

# 1 Основные определения

*Стохастической КС-грамматикой* [3] называется система  $G = \langle V_T, V_N, R, s \rangle$ , где  $V_T$  и  $V_N$  — конечные множества терминальных и нетерминальных символов (терминалов и нетерминалов) соответственно,  $s \in V_N$  — аксиома,  $R$  — множество правил. Множество  $R$  можно представить в виде  $R = \cup_{i=1}^k R_i$ , где  $k$  — мощность алфавита  $V_N$  и  $R_i = \{r_{i1}, \dots, r_{i,n_i}\}$ . Каждое правило  $r_{ij}$  из  $R_i$  имеет вид

$$r_{ij} : A_i \xrightarrow{p_{ij}} \beta_{ij}, \quad j = 1, \dots, n_i, \quad (1)$$

где  $A_i \in V_N$ ,  $\beta_{ij} \in (V_N \cup V_T)^*$  и  $p_{ij}$  — вероятность применения правила  $r_{ij}$ , причём

$$0 < p_{ij} \leq 1, \quad \sum_{j=1}^{n_i} p_{ij} = 1. \quad (2)$$

Для  $\alpha, \gamma \in (V_N \cup V_T)^*$  будем обозначать  $\alpha \Rightarrow \gamma$ , если существуют  $\alpha_1, \alpha_2 \in (V_N \cup V_T)^*$ , для которых  $\alpha = \alpha_1 A_i \alpha_2$ ,  $\gamma = \alpha_1 \beta_{ij} \alpha_2$  и в грамматике имеется правило  $A_i \xrightarrow{p_{ij}} \beta_{ij}$ . Через  $\Rightarrow_*$  обозначим рефлексивное транзитивное замыкание отношения  $\Rightarrow$ . Грамматика  $G$  задаёт контекстно-свободный язык  $L_G = \{\alpha \in V_T^* : s \Rightarrow_* \alpha\}$ .

Выводом слова  $\alpha$  назовём последовательность правил  $\omega(\alpha) = (r_{i_1 j_1}, r_{i_2 j_2}, \dots, r_{i_q j_q})$ , с помощью последовательного применения которых слово  $\alpha$  выводится из аксиомы  $s$ . Если при этом каждое правило применяется к самому левому нетерминалу в слове, такой вывод называется левым. Для вывода  $\omega(\alpha) = (r_{i_1 j_1}, \dots, r_{i_q j_q})$  определим величину  $p(\omega(\alpha)) = p_{i_1 j_1} \cdot \dots \cdot p_{i_q j_q}$ .

Важное значение имеет понятие *дерева вывода* [4]. Дерево вывода для слова  $\alpha$  строится следующим образом. Корень дерева помечается аксиомой  $s$ . Далее последовательно рассматриваются правила левого вывода слова  $\alpha$ . Пусть на очередном шаге рассматривается правило  $A_i \xrightarrow{p_{ij}} b_{i1} b_{i2} \dots b_{im}$ , где  $b_{il} \in (V_N \cup V_T)$  ( $l = 1, \dots, m$ ). Тогда из самой левой вершины-листа дерева, помеченной символом  $A_i$ , проводится  $m$  дуг в вершины следующего яруса, которые помечаются слева направо символами  $b_{i1}, \dots, b_{im}$  соответственно. После построения дуг и вершин для всех правил в выводе листья дерева помечены терминальными символами (либо пустым словом  $\lambda$ , если применяется правило вида  $A_i \xrightarrow{p_{ij}} \lambda$ ) и само слово получается при обходе листьев дерева слева направо. *Высотой* дерева вывода будем называть максимальную длину пути от корня к листу.

Обозначим  $p(\alpha) = \sum \omega(\alpha)$ , где сумма берётся по всем левым выводам слова  $\alpha$ . Грамматика  $G$  называется *согласованной*, если

$$\lim_{n \rightarrow \infty} \sum_{\substack{\alpha \in L_G \\ |\alpha| \leq n}} p(\alpha) = 1. \quad (3)$$

Согласованная грамматика  $G$  задаёт распределение вероятностей  $P$  на множестве  $L_G$  и определяет *стохастический КС-язык*  $\mathcal{L} = (L_G, P)$ . В этом случае величина  $p(\alpha)$  — вероятность слова  $\alpha$  в  $L_G$ . В дальнейшем будем рассматривать только согласованные грамматики.

