Артем Копецкий, 152

HW-4 «Автоматический синтаксический анализ»

**1. Деревья и КС-грамматика**

Для выполнения задания я выбрал следующие 4 предложения:

1. *Он идет по улице (неперех.)*

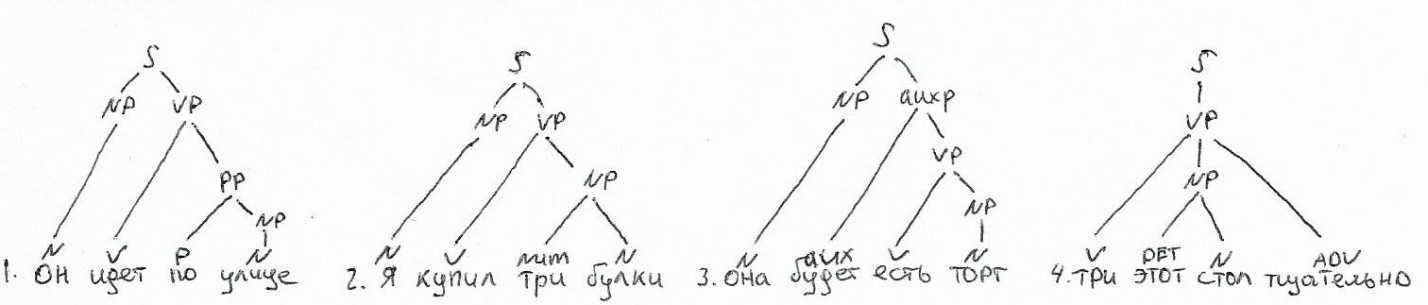
2. *Я купил три булки*

3. *Она будет есть торт*

4. *Три этот стол тщательно (императив + омонимия)*

Для простоты я буду считать местоимения именами, связку *быть* и отрицательную частицу *не* как *aux* и инфинитив – глаголом (V).

Ниже приведены построенные НС-деревья разбора.



Построим КС-грамматику на основе полученных деревьев. Я буду придерживаться конвенции, описанной в Jufarsky, считая терминальными символами только реальные слова языка (лексикон).

S → NP VP | NP AuxP | VP

AuxP → Aux VP

VP → V PP | V NP | V NP ADV

PP → P NP

NP → N | NUM N | DET N

V → {идет, купил, есть, ест, три}

AUX → {будет, не}

N → {он, улице, я, булки, она, торт, стол}

P → {по}

DET → {этот}

NUM → {три}

ADV → {тщательно, вечером}

Теперь для выполнения дальнейших заданий переведем полученную КС-грамматику в нормальную форму Хомского. Перевод в НФХ подразумевает преобразование правил таким образом, чтобы остались только правила вида «НЕТЕРМ → НЕТЕРМ НЕТЕРМ» и «НЕТЕРМ → терм».

