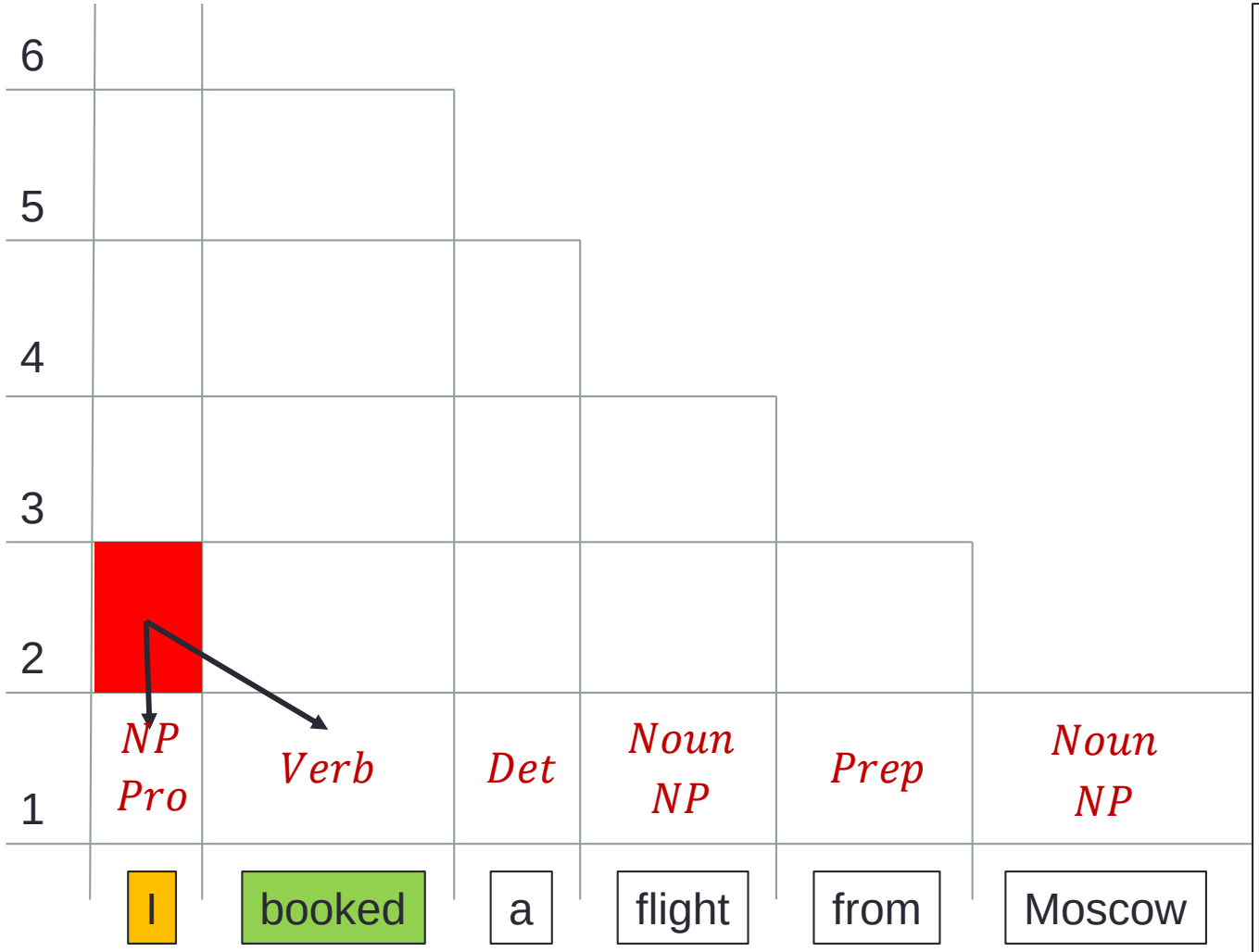
- Решаем задачу динамическим программированием
  - Будем рассматривать все подстроки длины m для $m = \overline{1..n}$
  - Для каждой подстроки [i..j] длины m рассмотрим все разбиения [i..k][k+1..j]
  - Для каждого правила $A \to B\ C$, если
    $d[B, i, k] = 1$ и $d[C, k+1, j] = 1$, то $d[A, i, j] = 1$

Алгоритм CYK



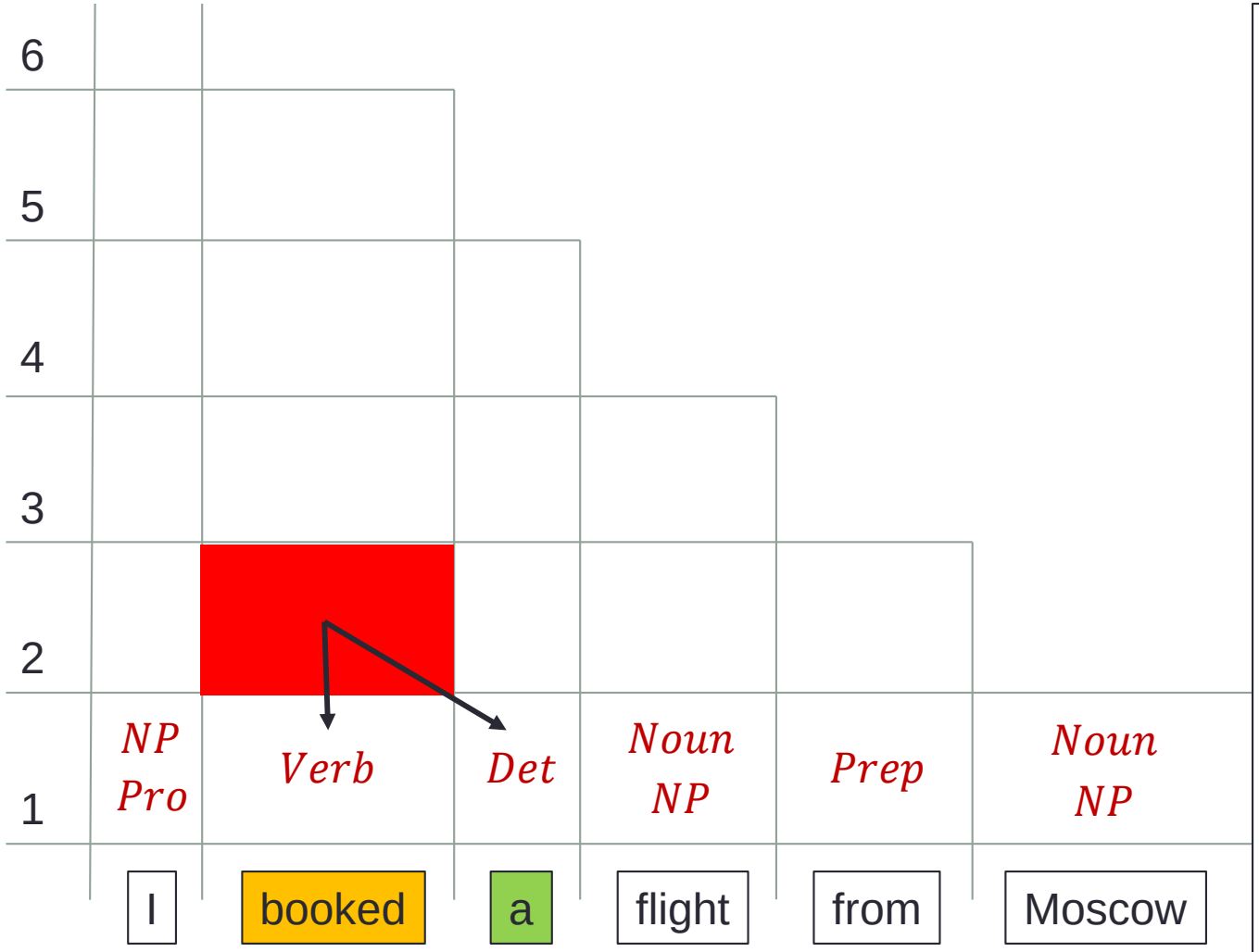
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



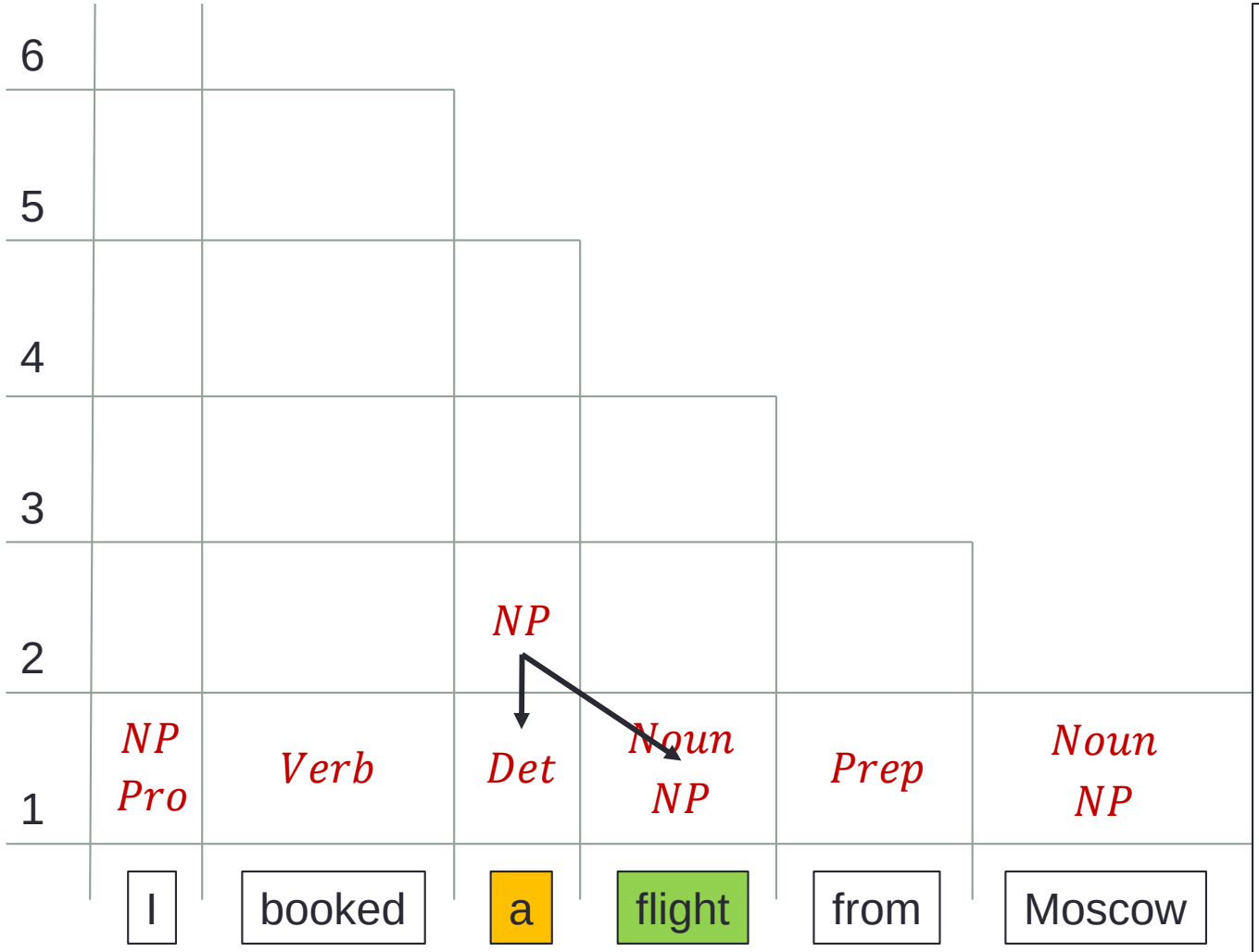
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



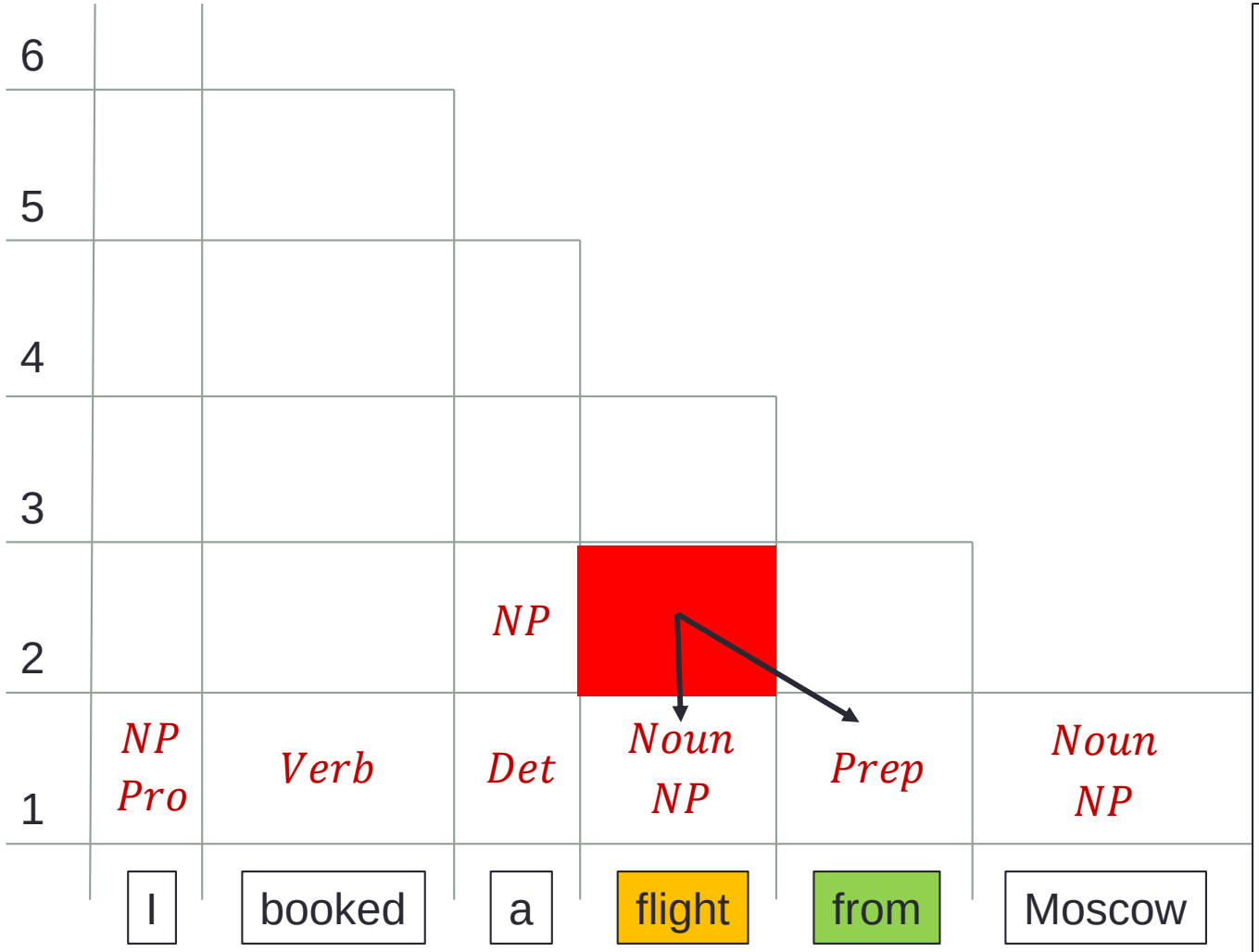
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



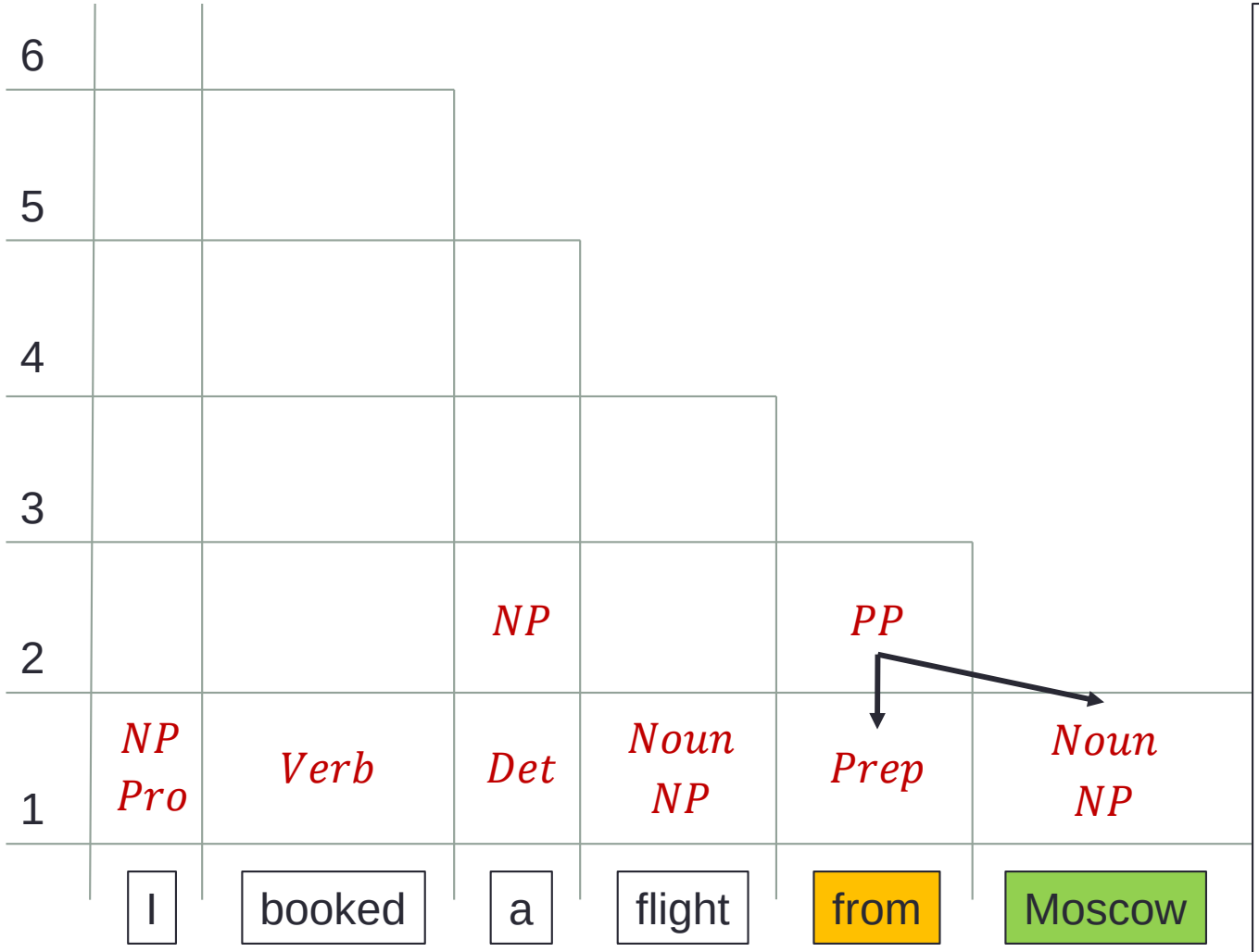
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $\underline{NP \rightarrow Det\ NP}$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