Определим многомерные производящие функции [3]:

$$F_i(s_1, s_2, \dots, s_k) = \sum_{j=1}^{n_i} p_{ij} s_1^{l_1} s_2^{l_2} \dots s_k^{l_k} \quad (1 \leq i \leq k), \quad (4)$$

где  $n_i$  — число правил вывода с нетерминалом  $A_i$  в левой части, и  $l_m = l_m(i, j)$  — число вхождений нетермина  $A_m$  в правую часть правила  $A_i \xrightarrow{p_{ij}} \beta_{ij}$ .

Величины

$$a_j^i = \frac{\partial F_i(s_1, s_2, \dots, s_k)}{\partial s_j} \Big|_{s_1=\dots=s_k=1} \quad (5)$$

называются первыми моментами грамматики и образуют матрицу первых моментов  $A = (a_j^i)$ . Аналогично определяются вторые моменты

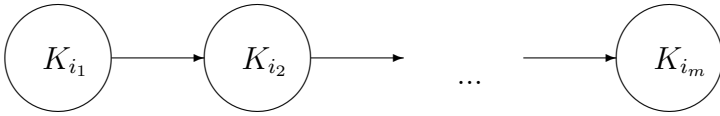
$$b_{jl}^i = \frac{\partial^2 F_i(s_1, s_2, \dots, s_k)}{\partial s_l \partial s_j} \Big|_{s_1=\dots=s_k=1} \quad (6)$$

Для нетерминалов  $A_i, A_j$  будем обозначать  $A_i \rightarrow A_j$ , если в грамматике имеется правило  $A_i \xrightarrow{p_{ij}} \alpha_1 A_j \alpha_2$ , где  $\alpha_1, \alpha_2 \in (V_N \cup V_T)^*$ . Рефлексивное транзитивное замыкание отношения  $\rightarrow$  обозначим  $\rightarrow_*$ . Если одновременно  $A_i \rightarrow_* A_j$  и  $A_j \rightarrow_* A_i$ , будем обозначать  $A_i \leftrightarrow_* A_j$ . Отношение  $\leftrightarrow_*$  разбивает множество нетерминалов грамматики на классы

$$K_1, K_2, \dots, K_m. \quad (7)$$

Множества номеров нетерминалов, входящих в класс  $K_j$  обозначим через  $I_j$ . При  $m \geq 2$  грамматика называется *разложимой*.

Разложимой грамматике соответствует разложимая матрица [6] первых моментов. Обозначим  $K_i \prec K_j$ , если  $i \neq j$  и существуют такие  $A_1 \in K_i$  и  $A_2 \in K_j$ , что  $A_1 \rightarrow A_2$ . Будем говорить, что грамматика имеет вид «цепочки», если она разложима, и для множества классов выполняется соотношение  $K_1 \prec K_2 \prec \dots \prec K_m$ . При этом граф, построенный на множестве классов по отношению  $\prec$ , имеет вид:



Назовём класс  $K$  *особым*, если он содержит ровно один нетерминал  $A_i$ , и в грамматике отсутствует правило вида  $A_i \xrightarrow{p_{ij}} \alpha_1 A_i \alpha_2$ , где  $\alpha_1, \alpha_2 \in (V_N \cup V_T)^*$ . Не уменьшая общности, будем считать, что грамматика не имеет особых классов.

Матрицу первых моментов грамматики, имеющей вид «цепочки» можно представить следующим образом.

$$A = \begin{pmatrix} A_{11} & A_{12} & 0 & \dots & 0 & 0 \\ 0 & A_{22} & A_{23} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_{m-1,m-1} & A_{m-1,m} \\ 0 & 0 & 0 & \dots & 0 & A_{m,m} \end{pmatrix}. \quad (8)$$

Блок  $A_{ii}$  соответствует классу  $K_i$  и является неразложимой неотрицательной матрицей.

