

Задание по теме «Автоматический синтаксический анализ»
HW4.

4. Алгоритм СΥΚ

4.1. Построить КС-грамматику (можно от руки)

РЕШЕНИЕ

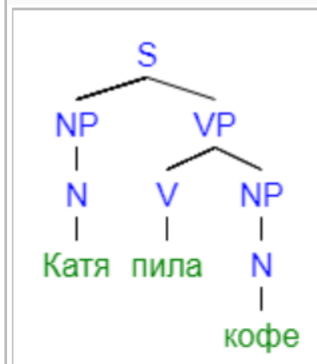
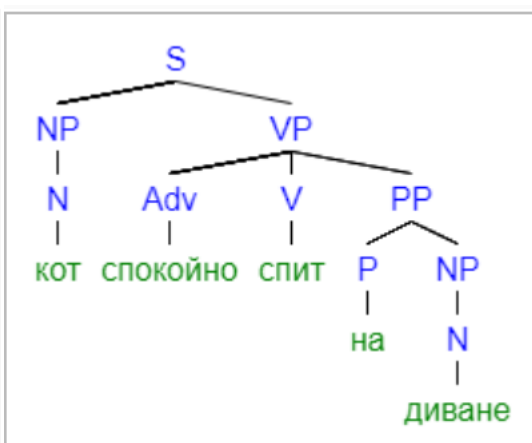
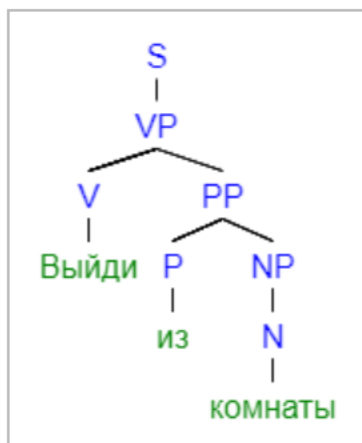
Предложения возьмём из русского языка:

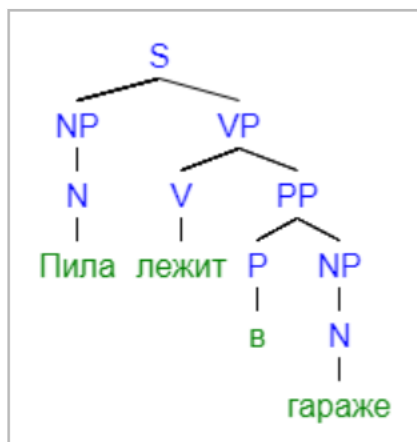
Выйди из комнаты.

Кот спокойно спит на диване.

Катя пила кофе.

Пила лежит в гараже.





Составим правила контекстно-свободной грамматики:

$S \rightarrow NP VP$
 $S \rightarrow VP$
 $NP \rightarrow N$
 $VP \rightarrow V PP$
 $VP \rightarrow Adv V PP$
 $VP \rightarrow V NP$
 $PP \rightarrow P NP$

Терминальные узлы:

$V \rightarrow$ лежать, спать, пить, выйти
 $P \rightarrow$ в, из, на
 $N \rightarrow$ пила, гараж, комната, кот, диван, Катя, кофе
 $Adv \rightarrow$ спокойно

Приведем эту грамматику к нормальной форме Хомского ($A \rightarrow BC$, $A \rightarrow a$ и $S \rightarrow \epsilon$, где A, B и C нетерминальные символы, a терминальный символ, а ϵ пустая строка)

Терминальные узлы остаются такими же. Нетерминальные узлы:

$S \rightarrow NP VP \mid \epsilon$
 $NP \rightarrow N$
 $PP \rightarrow P NP$
 $VP \rightarrow V PP \mid V NP \mid AdvP V$
 $AdvP \rightarrow Adv$

Проведем анализ одного из этих предложений при помощи алгоритма СΥΚ (алгоритма Кока-Янгера-Касами).

Пила лежит в гараже.

S				4
0	VP			3
0	0	PP		2
NP	0	0	NP	1
N пила	V лежит	P в	N гараже	0

Теперь придумаем какое-то новое предложение и попробуем описать его с правилами нашей КС-грамматики алгоритмом СΥΚ.

Катя спит в детской комнате.

Для того, что построить такую же матрицу для нового предложения нам придется добавить еще правил, поскольку в этом предложении есть группа прилагательного.

NP → AdjP NP \ N вместо NP → N

AdjP → Adj

Adj → детский

Остальные правила сохраняются.

Теперь построим матрицу для алгоритма Кока-Янгера-Касами:

S					5
0	VP				4
0	0	PP			3
0	0	0	NP		2
NP	0	0	AdjP	NP	1
N Катя	V спит	P в	Adj детской	N комнате	0

И последнее задание: Обозначьте пунктирными стрелками возможные пути разбора и сплошной стрелкой / стрелками (если допустимы 2 разбора) успешный разбор.

Красными стрелками показаны неосуществленные/неправильные разборы. «спит» могло перейти в VP на первом уровне, и тогда бы связь NP VP \rightarrow S образовалась бы на втором уровне.

Больше, на мой взгляд, никаких дополнительных разборов, следуя правилам нашей КС-грамматики получится не может.

Успешные разборы обозначать стрелками не буду. Из матрицы видно, что свертывание происходит снизу вверх и слева направо. То есть мы всегда смотрим на то, что в клетке ниже и в клетке ниже справа (если там 0, то можно смотреть ещё ниже), и проверяем, есть ли у нас соответствующее правило. Если правила нет, то пишем в клетку 0.