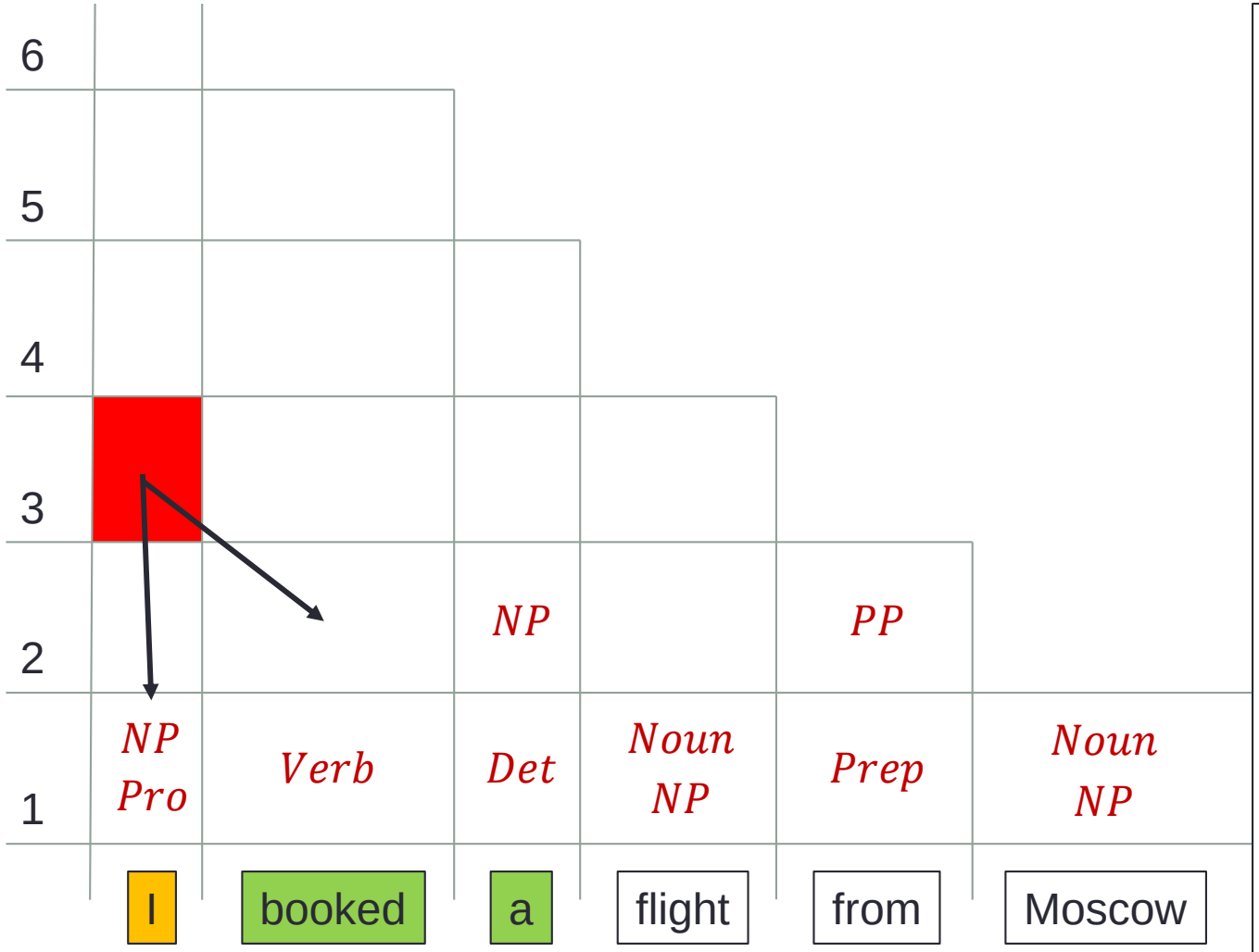
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



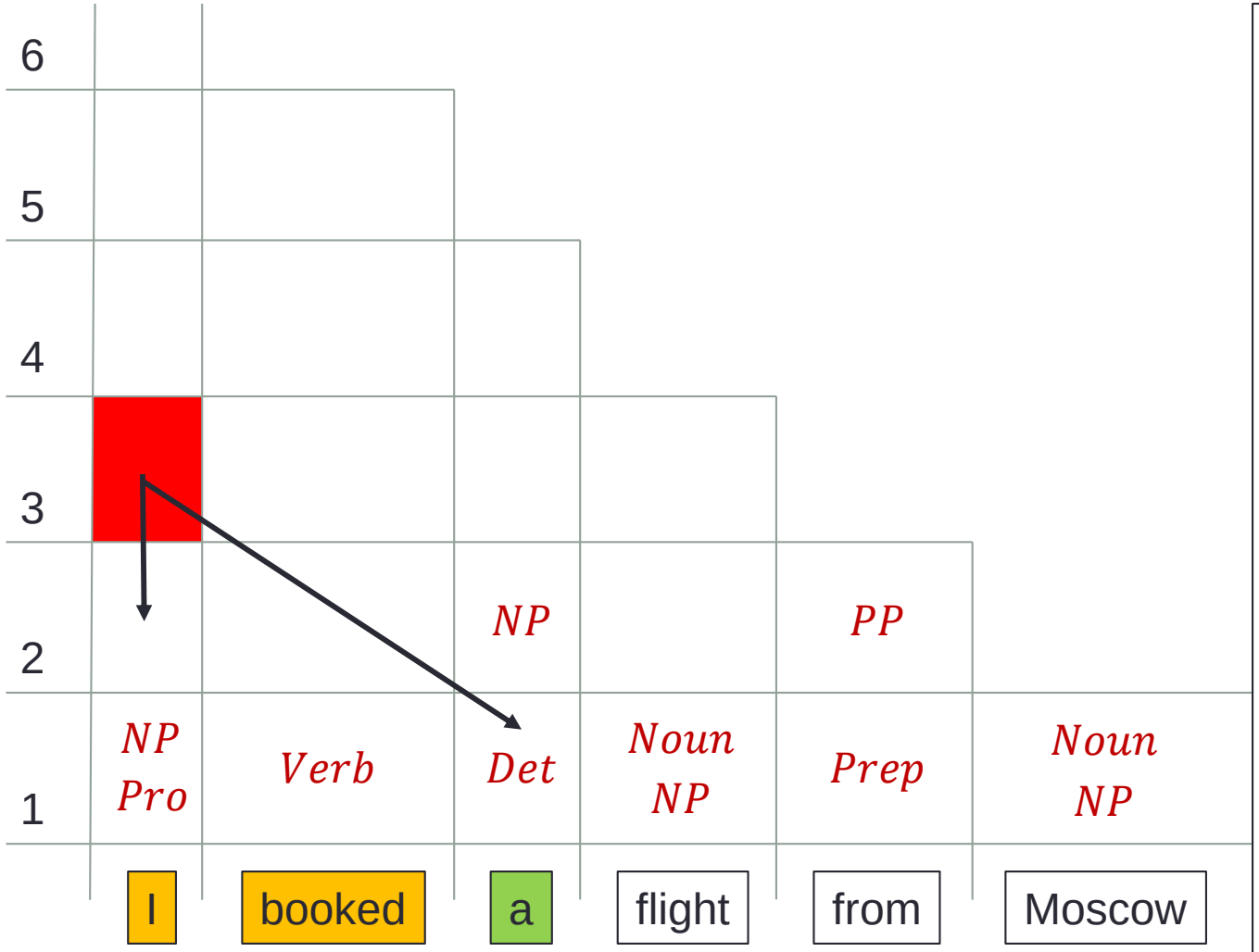
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



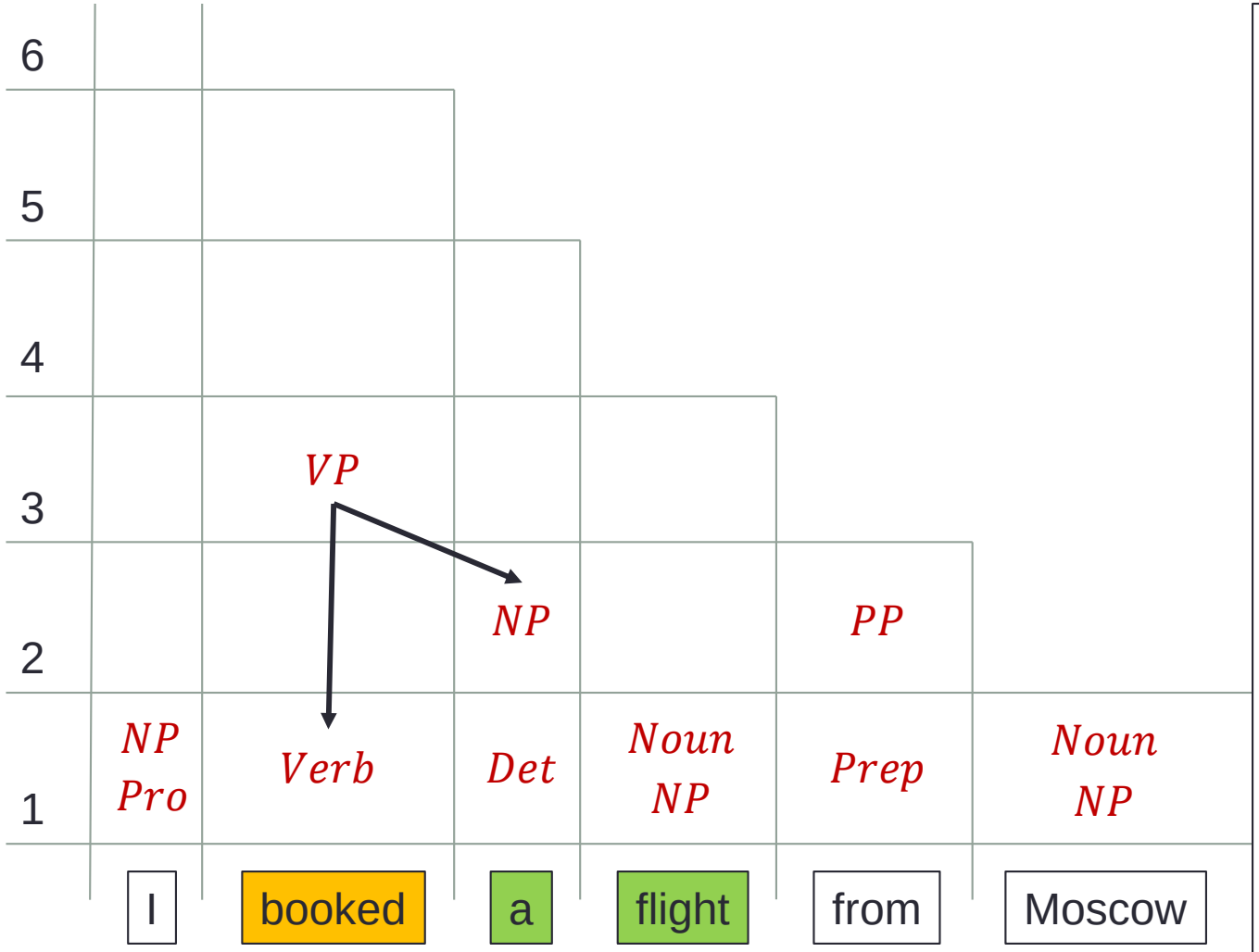
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



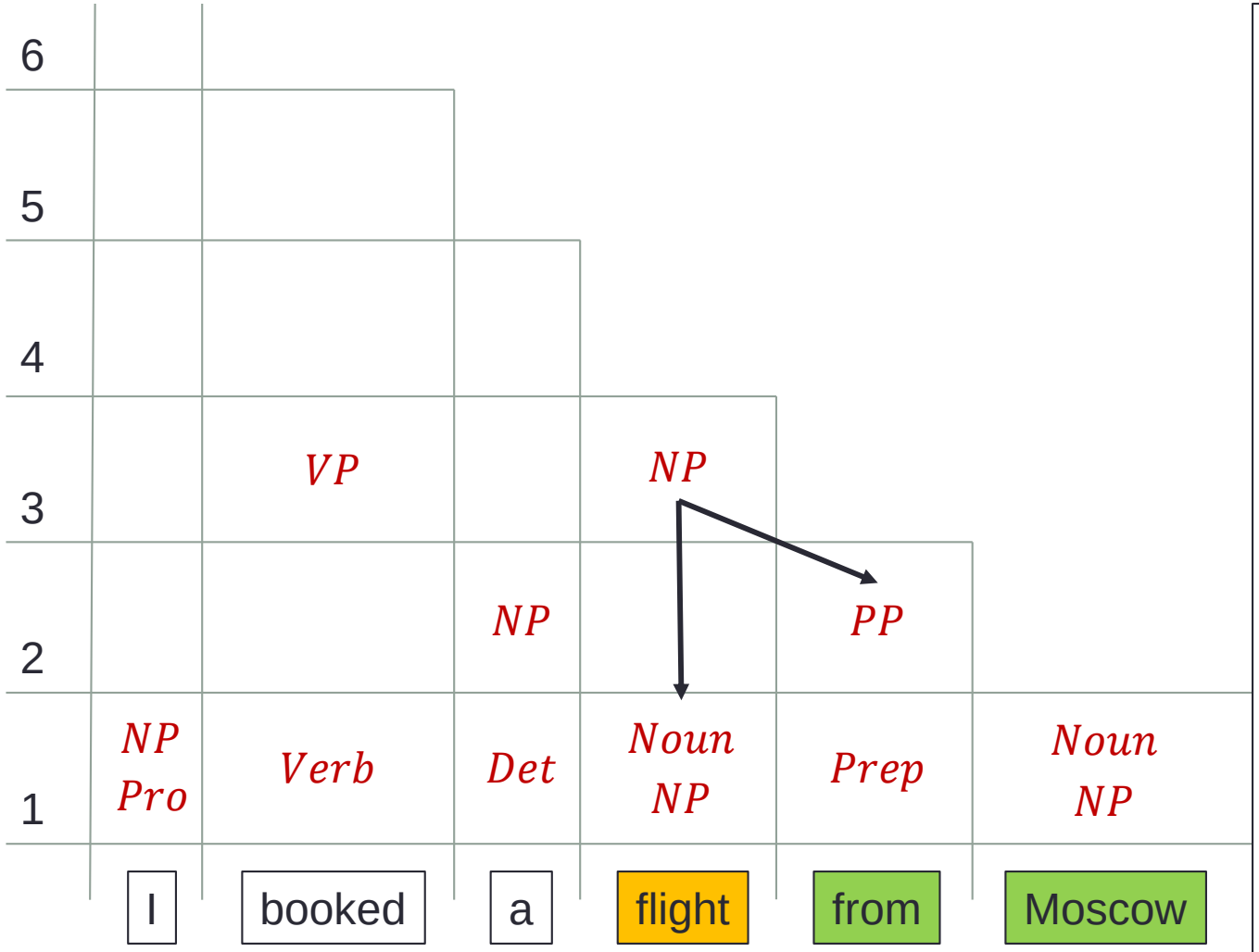
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



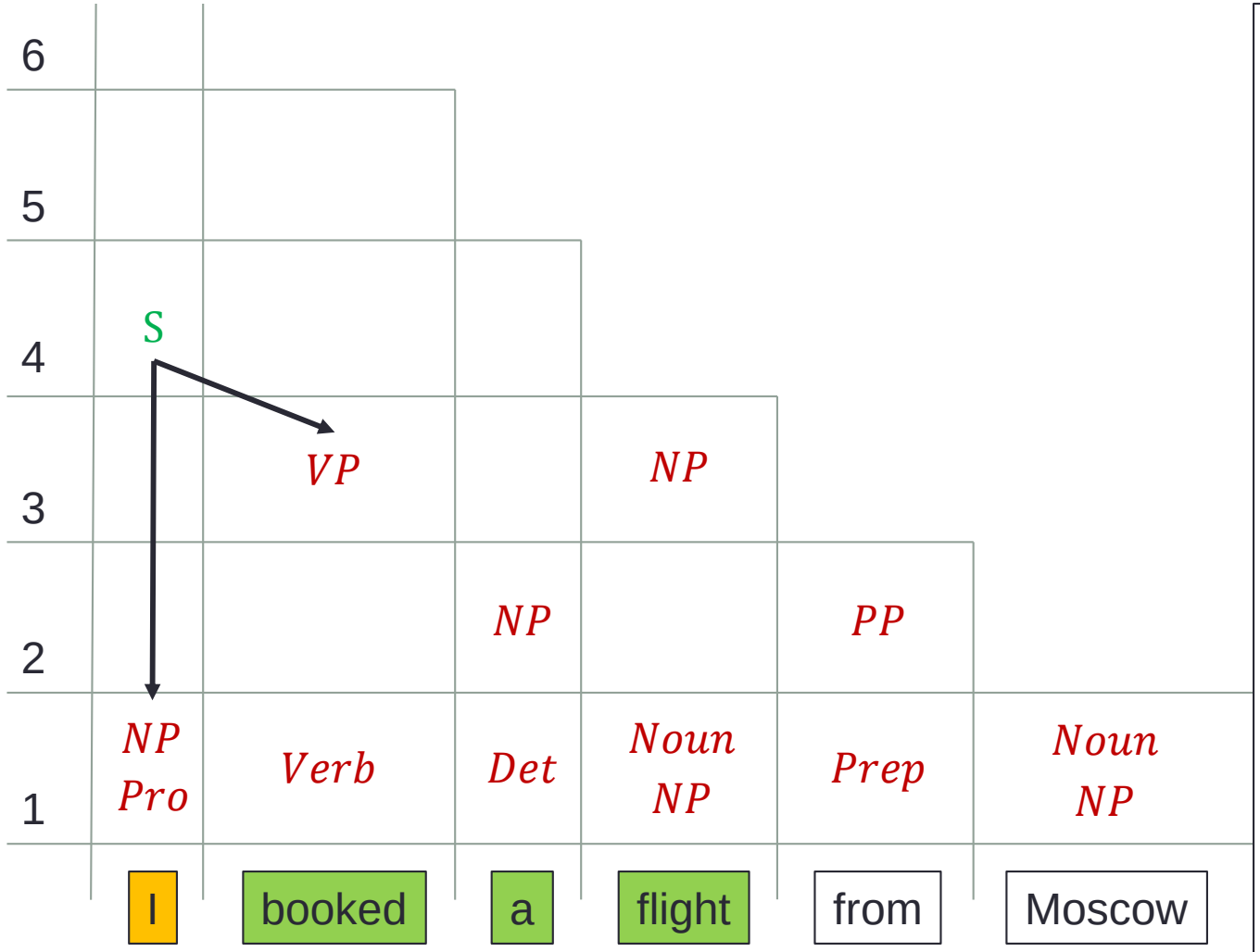
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