S → NP VP

S → NP AuxP

S → V PP

S → V NP

S → V X

X → NP ADV

AuxP → Aux VP

VP → V PP

VP → V NP

VP → V X

PP → P NP

NP → NUM N

NP → DET N

NP → {он, улице, я, булки, она, торт, стол}

V → {идет, купил, есть, ест, три}

AUX → {будет, не}

N → {он, улице, я, булки, она, торт, стол}

P → {по}

DET → {этот}

NUM → {три}

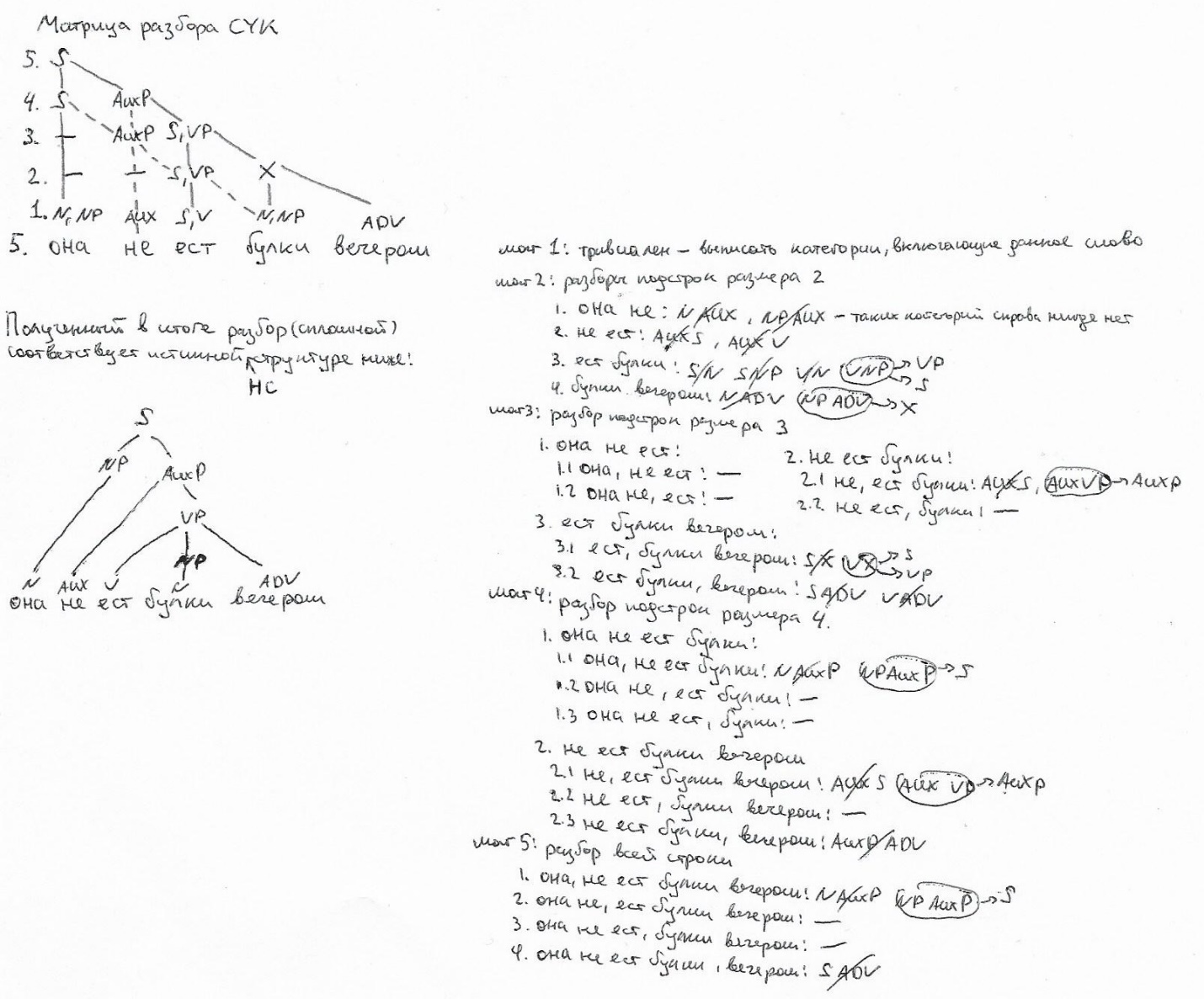
ADV → {тщательно, вечером}

**2 – 3. Парсинг с помощью алгоритма CYK**

Составим на основе имеющегося лексикона новое предложение вида:

5. *Она не ест булки вечером*

Теперь применим алгоритм CYK для выяснения возможных разборов этого предложения и его принадлежности к имеющейся у нас КС-грамматике. Последнее будет справедливо, если в итоге мы придем в ячейку с коренным узлом S.



Алгоритм CYK в целом несет следующую идею: на каждом шаге мы получаем возможные разборы подстрок длины k, постепенно доходя до всего предложения целиком. На каждом шаге возможные разборы устанавливаются как комбинации разборов всех бинарных разбиений подстроки текущего размера. Если на финальном шаге (соответствующем разбору всего предложения) удастся получить коренной узел (S), значит, данное предложение валидно в рамках используемой КС-грамматики. Помимо этого, опускаясь бинарного из корневого узла в полученной матрице можно получить все возможные разборы данного предложения.

Как видно, придуманное предложение 5 валидно в терминах имеющейся грамматики. Истинный разбор (соответствующий разбору в терминах НС) я отметил сплошной линией, а второй возможный – пунктиром.