По определению (5), матрицы  $A_{11}, A_{22}, \dots, A_{m,m}$  неразложимы и неотрицательны. Согласно теореме Фробениуса [6], неразложимая неотрицательная матрица всегда имеет простое положительное максимальное по модулю собственное число (перронов корень). Обозначим перроновы корни матриц  $A_{11}, \dots, A_{m,m}$  через  $r_1, \dots, r_m$  соответственно. Тогда  $r = \max\{r_1, \dots, r_m\}$  — перронов корень всей матрицы  $A$ . В данной работе рассматривается случай  $r = 1$ . По аналогии с теорией ветвящихся процессов [5] будем называть этот случай *критическим*. Классы, для которых перронов корень соответствующих подматриц равен 1, также будем называть критическими.

Обозначим  $s_{lh}$  (при  $l \leq h$ ) — число критических классов среди подцепочки  $K_l, K_{l+1}, \dots, K_h$ . Разобьём последовательность классов  $K_1, K_2, \dots, K_m$  на группы  $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_w$ , где  $w = s_{1m}$ . Класс  $K_l$  отнесём к группе  $\mathcal{M}_w$  при  $s_{lw} \leq 1$  и к группе  $\mathcal{M}_{w-j+1}$  при  $s_{lw} = j$  ( $j = 2, \dots, w$ ).

Тогда матрицу  $A$  можно представить в виде:

$$A = \begin{pmatrix} B_{11} & B_{12} & 0 & \cdots & 0 & 0 \\ 0 & B_{22} & B_{23} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & B_{w-1,w-1} & B_{w-1,w} \\ 0 & 0 & 0 & \cdots & 0 & B_{w,w} \end{pmatrix},$$

где матрица  $B_{lh}$  находится на пересечении строк для классов из группы  $\mathcal{M}_l$  и столбцов для классов из группы  $\mathcal{M}_h$ . Матрицы  $B_{lh}$ , в свою очередь, имеют вид

$$B_{lh} = \begin{pmatrix} C_{11} & C_{12} & 0 \\ 0 & C_{22} & C_{23} \\ 0 & 0 & C_{33} \end{pmatrix},$$

где  $C_{22}$  — блок, стоящий на пересечении строк для  $l$ -го критического класса и столбцов для  $h$ -го критического класса. При  $l = h$  этот блок является неразложимой матрицей. Блоки  $C_{11}$  и  $C_{33}$  стоят на пересечении строк и столбцов, соответствующих докритическим классам. При  $l, h < w$  блок  $B_{lh}$  имеет вид

$$B_{lh} = \begin{pmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{pmatrix}.$$

Блок, находящийся на позиции блока  $B_{lh}$  в матрице  $A^t$ , обозначим  $B_{lh}^{(t)}$ .

В [7] доказана следующая теорема.

**Теорема 1** При  $t \rightarrow \infty$

$$B_{lh}^{(t)} = \begin{pmatrix} 0 & b \cdot U_I^{(l)} V_{II}^{(h)} & b \cdot U_I^{(l)} V_{III}^{(h)} \\ 0 & b \cdot U_{II}^{(l)} V_{II}^{(h)} & b \cdot U_{II}^{(l)} V_{III}^{(h)} \\ 0 & 0 & 0 \end{pmatrix} = b \cdot U^{(l)} V^{(h)} t^{s_{lh}-1} r^t \cdot (1 + o(1)),$$

где  $U^{(a)}$  и  $V^{(a)}$  — правый и левый собственные векторы матрицы  $B_{aa}$ , и  $b = V^{(l)} B_{lh} U^{(h)}$ .