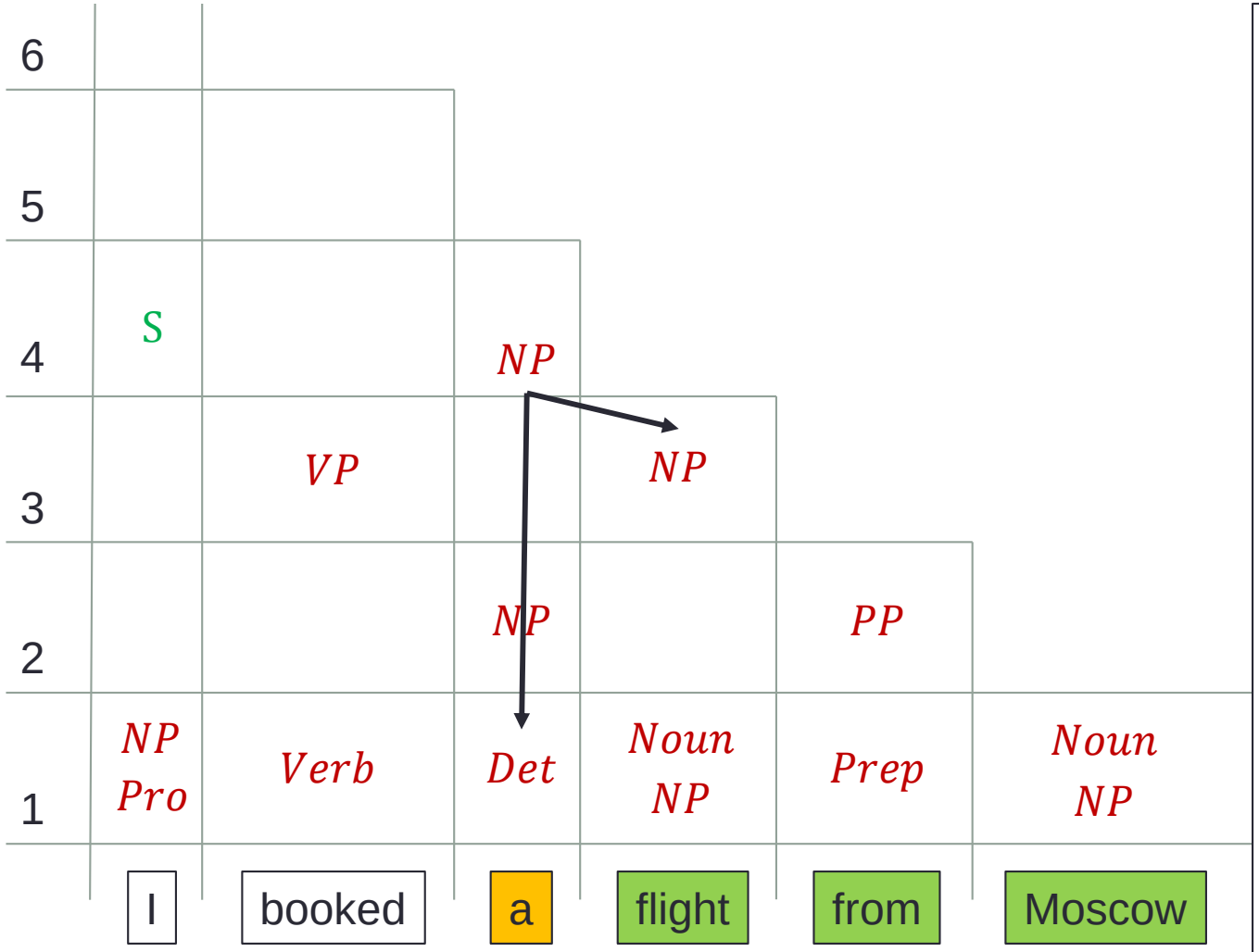
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $\underline{NP \rightarrow Noun\ PP}$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



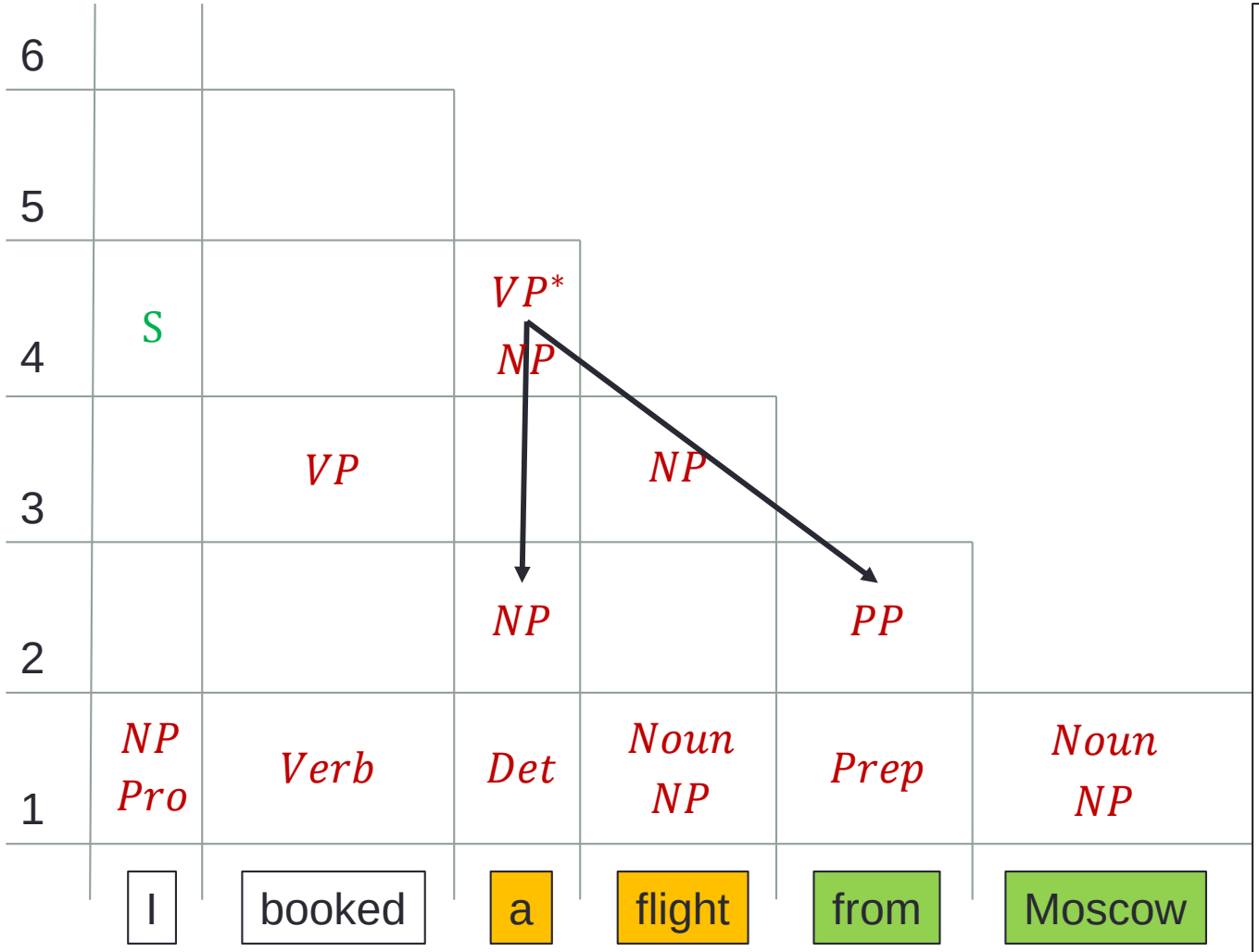
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



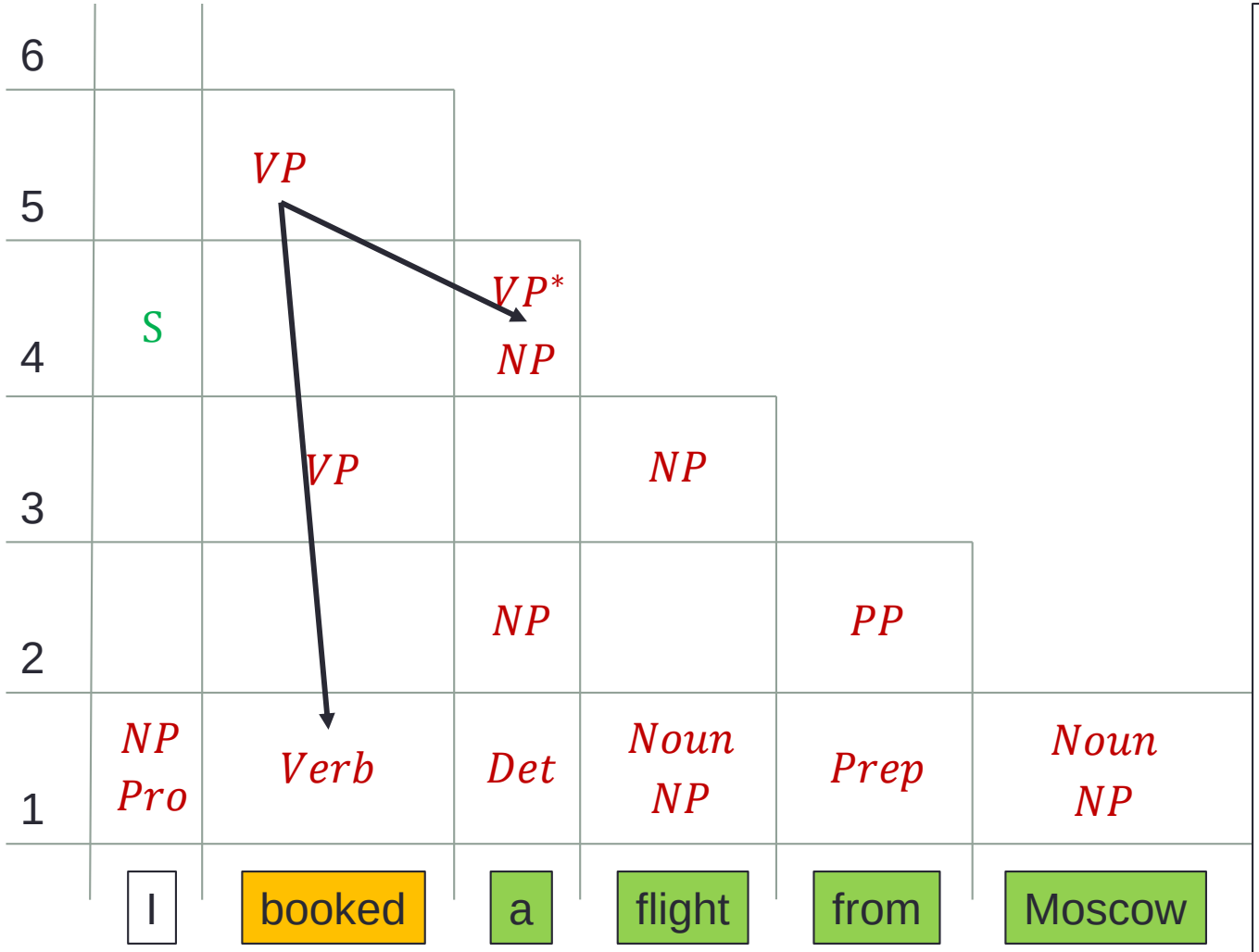
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



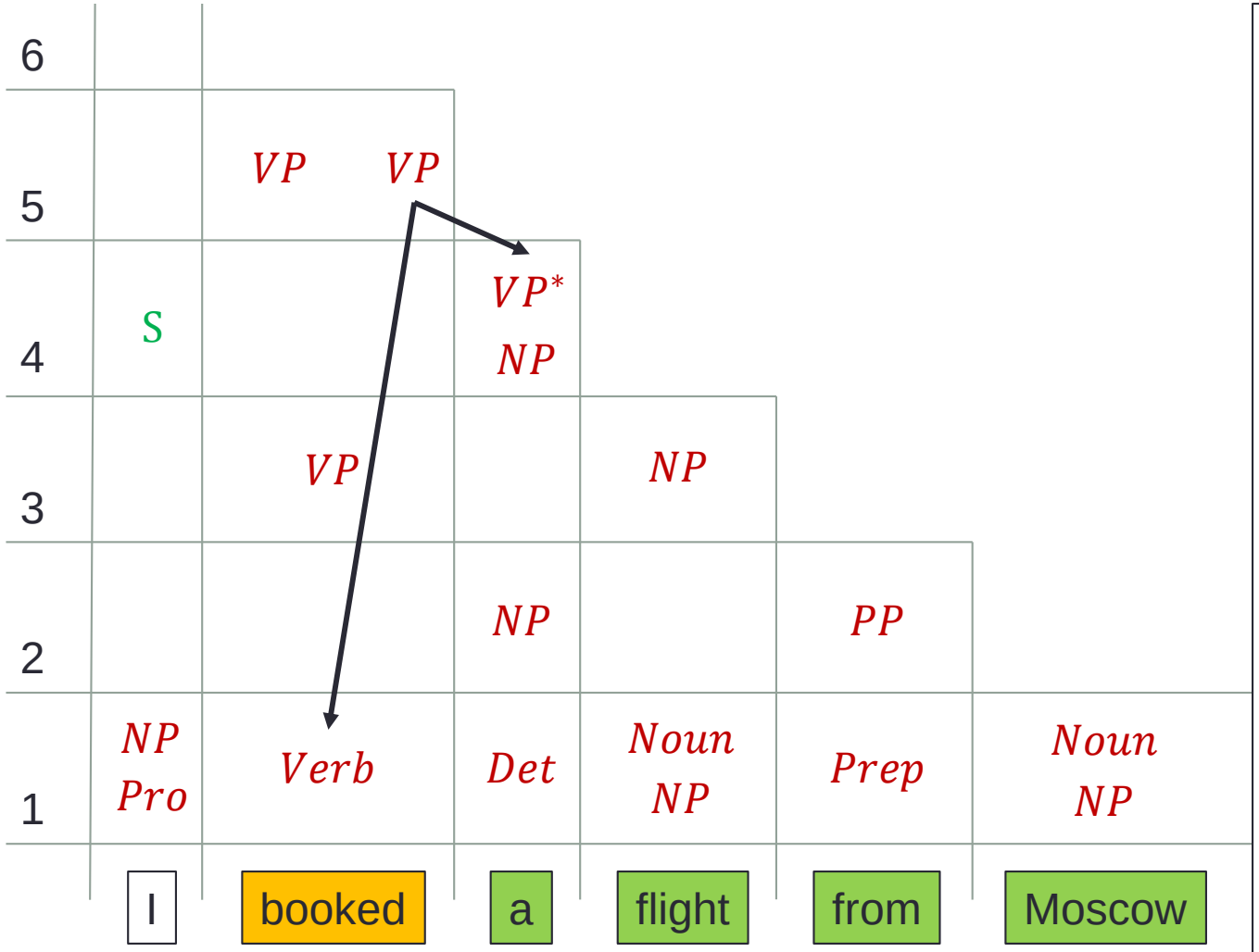
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $\underline{VP^* \rightarrow NP\ PP}$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



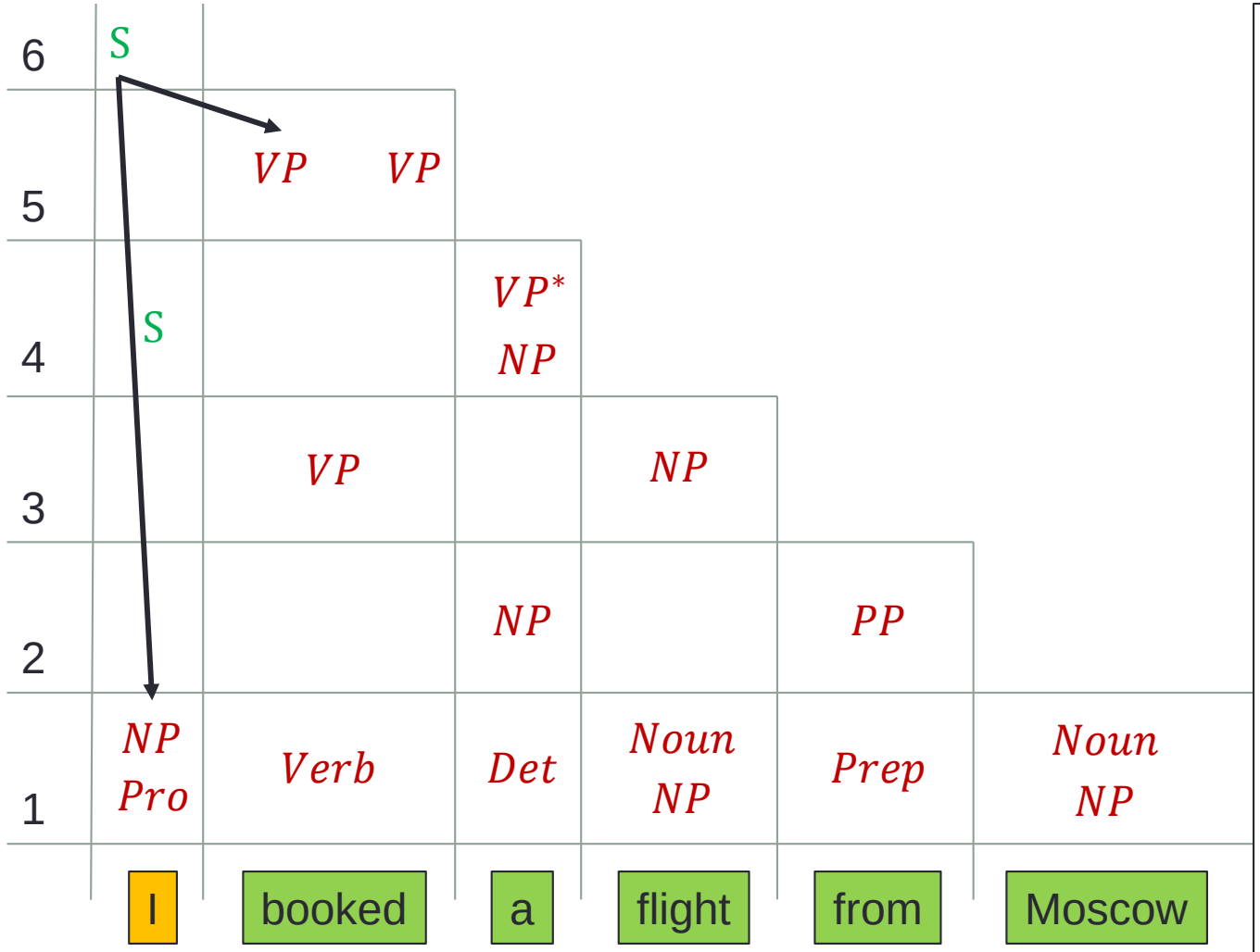
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



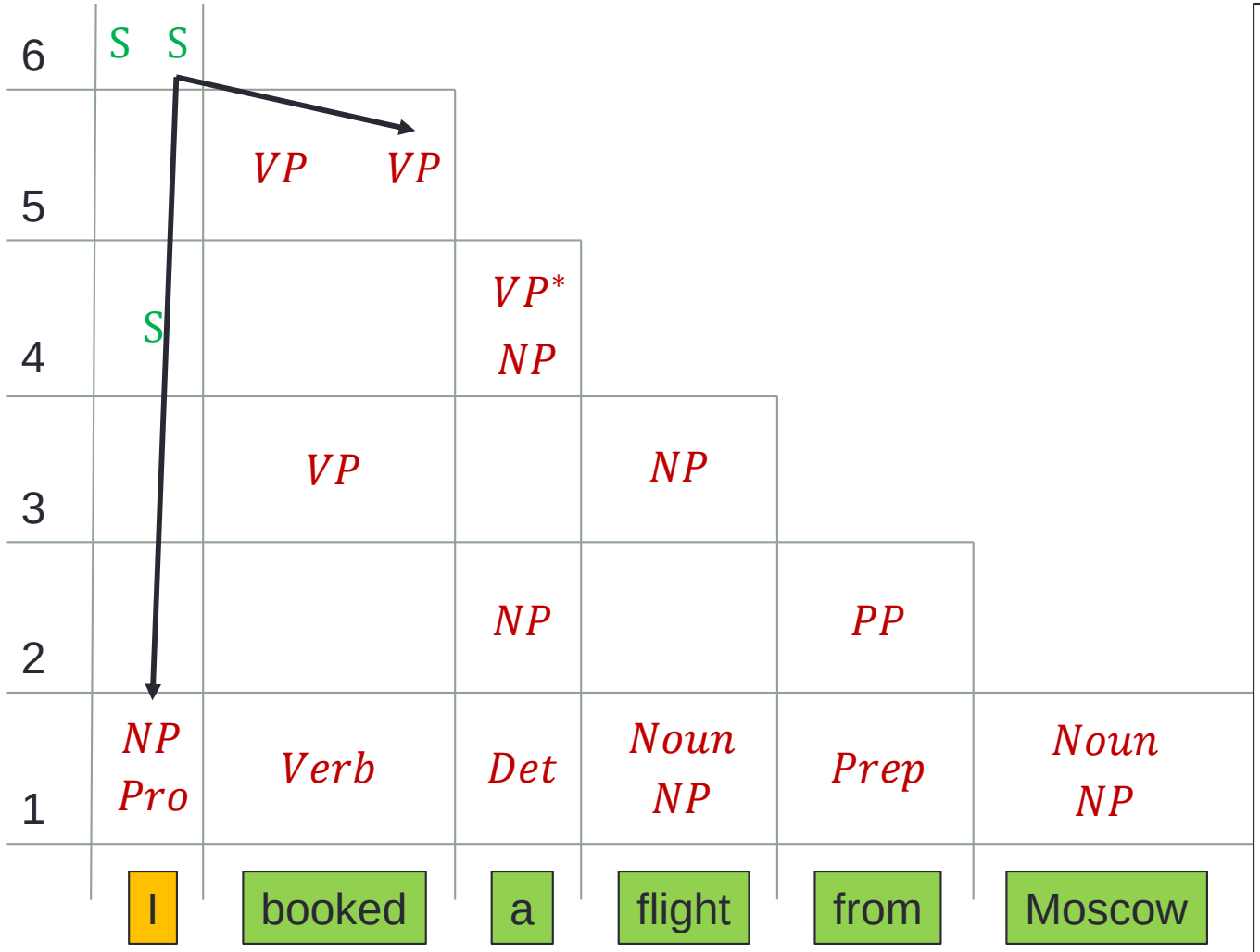
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



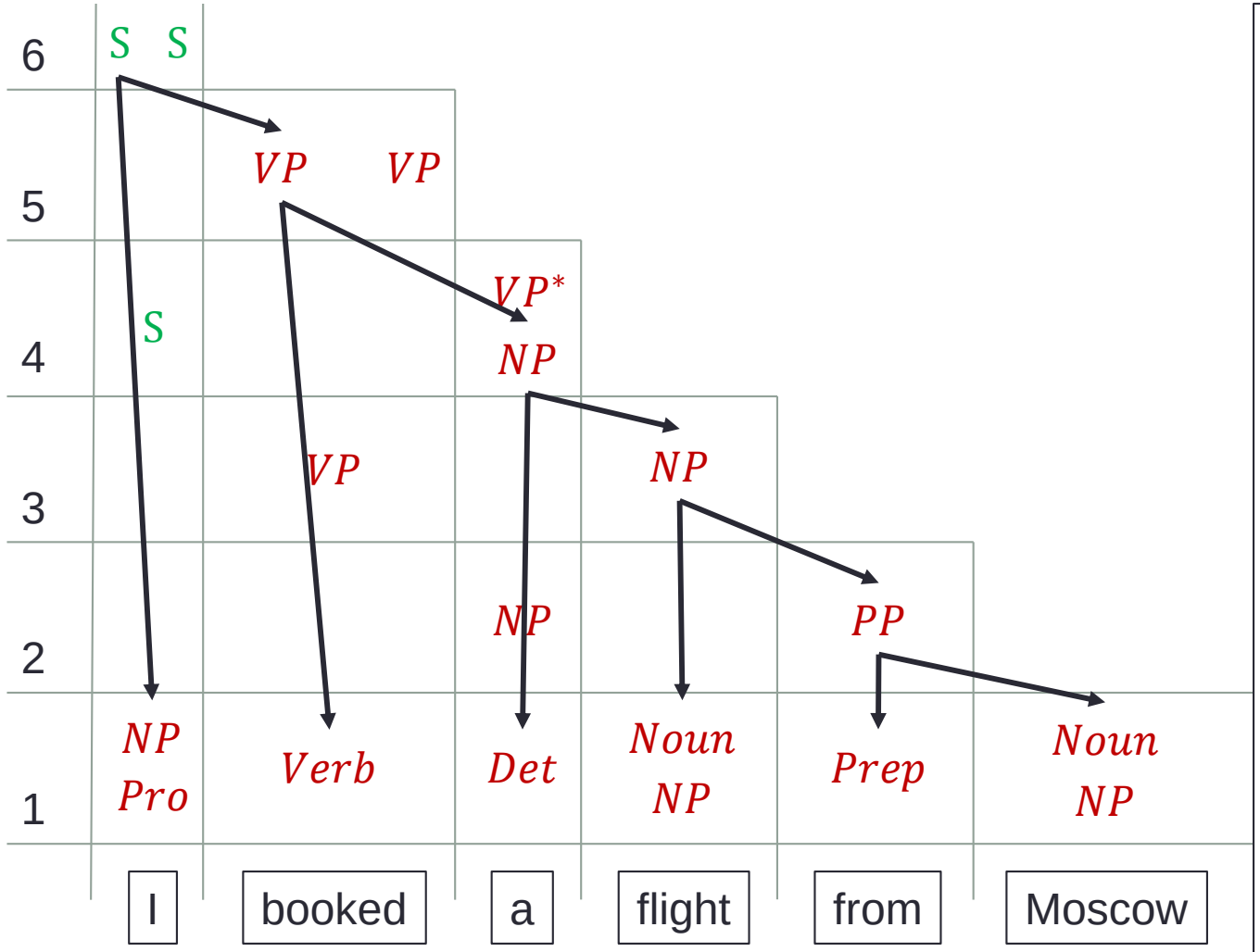
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



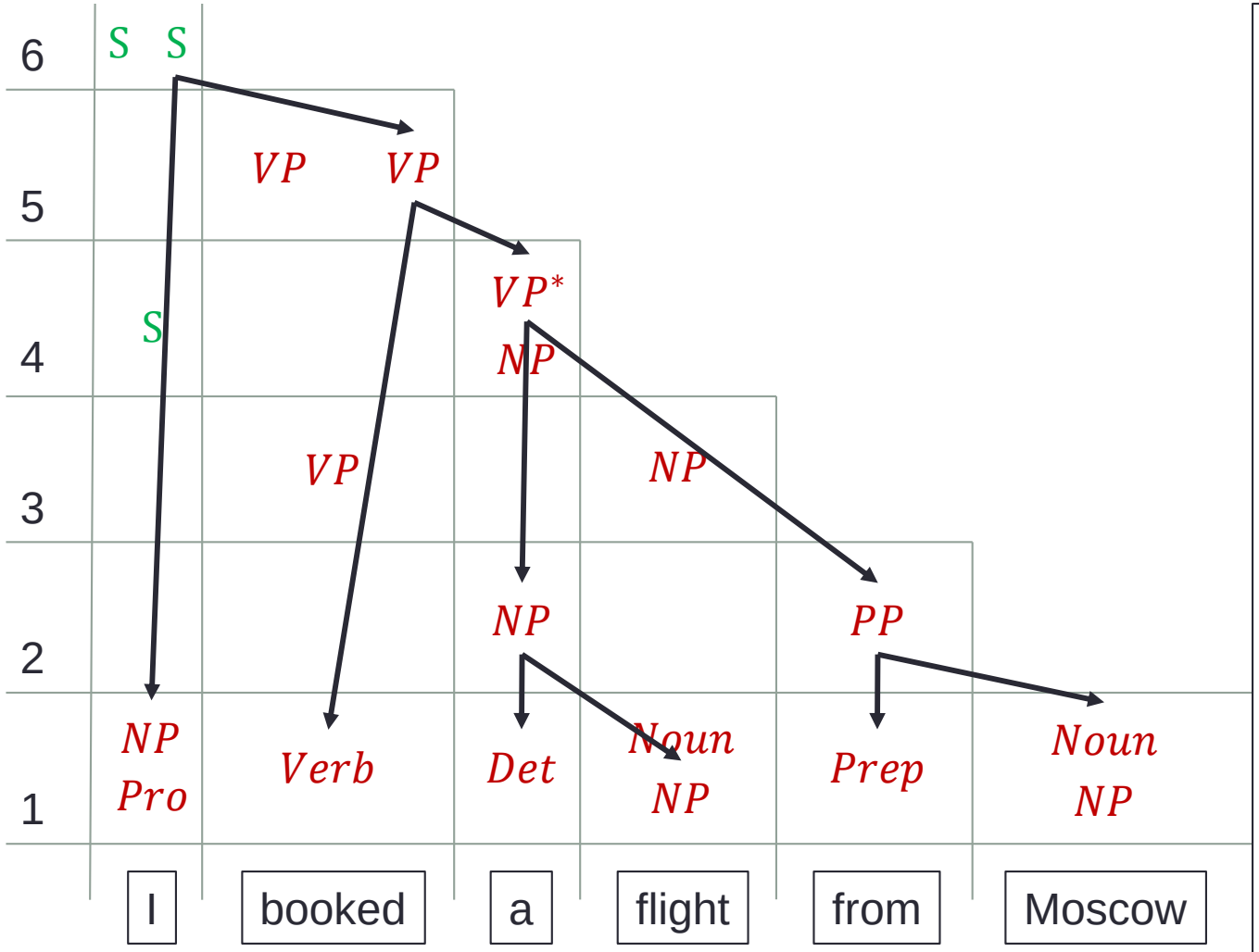
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Алгоритм CYK



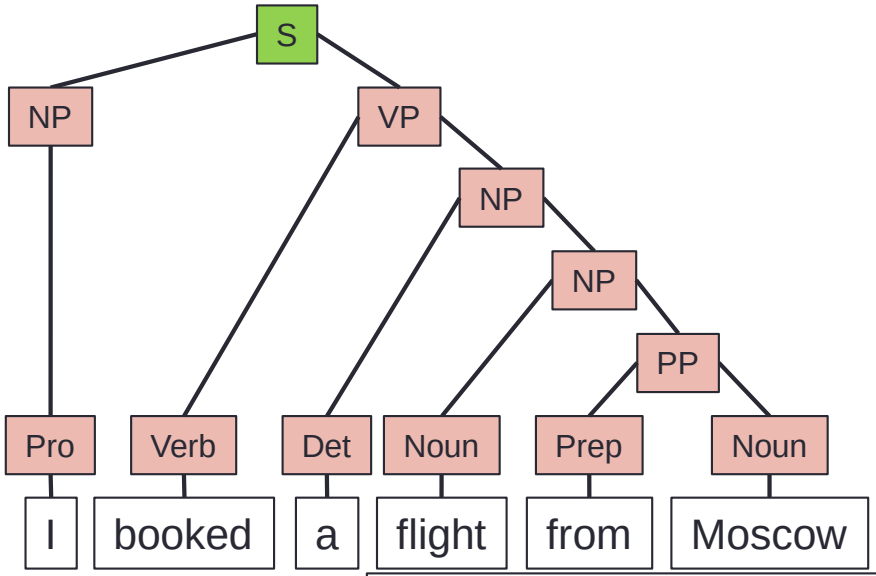
1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

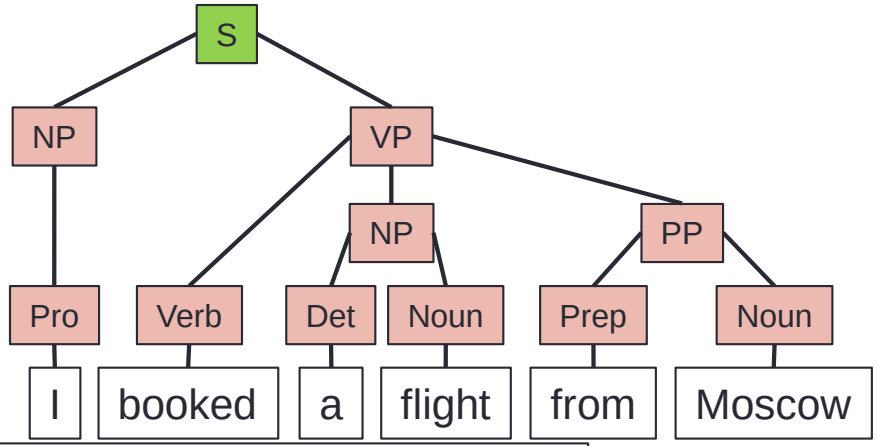
Алгоритм CYK



1. $S \rightarrow NP\ VP$
2. $VP \rightarrow Verb\ NP$
3. $VP \rightarrow Verb\ VP^*$
4. $VP^* \rightarrow NP\ PP$
5. $NP \rightarrow I$
6. $NP \rightarrow Det\ NP$
7. $NP \rightarrow Noun\ PP$
8. $NP \rightarrow flight$
9. $NP \rightarrow Moscow$
10. $PP \rightarrow Prep\ Noun$
11. $Verb \rightarrow booked$
12. $Noun \rightarrow flight$
13. $Noun \rightarrow Moscow$
14. $Pro \rightarrow I$
15. $Det \rightarrow a$
16. $Det \rightarrow the$
17. $Prep \rightarrow from$

Неоднозначность языка





1. $S \rightarrow NP\ VP$
2. $\underline{VP \rightarrow Verb\ NP}$
3. $\underline{VP \rightarrow Verb\ NP\ PP}$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Noun \rightarrow flight$
10. $Verb \rightarrow booked$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Выбор лучшего дерева

- Зададим каждому правилу вероятность его применения
- Будем оценивать вероятность разбора, как произведение вероятностей правил, участвующих в нем

$$p(tree) = \prod_i p(\alpha_i \to \beta_i)$$

$$p(\alpha \to \beta) = \frac{Count(\alpha \to \beta)}{Count(\alpha)}$$

$$t^*(s) = \arg\max_{t \in \mathrm{T}(s)} p(t)$$

Выбор лучшего дерева

- Зададим каждому правилу вероятность его применения
- Будем оценивать вероятность разбора, как произведение вероятностей правил, участвующих в нем