## 2 Вероятности продолжения

Введём обозначения

$$\begin{aligned} F_i(\mathbf{s}) &= F_i(0, \mathbf{s}) = \sum_{j=1}^{n_i} p_{ij} s_1^{l_1} s_2^{l_2} \dots s_k^{l_k} \quad (1 \leq i \leq k), \\ F_i(t, \mathbf{s}) &= F_i(\mathbf{F}(t-1, \mathbf{s})) \quad (t > 0, 1 \leq i \leq k), \end{aligned} \quad (9)$$

где  $\mathbf{s} = (s_1, \dots, s_k)$ ,  $0 \leq s_j \leq 1$  и  $\mathbf{F}(t, \mathbf{s}) = (F_1(t, \mathbf{s}), \dots, F_k(t, \mathbf{s}))$ .

В силу согласованности грамматики  $F_i(t, \mathbf{0}) \rightarrow 1$  при  $t \rightarrow \infty$ . В самом деле, из (3) следует

$$\sum_{d \in D} p(d) = \lim_{t \rightarrow \infty} \sum_{d \in D \leq t} p(d) = 1,$$

откуда

$$\lim_{t \rightarrow \infty} \sum_{d \in D > t} p(d) = 0$$

и

$$F_i(t, \mathbf{0}) = 1 - \sum_{d \in D > t} p(d) \rightarrow 1.$$

Раскладывая  $F_i(\mathbf{s})$  в ряд Тейлора в окрестности  $\mathbf{s} = (1, \dots, 1)$ , и учитывая равенство  $F_i(1, 1, \dots, 1) = 1$ , получаем:

$$1 - F_i(\mathbf{s}) = \sum_{j=1}^{n_i} a_j^i (1 - s_j) - \frac{1}{2} \sum_{1 \leq j, l \leq n_i} b_{jl}^i (1 - s_j)(1 - s_l) + O\left(\sum_{j=1}^k |1 - s_j|^3\right), \quad (10)$$

Подставляя в качестве  $\mathbf{s}$  вектор  $\mathbf{F}(t, s) = (F_1(t, s), F_2(t, s), \dots, F_k(t, s))$ , получаем:

$$\begin{aligned} 1 - F_i(t+1, s) &= \sum_{j=1}^k a_j^i (1 - F_j(t, s)) - \frac{1}{2} \sum_{1 \leq j, l \leq k} b_{jl}^i (1 - F_j(t, s))(1 - F_l(t, s)) + \\ &\quad + O\left(\sum_{j=1}^k |1 - F_j(t, s)|^3\right) \end{aligned} \quad (11)$$

Введём вектор *вероятностей продолжения*  $Q(t) = (Q_1(t), Q_2(t), \dots, Q_k(t))^T$ , положив

$$Q_i(t) = 1 - F_i(t, \mathbf{s})|_{s_1=s_2=\dots=s_k=0} \quad (12)$$

В силу (2)  $Q_i(t) \rightarrow 0$  при  $t \rightarrow \infty$ .

Тогда уравнение (11) примет вид

$$Q_i(t+1) = \sum_{j=1}^k a_j^i Q_j(t) - \frac{1}{2} \sum_{1 \leq j, l \leq k} b_{jl}^i Q_j(t) Q_l(t) + O\left(\sum_{j=1}^k |Q_j(t)|^3\right) \quad (13)$$

Для каждого из классов  $K_n$  будем рассматривать вектор  $Q^{(n)}(t)$  — вектор-столбец, содержащий вероятности продолжения для нетерминалов из класса  $K_n$  в порядке их нумерации. Тогда

$$Q(t) = \begin{pmatrix} Q^{(1)}(t) \\ Q^{(2)}(t) \\ \vdots \\ Q^{(m)}(t) \end{pmatrix}, \quad Q^{(j)}(t) \in \mathbb{R}^{k_j}, \quad (14)$$

где  $k_j = |K_j|$ . Обозначим через  $I_n$  множество индексов нетерминалов, входящих в класс  $K_n$ . Используя это обозначение, уравнение (13) можно записать в виде

$$Q_i(t+1) = \sum_{j \in I_n} a_j^i Q_j(t) + \sum_{i \in I_{n+1}} a_j^i Q_j(t) \cdot (1 + o(1)) \quad (i \in I_n, n < m) \quad (15)$$

$$Q_i(t+1) = \sum_{j \in I_m} a_j^i Q_j(t) \cdot (1 + o(1)) \quad (i \in I_m) \quad (16)$$