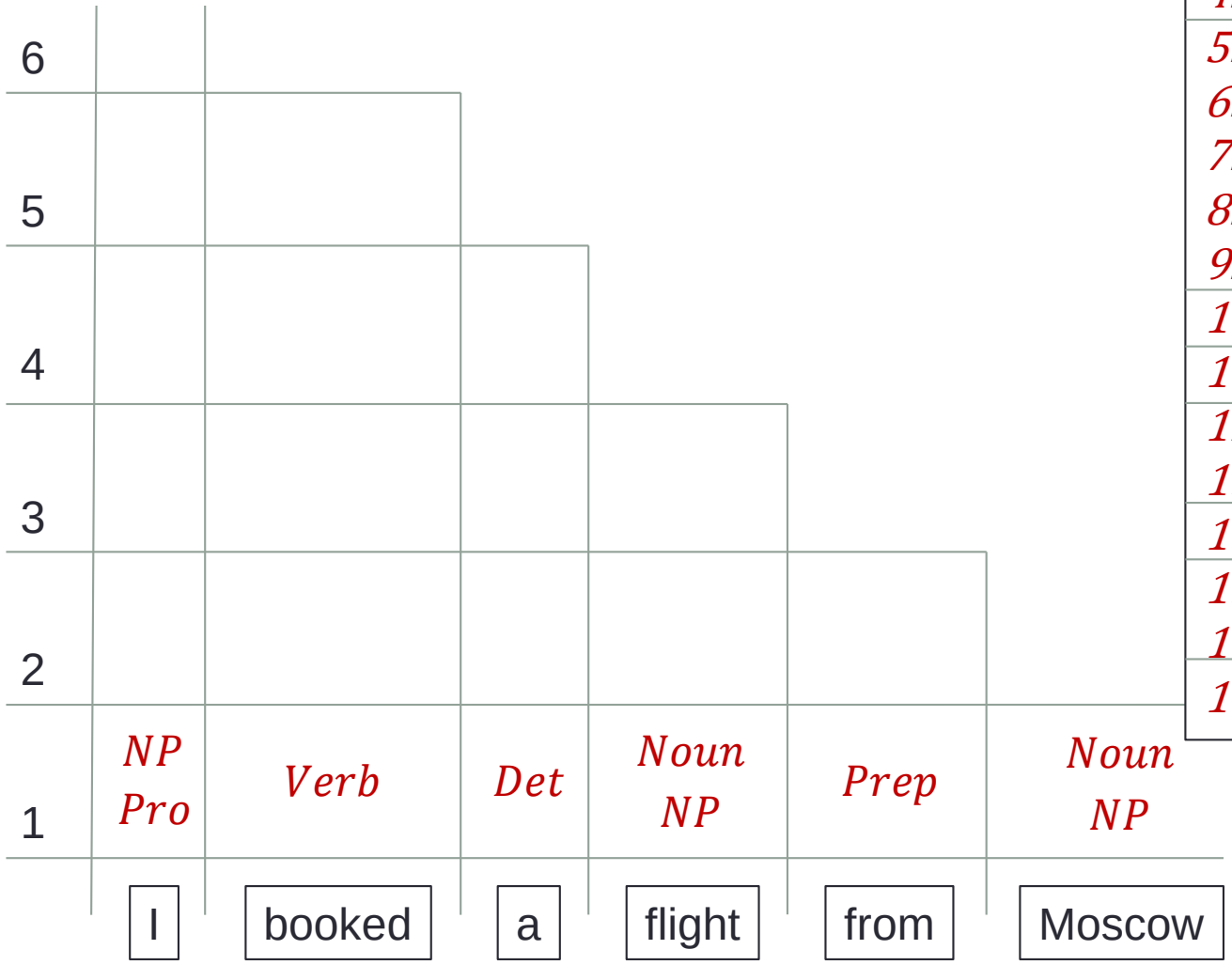
$$p(tree) = \prod_i p(\alpha_i \rightarrow \beta_i)$$

$$p(\alpha \rightarrow \beta) = \frac{Count(\alpha \rightarrow \beta)}{Count(\alpha)}$$

$$t^*(s) = \arg\max_{t \in \mathrm{T}(s)} p(t)$$

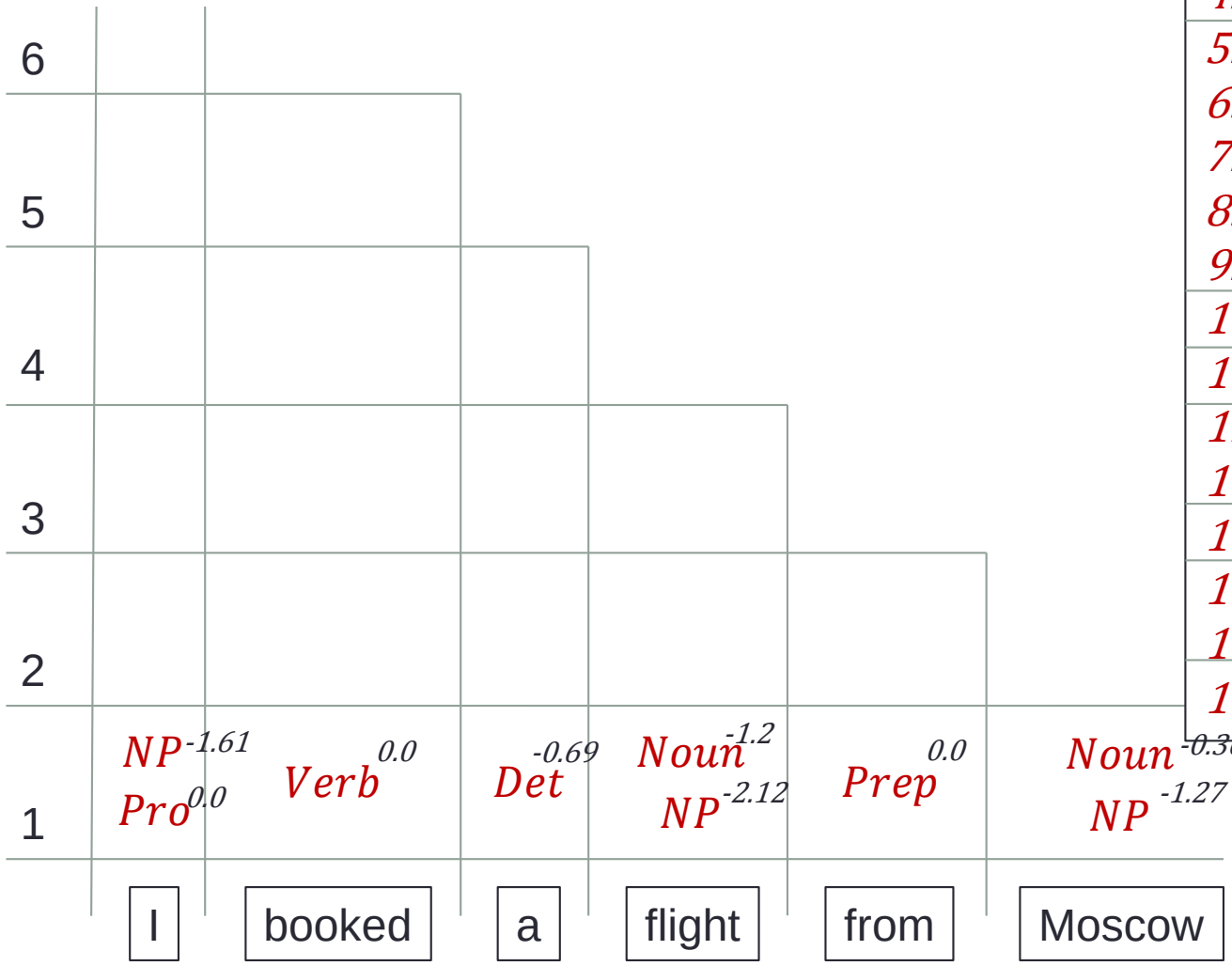
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ NP\ PP$	0.1
4.	$NP \rightarrow Pro$	0.2
5.	$NP \rightarrow Det\ NP$	0.3
6.	$NP \rightarrow Noun\ PP$	0.1
7.	$NP \rightarrow Noun$	0.4
8.	$PP \rightarrow Prep\ Noun$	1.0
9.	$Verb \rightarrow booked$	1.0
10.	$Noun \rightarrow flight$	0.3
11.	$Noun \rightarrow Moscow$	0.7

Алгоритм CYK



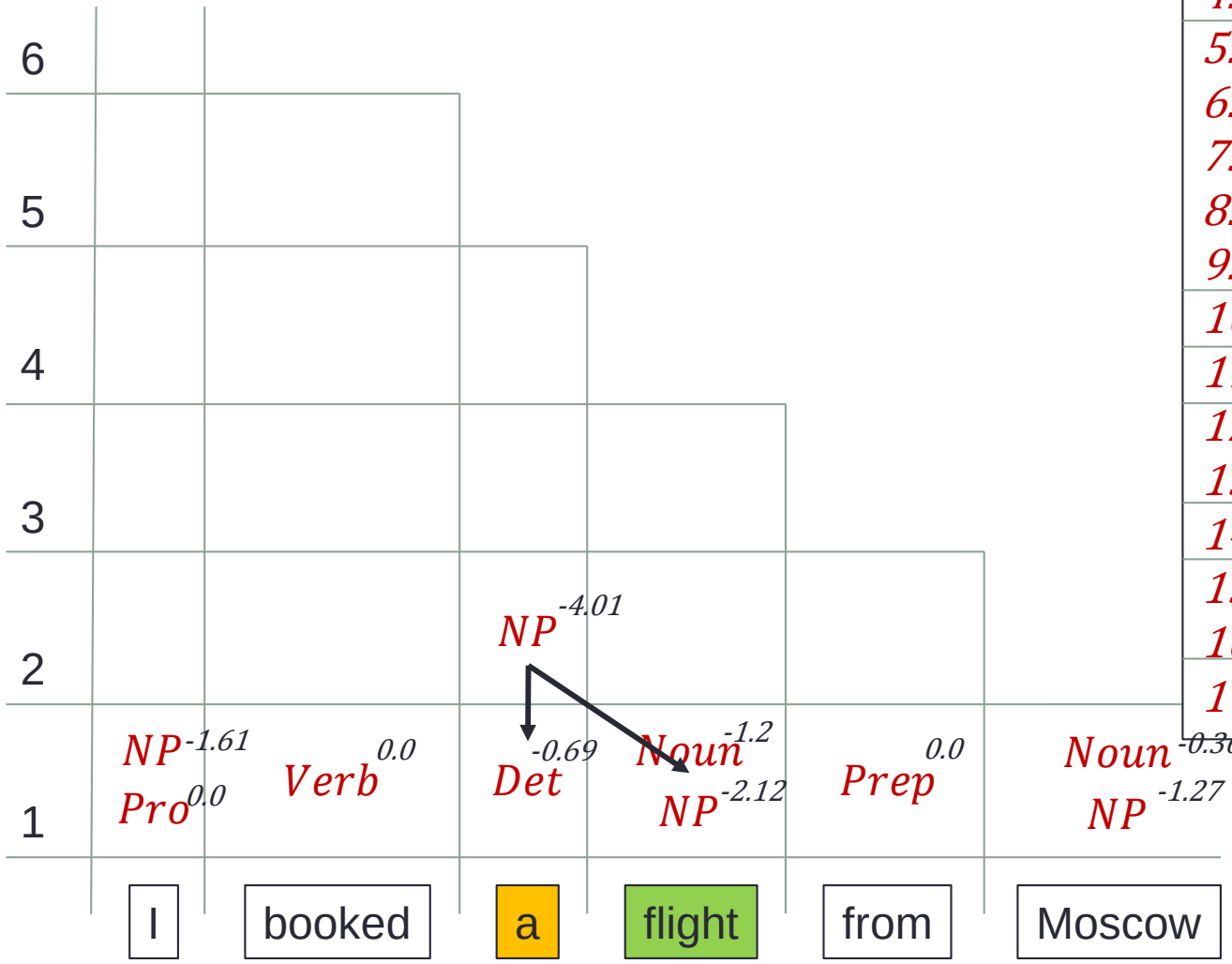
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



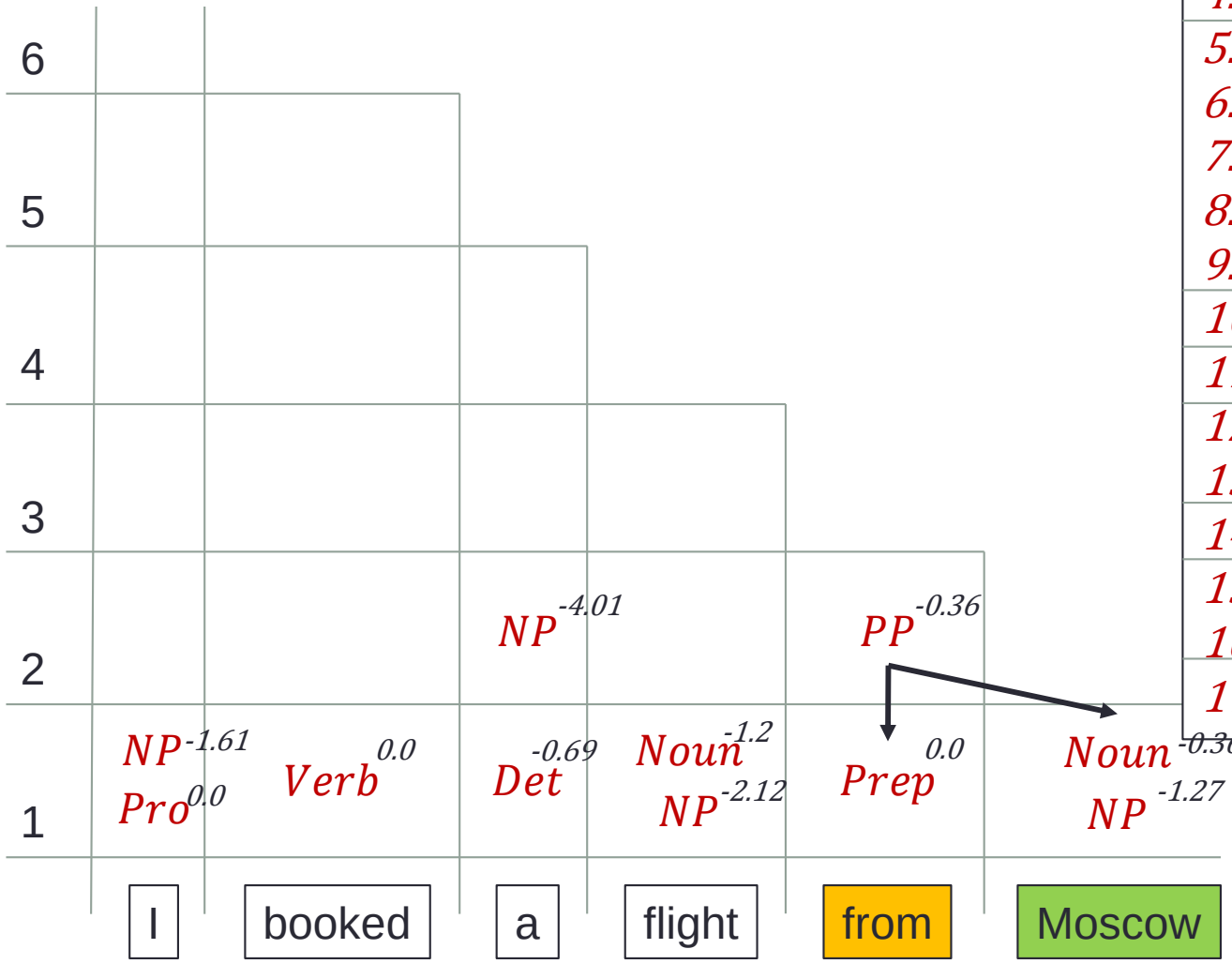
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



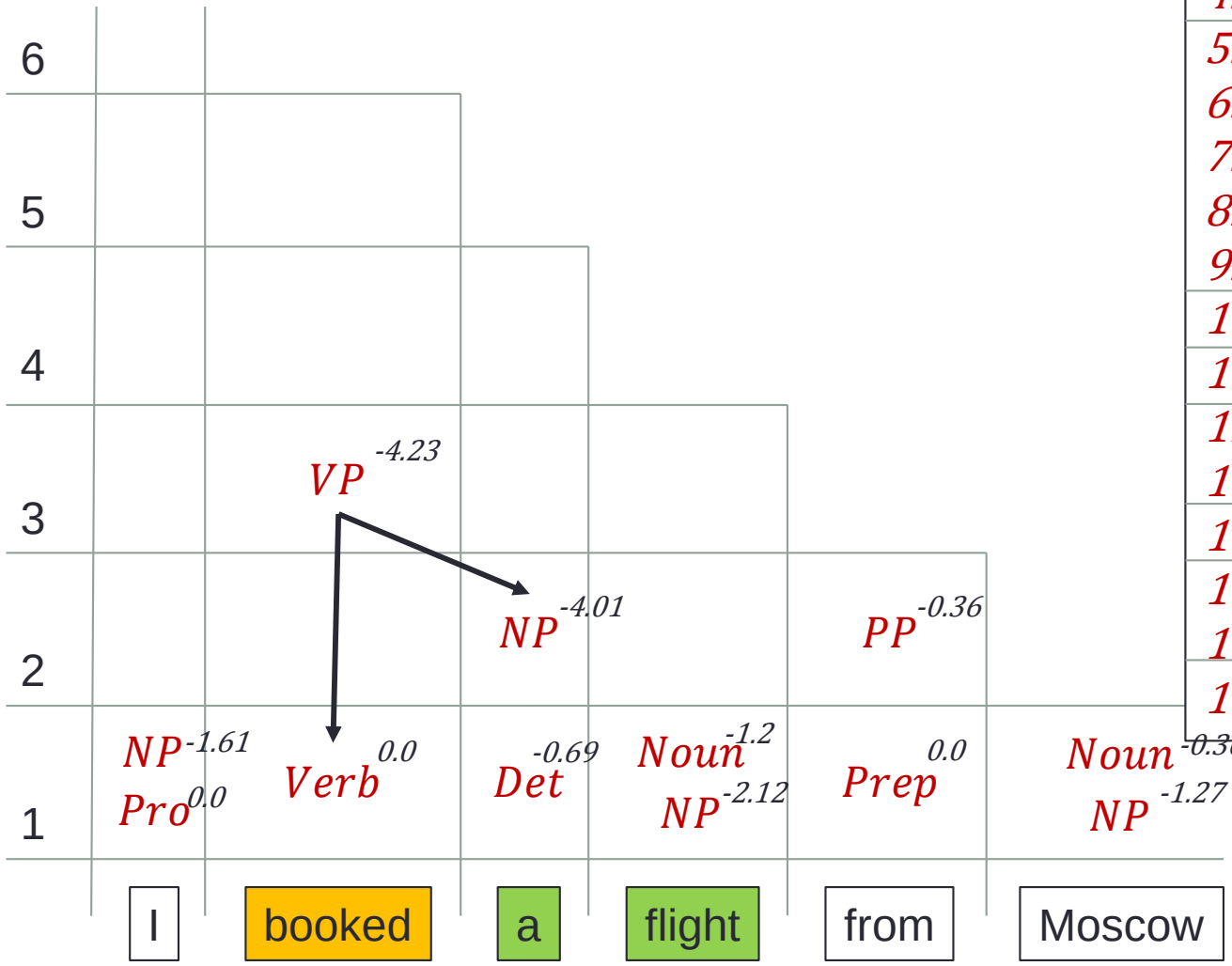
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



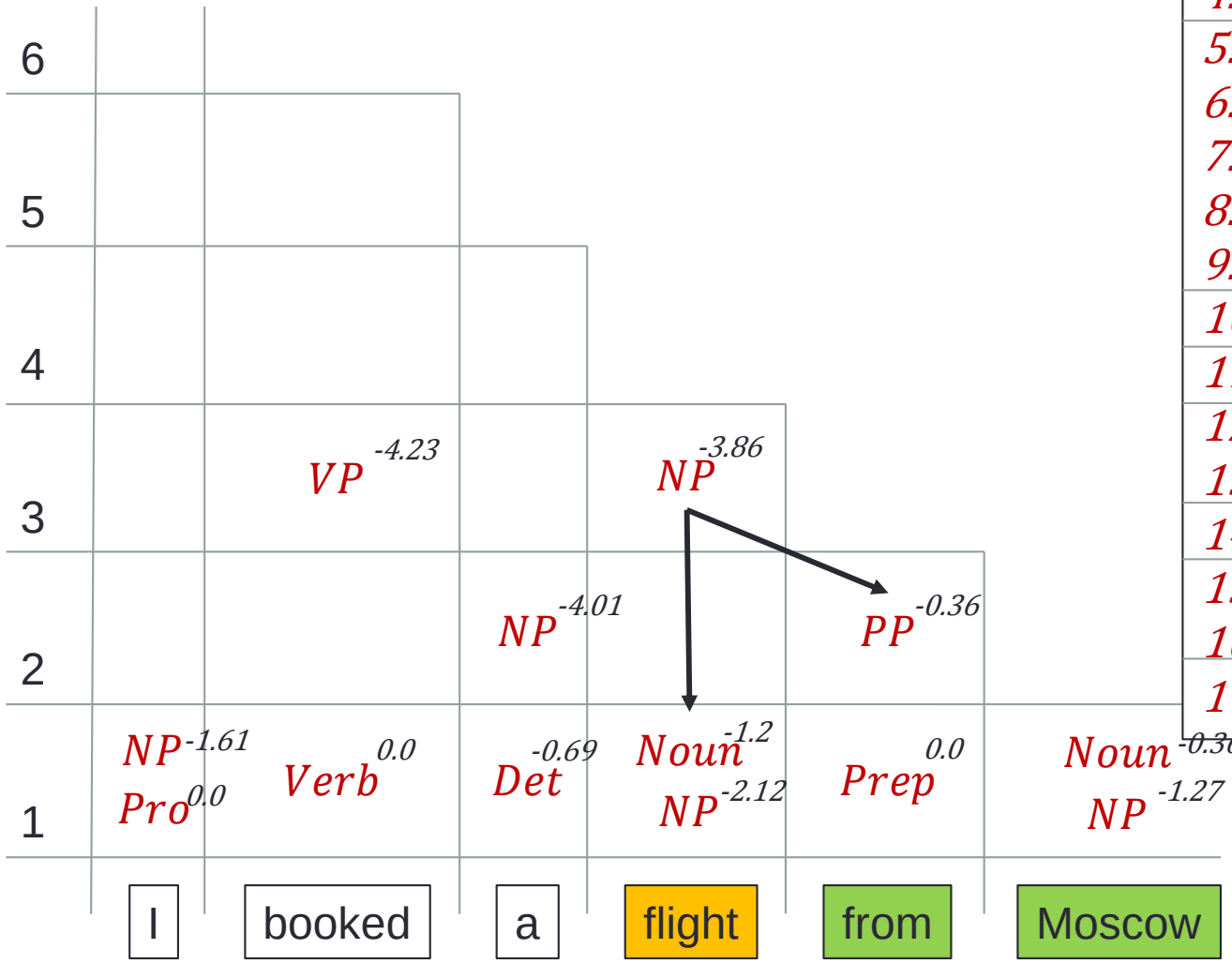
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



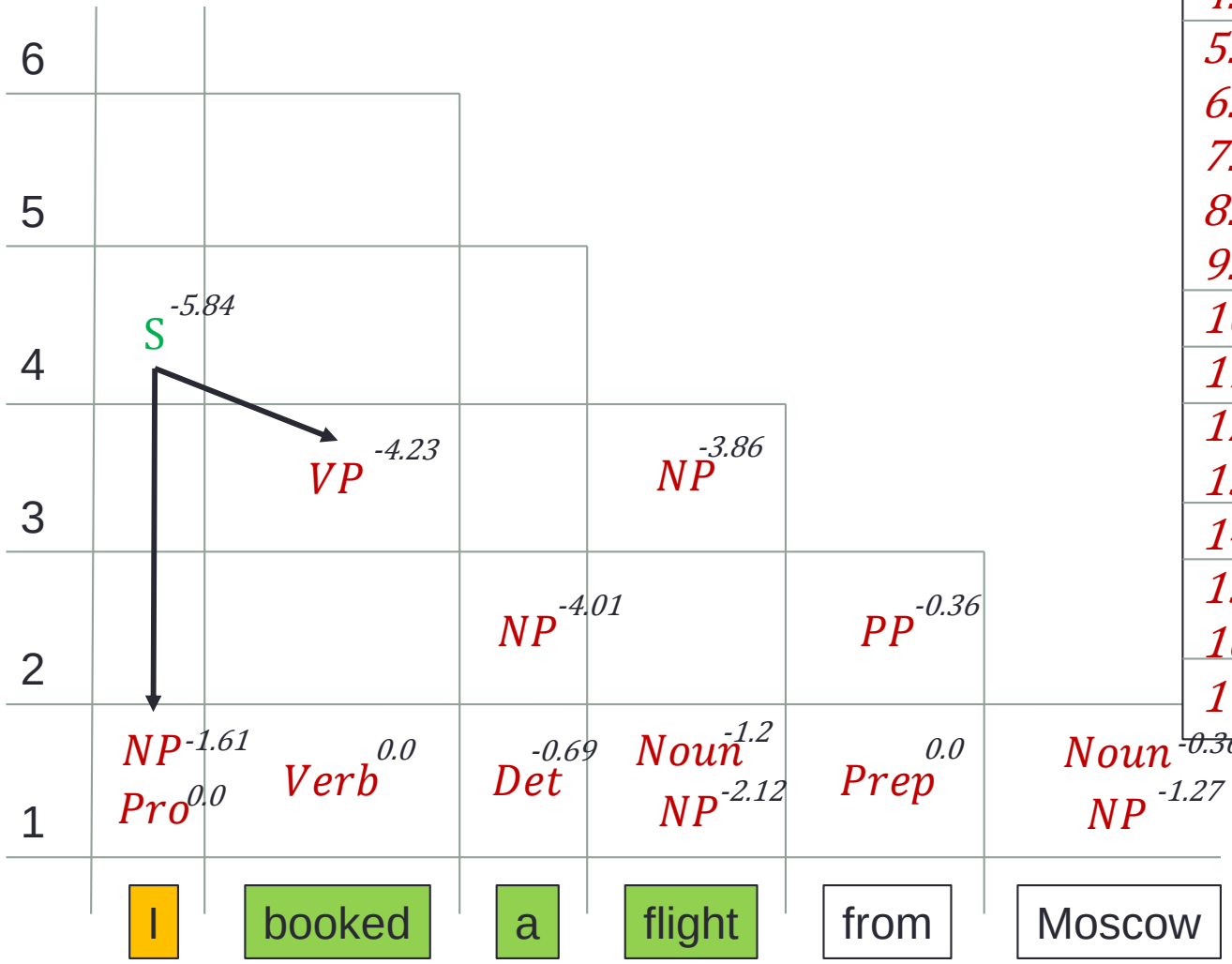
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



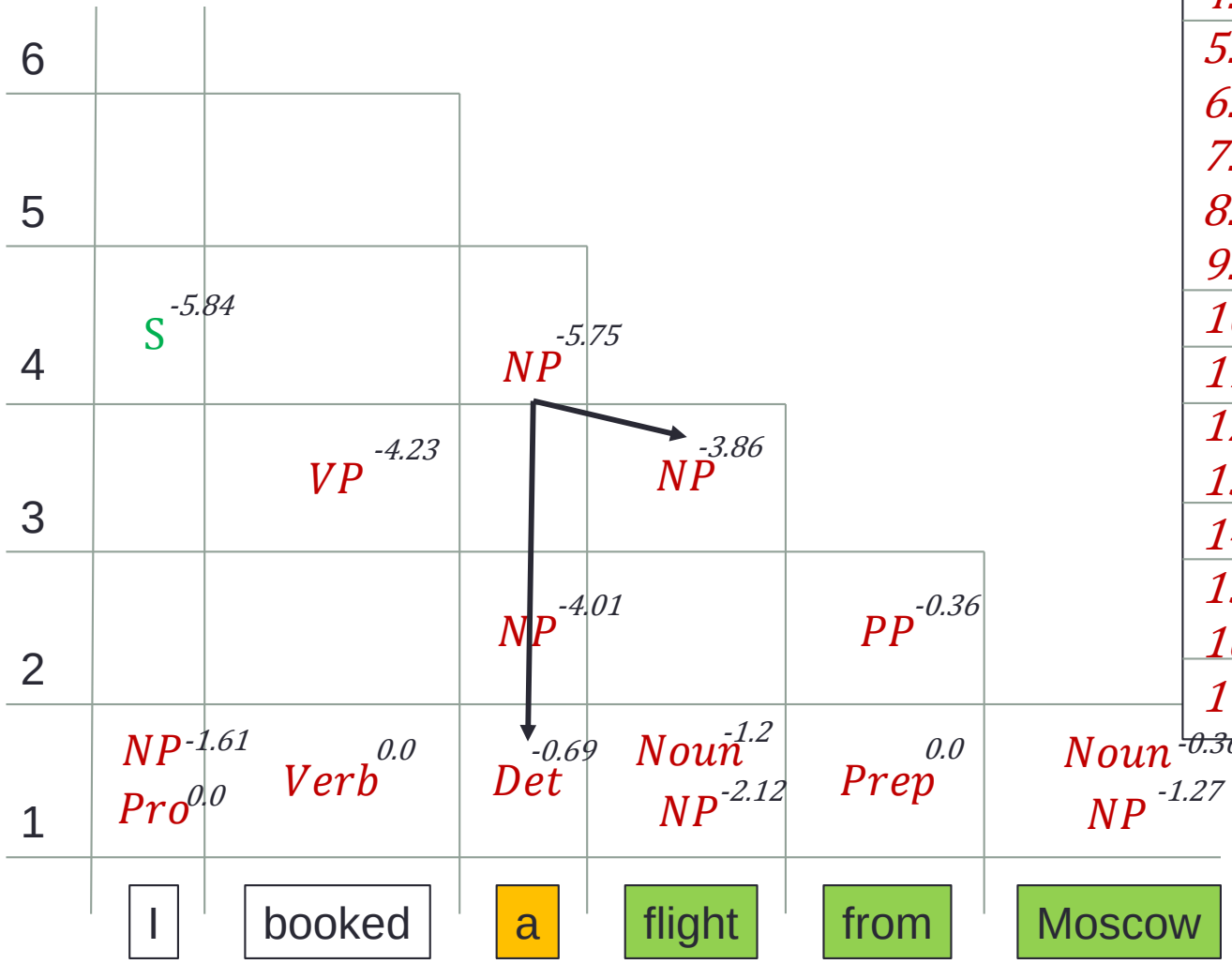
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



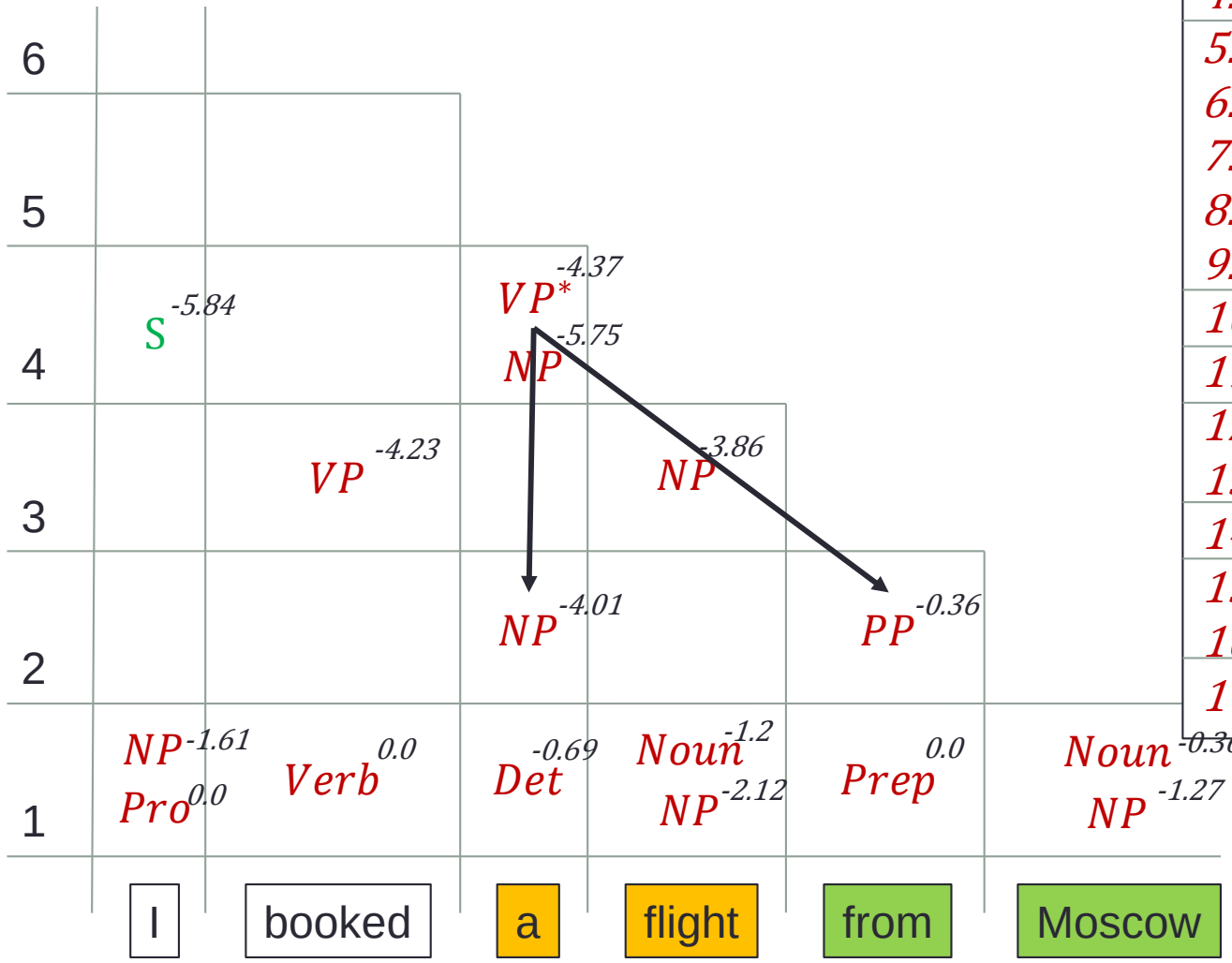
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