или, используя вид (8) матрицы первых моментов,

$$Q^{(n)}(t+1) = A_{n,n} Q^{(n)}(t) + A_{n,n+1} Q^{(n+1)}(t)(1 + o(1)) \quad (17)$$

Для всего вектора  $Q(t)$  верно равенство

$$Q(t+1) = (A - A(t))Q(t), \quad (18)$$

где  $A(t)$  — матрица, составленная из элементов  $a_{ij} = \frac{1}{2} \sum_{l=1}^k b_{jl}^i Q_l(t)$  ( $1 \leq i, j \leq k$ ). В силу согласованности грамматики  $Q(t) \rightarrow 0$  и, следовательно,  $A(t) \rightarrow 0$  при  $t \rightarrow \infty$ .

Докажем, что компоненты вектора  $Q^{(n)}(t)$  пропорциональны некоторому вектору  $U^{(n)}$ . Доказательство аналогичного факта для случая двух классов принадлежит А. Борисову. Здесь мы проведём похожие рассуждения.

Зафиксируем некоторое  $\tau \geq 0$ . Тогда из (18) получаем

$$Q(t+1) = (A - A(t)) \cdot \dots \cdot (A - A(\tau))Q(\tau) \quad (19)$$

Обозначим

$$\begin{aligned} A^*(t) &= (A - A(t)) \cdot (A - A(t-1)) \cdot \dots \cdot (A - A(\tau+1)) \\ \tilde{A}_{ij}^* &= \frac{A_{ij}^*(t)}{t^{s_{ij}}} \\ \tilde{A}_{ij} &= \frac{A_{ij}^{(t)}}{t^{s_{ij}}}, \end{aligned} \quad (20)$$

где  $A_{ij}^{(t)}$  — блоки, расположенные на месте блоков  $A_{ij}$  в матрице  $A^t$  и  $s_{ij}$  — число критических классов в подцепочке  $K_i, K_{i+1}, \dots, K_j$ .

Из исследования асимптотики матрицы  $A^t$  известно [7], что  $\tilde{A}_{ij}(t) \rightarrow \tilde{a}_{ij} U^{(i)} V^{(j)}$ , где  $\tilde{a}_{ij}$  — некоторые константы,  $U^{(i)}$  — вектор-строка длины  $k_i$ , а  $V^{(j)}$  — вектор-столбец длины  $k_j$ .

Выберем произвольные  $\varepsilon_1, \varepsilon_2$ , такие что  $0 < \varepsilon_1, \varepsilon_2 < 1$ . Тогда существуют функции  $l(\varepsilon_1)$  и  $n(\varepsilon_2)$ , такие что

$$\begin{aligned} & \left| \tilde{A}_{ij}(l(\varepsilon_1)) - \tilde{a}_{ij}U^{(i)}V^{(j)} \right| < \varepsilon_1 E \\ & \forall t \geq n(\varepsilon_2) \quad A(t) < \varepsilon_2 A \end{aligned} \quad (21)$$

Рассмотрим произвольный вектор-столбец  $x > \mathbf{0}$  длины  $k$ . Тогда выполняется оценка

$$(1 - \varepsilon_2)^l A^l x^{(\tau)} \leq A^*(t) x^{(\tau)} \leq A^l x^{(\tau)}, \quad (22)$$

где  $x^{(\tau)} = (A - A(\tau))x$ . Записывая это неравенство отдельно для блоков  $A_{ij}$ , получаем

$$(1 - \varepsilon_2)^l A_{ij}^l x_j^{(\tau)} \leq A_{ij}^*(l) x_j^{(\tau)} \leq A_{ij}^{(l)} x_j^{(\tau)}, \quad (23)$$