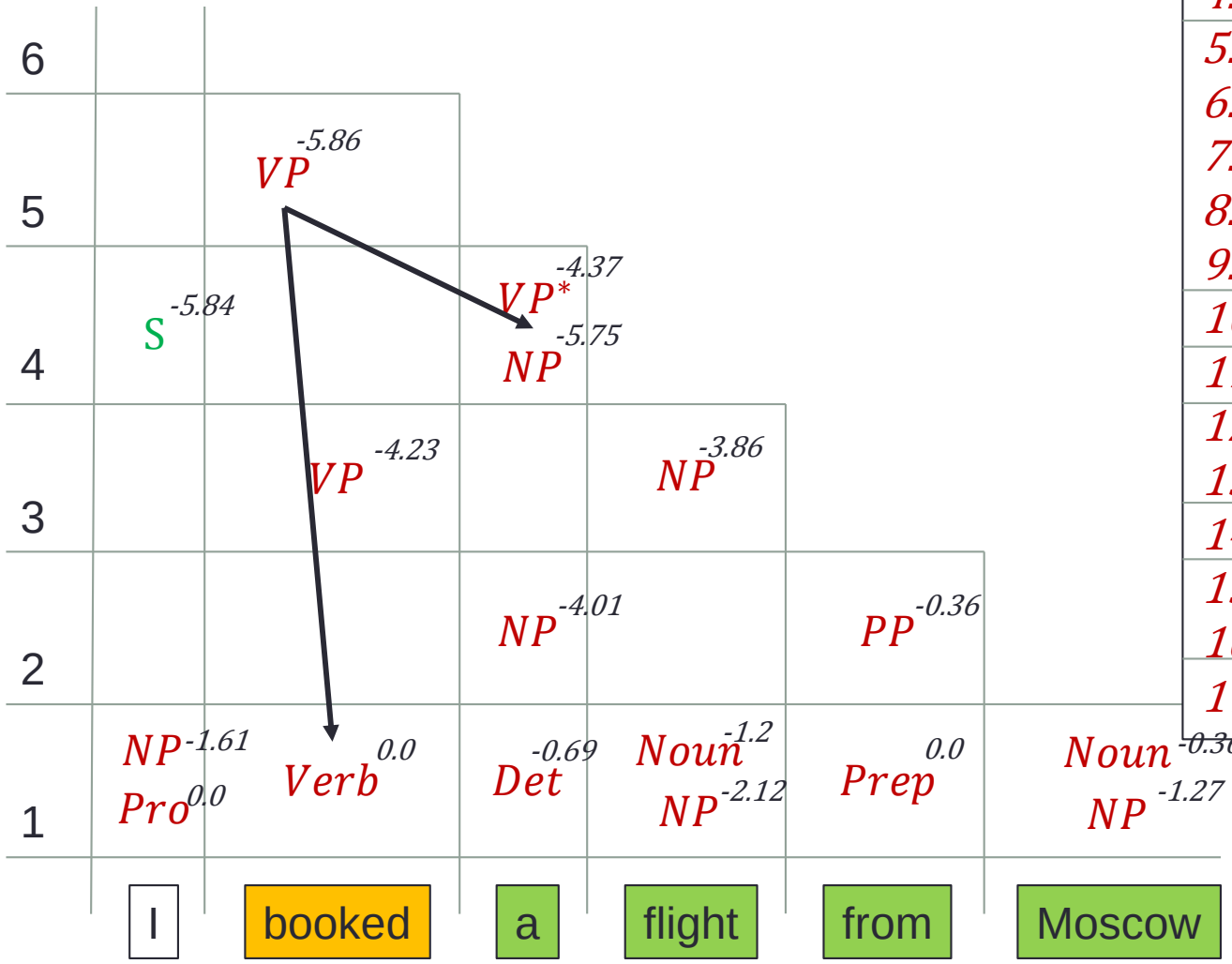
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



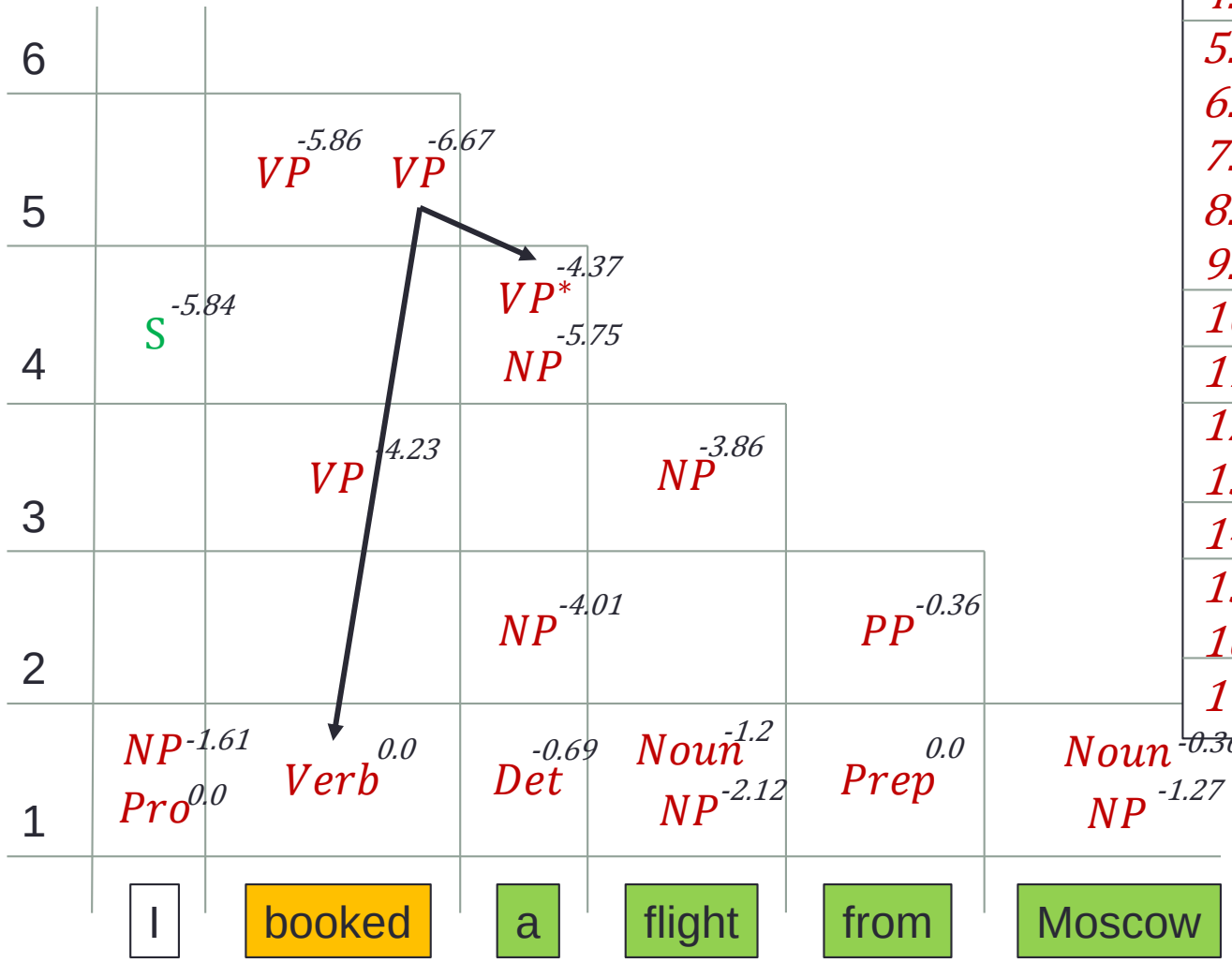
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



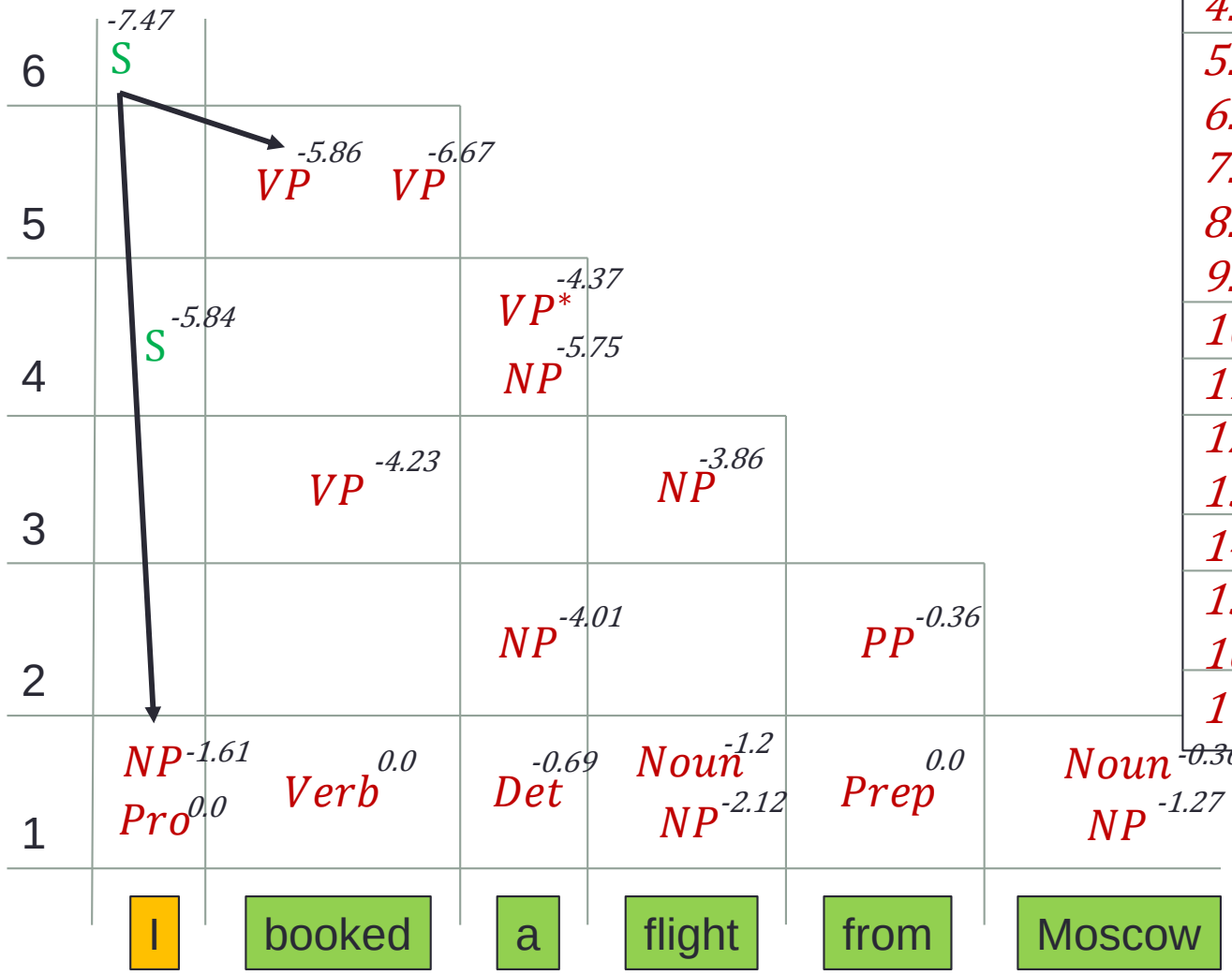
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



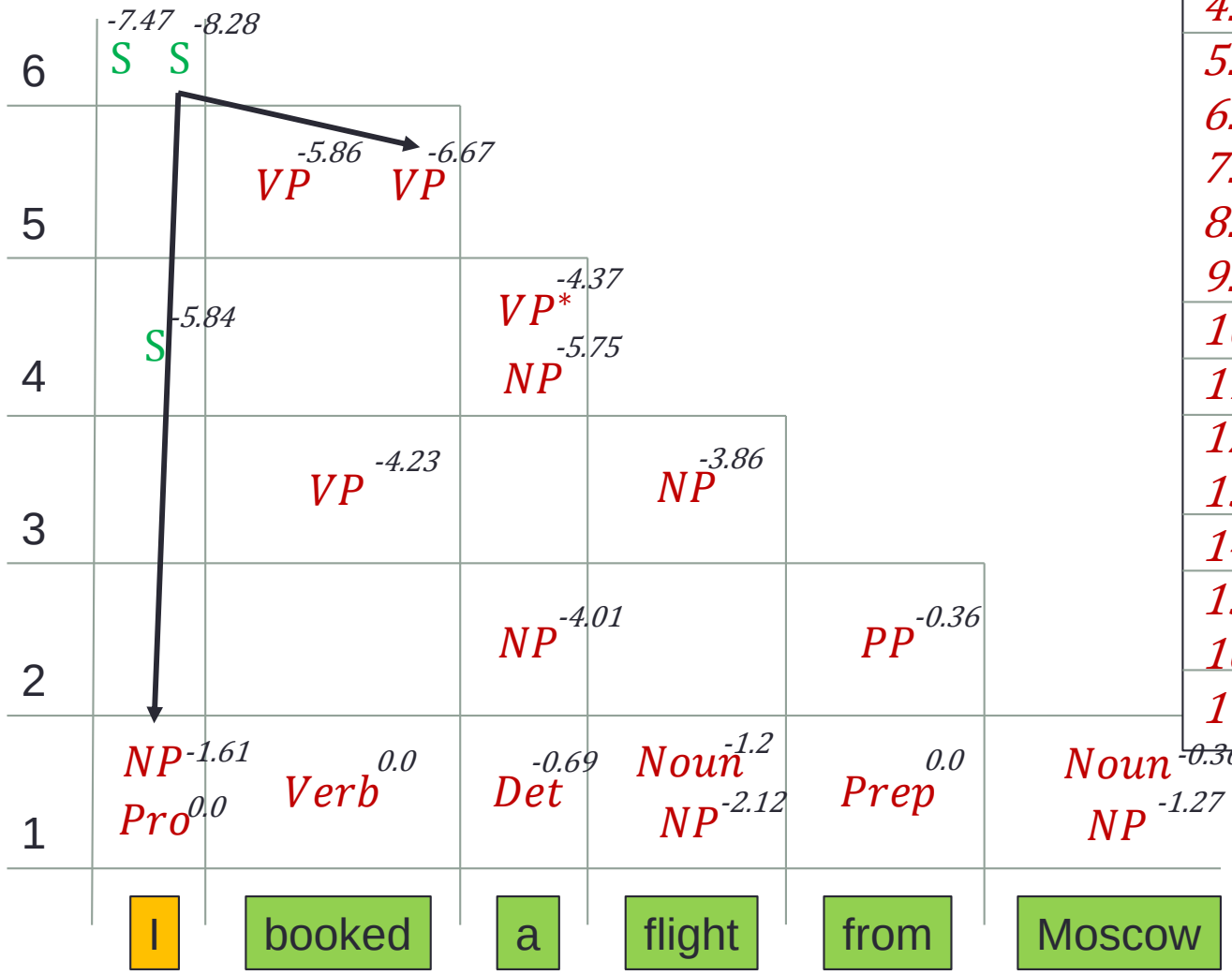
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Алгоритм CYK



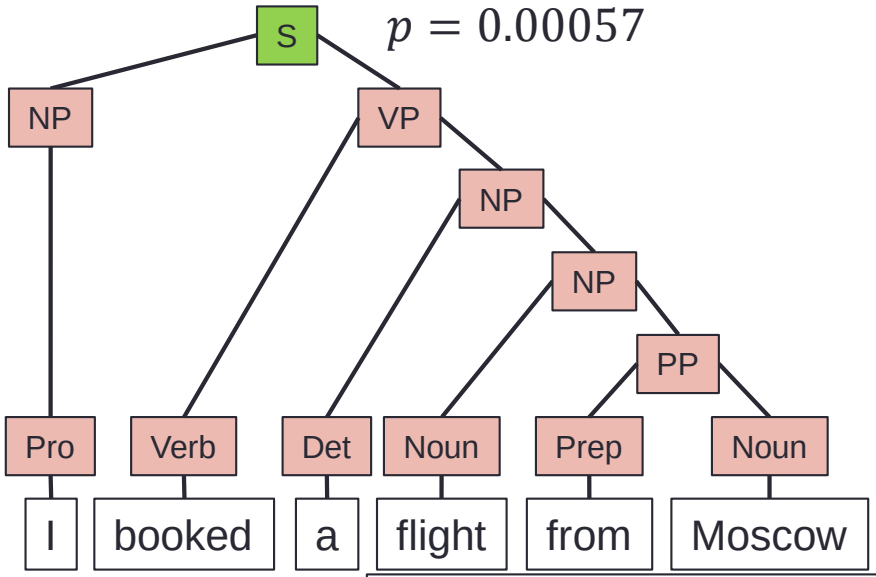
1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

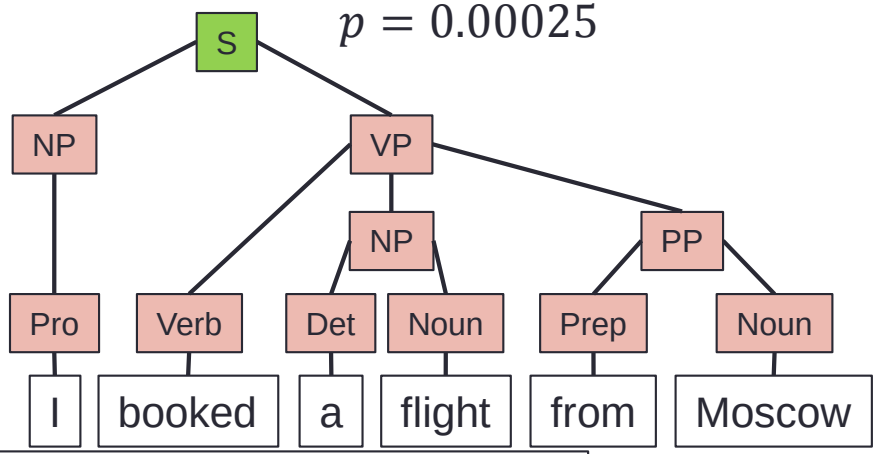
Алгоритм CYK



1.	$S \rightarrow NP\ VP$	1.0
2.	$VP \rightarrow Verb\ NP$	0.9
3.	$VP \rightarrow Verb\ VP^*$	0.1
4.	$VP^* \rightarrow NP\ PP$	1.0
5.	$NP \rightarrow I$	0.2
6.	$NP \rightarrow Det\ NP$	0.3
7.	$NP \rightarrow Noun\ PP$	0.1
8.	$NP \rightarrow flight$	0.12
9.	$NP \rightarrow Moscow$	0.28
10.	$PP \rightarrow Prep\ Noun$	1.0
11.	$Verb \rightarrow booked$	1.0
12.	$Noun \rightarrow flight$	0.3
13.	$Noun \rightarrow Moscow$	0.7
14.	$Pro \rightarrow I$	1.0
15.	$Det \rightarrow a$	0.5
16.	$Det \rightarrow the$	0.5
17.	$Prep \rightarrow from$	1.0