откуда

$$(1 - \varepsilon_2)^l \tilde{A}_{ij}(l) x^{(\tau)} \leq \tilde{A}_{ij}^*(l) x_j^{(\tau)} \leq \tilde{A}_{ij}(l) x^{(\tau)} \quad (24)$$

Вычитая из всех частей неравенства  $\tilde{A}_{ij}(l) x_j^{(\tau)}$ , получаем оценку

$$\left| \left( \tilde{A}_{ij}^*(l) - \tilde{A}_{ij}(l) \right) x_j^{(\tau)} \right| \leq (1 - (1 - \varepsilon_2)^l) \tilde{A}_{ij}(l) x^{(\tau)} \quad (25)$$

Используя эту оценку, можем записать

$$\begin{aligned} & \left| \tilde{A}_{ij}^*(t) - \tilde{a}_{ij}U^{(i)}V^{(j)} x_j^{(\tau)} \right| \leq \left| \left( \tilde{A}_{ij}^*(t) - \tilde{A}_{ij}(t) \right) x^{(\tau)} \right| + \\ & + \left| \left( \tilde{A}_{ij}(t) - \tilde{a}_{ij}U^{(i)}V^{(j)} \right) x_j^{(\tau)} \right| \leq (1 - (1 - \varepsilon_2)^l) \tilde{A}_{ij}(l) x_j^{(\tau)} + \varepsilon_1 x_j^{(\tau)} \leq \\ & \leq (1 - (1 - \varepsilon_2)^l) h k_j x_j^{(\tau)} + \varepsilon_1 x_j^{(\tau)} \leq ((1 - 1 - \varepsilon_2)^l) h k_j + \varepsilon_1 x_j^*(\tau), \end{aligned} \quad (26)$$

где  $h = \max_{i,j,l} \left\{ \tilde{A}_{ij}(l) \right\}$  и  $x_j^*(\tau) = \max_i (x_j^{(\tau)})_i$ .

Устремляем  $\varepsilon_2$  к нулю, затем  $\varepsilon_1$  к нулю таким образом, чтобы выполнялось условие

$$l(\varepsilon_1) \log(1 - \varepsilon_2) \rightarrow -\infty \quad (27)$$

Тогда

$$\left| \tilde{A}_{ij}^*(t) - \tilde{a}_{ij}U^{(i)}V^{(j)} x_j^{(\tau)} \right| \leq \varepsilon x_j^*(\tau) \quad (\varepsilon \rightarrow 0). \quad (28)$$

Домножая слева на  $V^{(i)}$ , имеем

$$\left| V^{(i)} \tilde{A}_{ij}^*(t) x_j^{(\tau)} - \tilde{a}_{ij} V^{(j)} x_j^{(\tau)} \right| \leq \varepsilon k_i \max \{ (V^{(i)}) \} x_j^*(\tau) \leq \varepsilon^* V^{(j)} x_j^{(\tau)}. \quad (29)$$

Отсюда,

$$\left| \frac{\tilde{A}_{ij}^*(t) x_j^{(\tau)}}{V^{(i)} \tilde{A}_{ij}^*(t) x_j^{(\tau)}} - \frac{\tilde{a}_{ij} U^{(i)} V^{(j)} x_j^{(\tau)}}{\tilde{a}_{ij} V^{(j)} x_j^{(\tau)}} \right| = \left| \frac{\tilde{A}_{ij}^*(t) x_j^{(\tau)}}{V^{(i)} \tilde{A}_{ij}^*(t) x_j^{(\tau)}} - U^{(i)} \right| \rightarrow 0 \quad (30)$$

или же

$$\left| \frac{A_{ij}^*(t)x_j^{(\tau)}}{V^{(i)}A_{ij}^*(t)x_j^{(\tau)}} - U^{(i)} \right| \rightarrow 0, \quad (31)$$

откуда

$$(A - A(t)) \cdot \dots \cdot (A - A(\tau)) \cdot x_j = U^{(i)} V^{(i)} (A - A(t)) \cdot \dots \cdot (A - A(\tau)) \cdot x_j \cdot (1 + o(1)) \quad (32)$$

Ввиду полученного выражения и (19) компоненты каждого из векторов  $Q^{(n)}(t)$  пропорциональны компонентам вектора  $U^{(n)}$ .

Оценим теперь асимптотику элементов вектора  $Q^{(n)}(t)$  при  $t \rightarrow \infty$ .

Положим  $V^{(n)}Q^{(n)}(t) = Q_*^{(n)}(t)$ , и домножим уравнение (13) скалярно на  $V^{(n)}$ . Заметим, что

$$Q^{(n)}(t) = U^{(n)} Q_*^{(n)}(t) (1 + o(1)). \quad (33)$$

$$\begin{aligned} Q_*^{(n)}(t+1) &= Q(n)_*(t) + V^{(n)} B_{n,n+1} U^{(n+1)} Q_*^{(n+1)}(t) - \\ &\quad - \frac{1}{2} \sum_{1 \leq i, j, l \leq k_n} V_i^{(n)} b_{jl}^i(n) U_j^{(n)} U_l^{(n)} (Q_*^{(n)}(t))^2 (1 + o(1)). \end{aligned} \quad (34)$$

Обозначим  $\delta Q_*^{(n)}(t) = Q_*^{(n)}(t+1) - Q_*^{(n)}(t)$ , а также

$$\begin{aligned} b_n &= V^{(n)} B_{n,n+1} U^{(n+1)} \\ B_n &= \sum_{1 \leq i, j, l \leq k_n} V_i^{(n)} b_{jl}^i(n) U_j^{(n)} U_l^{(n)} \end{aligned}$$

Тогда уравнение (34) переписется как

$$\delta Q_*^{(n)}(t) = b_n Q_*^{(n+1)}(t) - \frac{1}{2} B_n (Q_*^{(n)}(t))^2 (1 + o(1)) \quad (35)$$

Выражение для  $\delta Q_*^{(n)}(t)$  также можно получить из (13), вычитая это уравнение из себя с заменой  $t \rightarrow t+1$ :

$$\begin{aligned} \delta Q_*^{(n)}(t+1) &= \sum_{j=1}^{k_n} a_j^i(n) \delta Q_j^{(n)}(t) + \sum_{j=1}^{k_{n+1}} a_j^i(n) \delta Q_j^{(n+1)}(t) - \\ &\quad - \frac{1}{2} \sum_{1 \leq j, l \leq k_n} b_{jl}^i(n) \left( Q_j^{(n)}(t+1) Q_l^{(n)}(t+1) - Q_j^{(n)}(t) Q_l^{(n)}(t) \right) (1 + o(1)) \end{aligned}$$

Скалярно домножая на  $V^{(n)}$ , получим

$$\begin{aligned} \delta Q_*^{(n)}(t+1) &= \delta Q_*^{(n)}(t) + b_n \delta Q_*^{(n+1)}(t) - \\ &\quad - \frac{1}{2} B_n \delta Q_*^{(n)}(t) (Q_*^{(n)}(t+1) + Q_*^{(n)}(t)) (1 + o(1)) \end{aligned} \quad (36)$$

Для последнего класса

$$Q_*^{(w)}(t) = c_w t^{-1} (1 + o(1)), \quad (37)$$

что следует из неразложимого случая. Проведём рассуждение по индукции. Пусть для группы с номером  $n + 1$  верно

$$Q_*^{(n+1)}(t) = c_{n+1}t^{-\alpha}(1 + o(1)),$$

где  $0 < \alpha \leq 1$ . Положим

$$z(t) = t^\alpha \delta Q_*^{(n)}(t)$$

Произведя замену в уравнении (36), и имея в виду, что  $Q_*^{(n)}(t + 1) = O(Q_*^{(n)}(t))$ , получаем

$$\frac{z(t+1)}{(t+1)^\alpha} - \frac{z(t)}{t^\alpha} = b_n \delta Q_*^{(n+1)}(t)(1 + o(1)) - \frac{1}{2} B_n \frac{z(t)}{t^\alpha} \cdot 2Q_*^{(n)}(t)(1 + o(1))$$