Неоднозначность языка

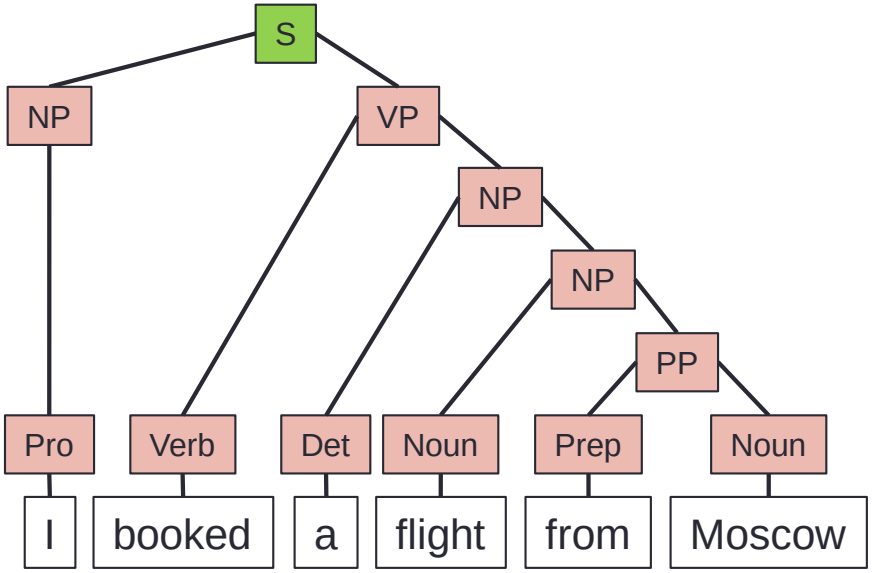




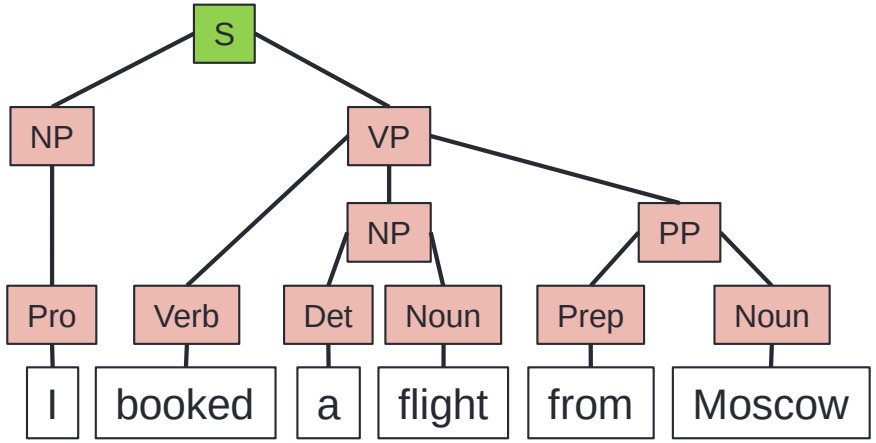
1. $S \rightarrow NP\ VP$
2. $\underline{VP \rightarrow Verb\ NP}$
3. $\underline{VP \rightarrow Verb\ NP\ PP}$
4. $NP \rightarrow Pro$
5. $NP \rightarrow Det\ NP$
6. $NP \rightarrow Noun\ PP$
7. $NP \rightarrow Noun$
8. $PP \rightarrow Prep\ Noun$
9. $Noun \rightarrow flight$
10. $Verb \rightarrow booked$
11. $Noun \rightarrow Moscow$
12. $Pro \rightarrow I$
13. $Det \rightarrow a$
14. $Det \rightarrow the$
15. $Prep \rightarrow from$

Оценка качества

Gold:

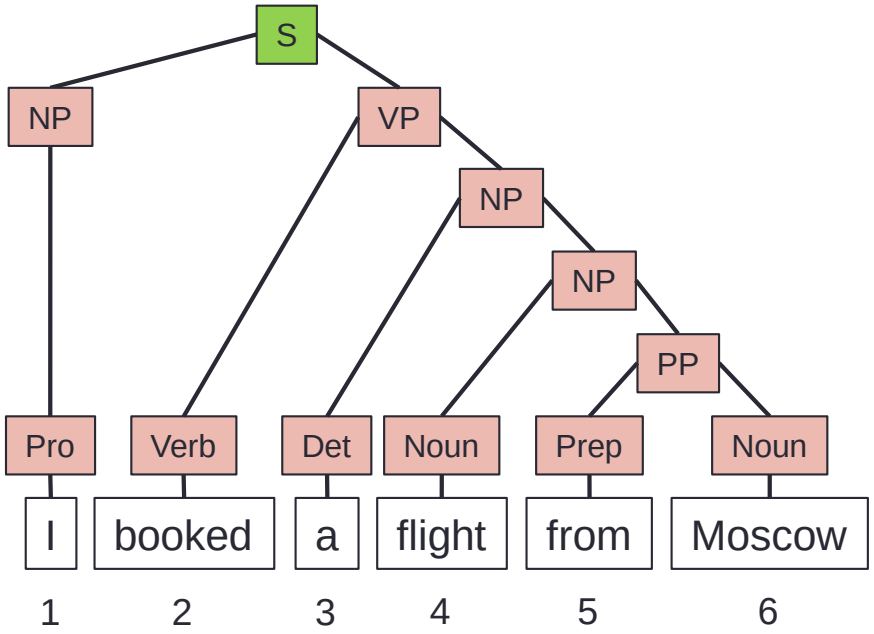


Predicted:

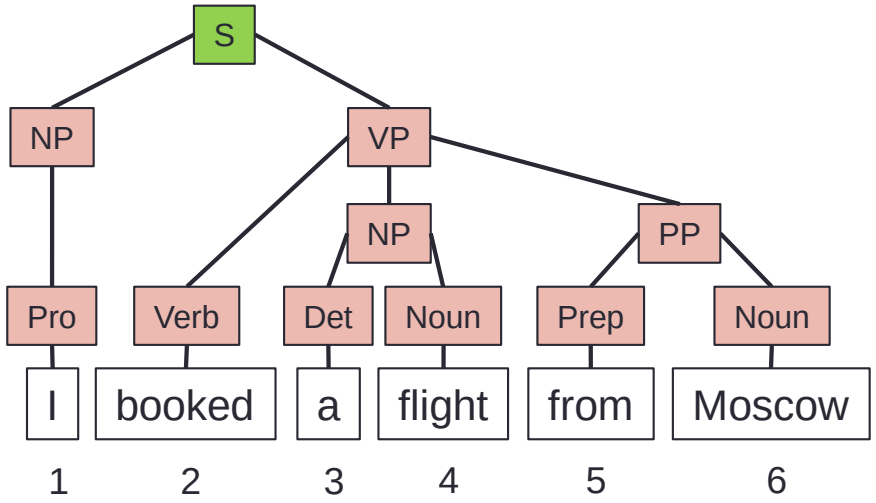


Оценка качества

Gold:



Predicted:



- Gold:

  S(1:6), NP(1:1), VP(2:6), NP(3:6), NP(4:6), PP(5:6)
- Predicted:

  S(1:6), NP(1:1), VP(2:6), NP(3:4), PP(5:6)

Оценка качества

- $Recall = \frac{|correct|}{|gold|}$;
- $Precision = \frac{|correct|}{|predicted|}$;
- $F1 = \frac{2 * Precision * Recall}{Precision + Recall}$;

  - Gold:
    S(1:6), NP(1:1), VP(2:6), NP(3:6), NP(4:6), PP(5:6)
  - Predicted:
    S(1:6), NP(1:1), VP(2:6), NP(3:4), PP(5:6)

$Recall = \frac{2}{3}$ $Precision = \frac{4}{5}$ $F1 = \frac{8}{11}$

Поверхностный синтаксический анализ

- Задача заключается в определении синтаксически связанных групп слов

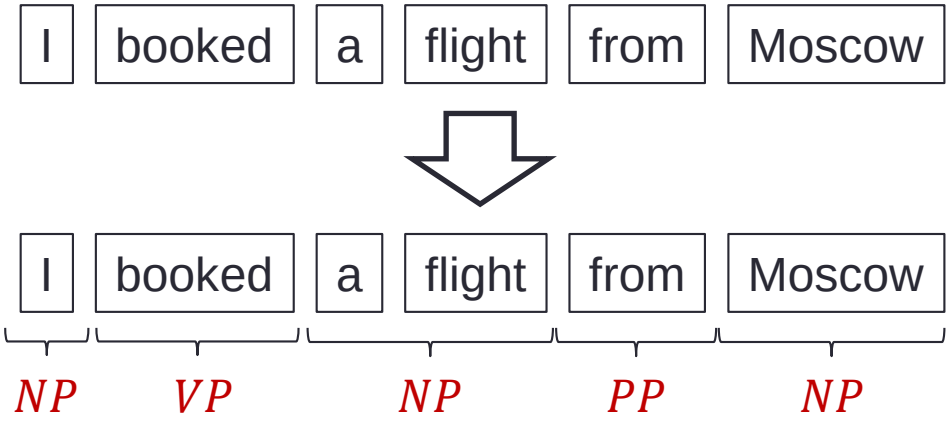
- На входе: последовательность слов предложения
- На выходе: группы синтаксически связанных слов

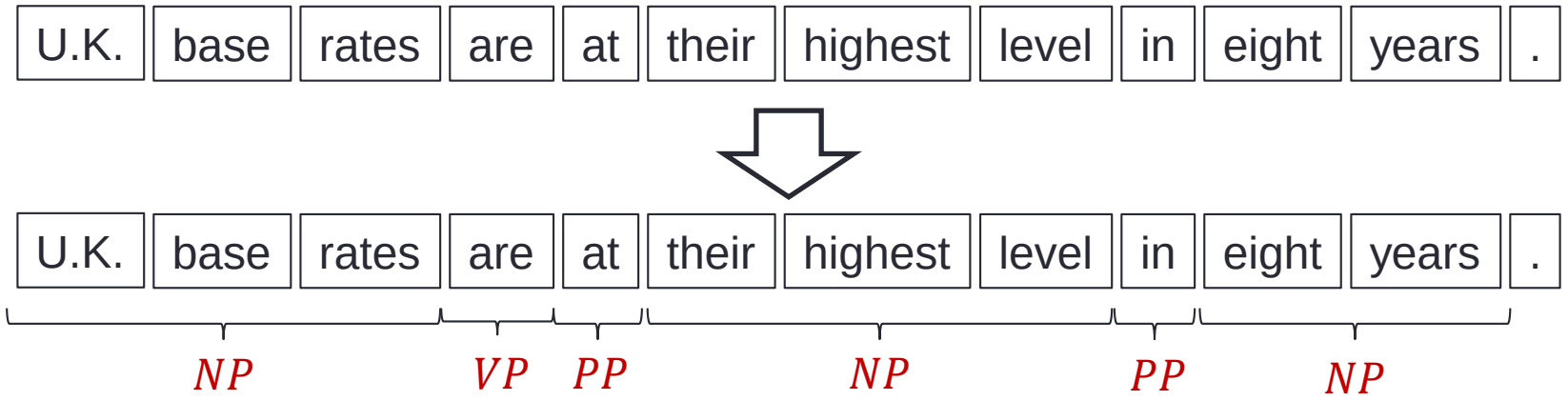
Синтаксические правила

- В зависимости от главного слова в словосочетании выделяют
  - Именные группы (главное – существительное)
  - Группа прилагательного
  - Наречная группа
  - Предложная группа
  - Глагольная группа

Поверхностный синтаксический анализ

- На входе: последовательность слов предложения
- На выходе: группы синтаксически связанных слов

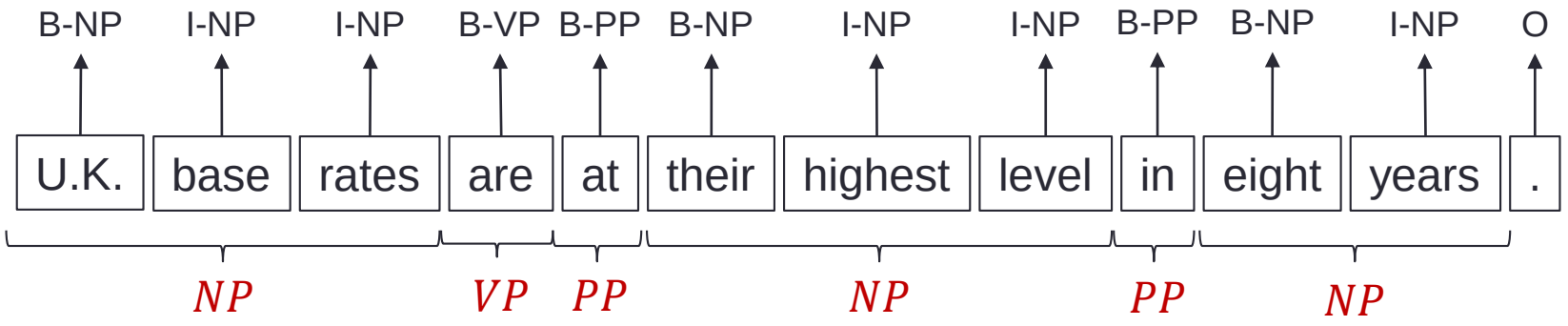




Поверхностный синтаксический анализ

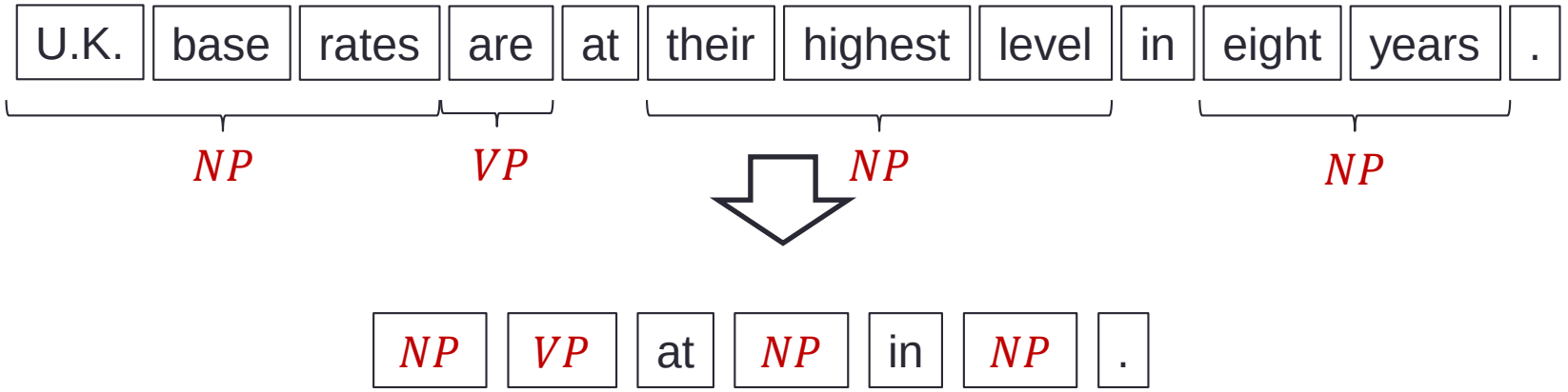
- На входе: последовательность слов предложения
- На выходе: группы синтаксически связанных слов

- Методы решения:
  - BIO кодировка + разметка последовательности



Поверхностный синтаксический анализ

- Применение:
  - Предварительный шаг обработки перед “глубоким” синтаксическим анализом

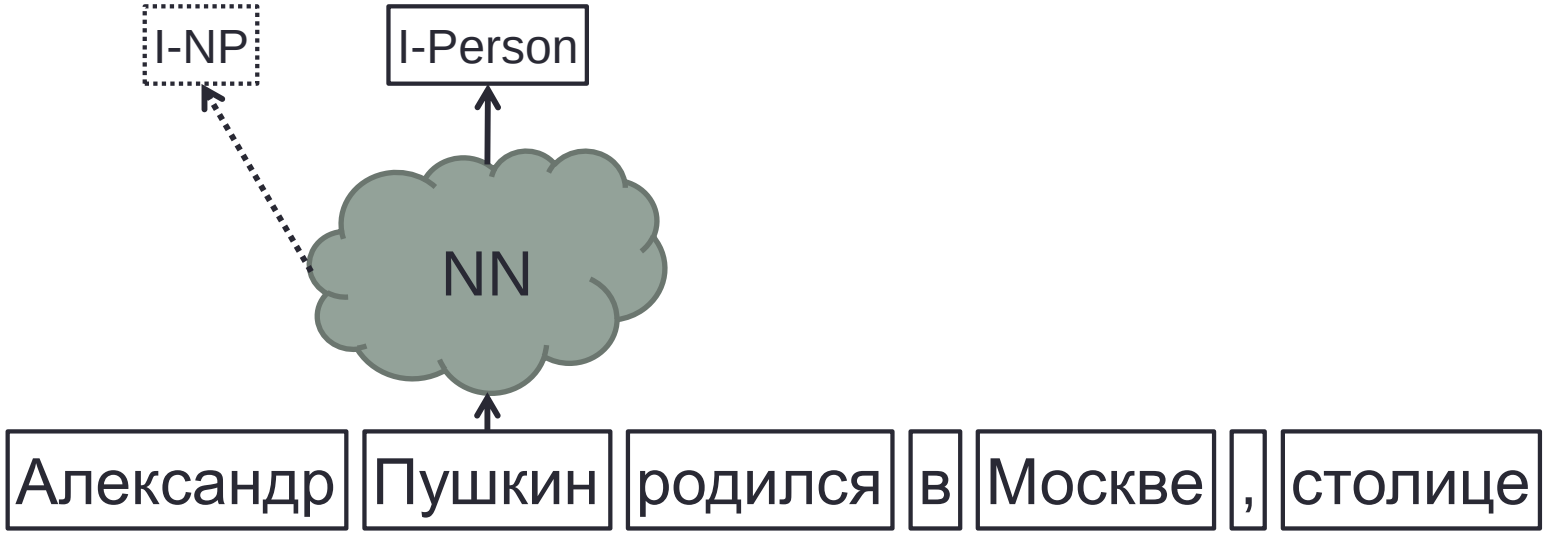


Поверхностный синтаксический анализ

- Применение:
  - Предварительный шаг обработки перед “глубоким” синтаксическим анализом
  - Дополнительные признаки для решения других задач обработки текстов

Поверхностный синтаксический анализ

- Применение:
  - Предварительный шаг обработки перед “глубоким” синтаксическим анализом
  - Дополнительные признаки для решения других задач обработки текстов
  - Вспомогательная задача при решении более высокоуровневых задач обработки текстов



Следующая лекция

Синтаксический анализ (часть 2)