Преобразуем выражение в левой части уравнения:

$$\begin{aligned} \frac{z(t+1)}{(t+1)^\alpha} - \frac{z(t)}{t^\alpha} &= \frac{t^\alpha z(t+1) - (t+1)^\alpha z(t)}{t^\alpha (t+1)^\alpha} = \\ &= \frac{t^\alpha z(t+1) - t^\alpha \left(1 + \frac{\alpha}{t} + o\left(\frac{1}{t}\right)\right) z(t)}{t^\alpha (t+1)^\alpha} = \frac{\delta z(t)}{(t+1)^\alpha} - \frac{\alpha z(t)(1 + o(1))}{t(t+1)^\alpha} \end{aligned}$$

Тогда

$$\frac{\delta z(t)}{(t+1)^\alpha} - \frac{\alpha z(t)(1 + o(1))}{t(t+1)^\alpha} = b_n \delta Q_*^{(n)}(t) - \frac{B_n}{t^\alpha} Q_*^{(n)}(t) z(t)(1 + o(1))$$

По предположению индукции,  $\delta Q_*^{(n+1)}(t) = -\frac{c_{n+1}\alpha}{t(t+1)^\alpha}(1 + o(1))$ , и тогда

$$\frac{\delta z(t)}{(t+1)^\alpha} - \frac{\alpha z(t)(1 + o(1))}{t(t+1)^\alpha} = -\frac{b_n \alpha c_{n+1}}{t(t+1)^\alpha} - \frac{B_n}{t^\alpha} Q_*^{(n)}(t) z(t)(1 + o(1))$$

Домножая на  $(t+1)^\alpha$ , получаем

$$\delta z(t) - \frac{\alpha z(t)}{t} = -\frac{b_n \alpha c_{n+1}}{t} - B_n Q_*^{(n)}(t) z(t)(1 + o(1))$$

Заметим, что, в силу предположения индукции,  $\frac{1}{t} \leq Q_*^{(n+1)}(t) = o(Q_*^{(n)}(t))$ , поэтому можно записать

$$\delta z(t) = -\frac{b_n \alpha c_{n+1}}{t} - B_n Q_*^{(n)}(t) z(t)(1 + o(1)) \quad (38)$$

Известна следующая лемма (доказательство леммы принадлежит А. Борисову).

**Лемма 1** Пусть последовательность  $z(t)$  ( $t = 1, 2, \dots$ ) удовлетворяет рекуррентному соотношению

$$\delta z(t) = f(t) - g(t)z(t),$$

где при  $t \rightarrow \infty$  выполняются условия

$$g(t) \rightarrow 0, \frac{f(t)}{g(t)} \rightarrow 0, \sum_{k=1}^t g(k) \rightarrow \infty.$$

Пусть также  $g(t) > 0$  при любом  $t > t_0$ . Тогда  $z(t) \rightarrow 0$  при  $t \rightarrow \infty$ .



Полагая в уравнении (38)  $f(t) = -\frac{b_n \alpha c_{n+1}}{t}(1 + o(1))$ ,  $g(t) = B_n Q_*^{(n)}(t)(1 + o(1))$ , замечаем, что для  $z(t)$  выполняются все условия леммы (1), и соответственно,  $z(t) \rightarrow 0$  при  $t \rightarrow \infty$ . Из определения  $z(t)$  получаем:

$$\delta Q_*^{(n)}(t) = o\left(\frac{1}{t^\alpha}\right).$$

Подставляя эту оценку в (35), получаем

$$o\left(\frac{1}{t^\alpha}\right) = \frac{b_n c_{n+1}}{t^\alpha}(1 + o(1)) - \frac{B_n}{2} (Q_*^{(n)}(t))^2 (1 + o(1))$$

Отсюда

$$\frac{b_n c_{n+1}}{t^\alpha}(1 + o(1)) = \frac{B_n}{2} (Q_*^{(n)}(t))^2 (1 + o(1))$$

Тогда для  $Q_*^{(n)}(t)$  получаем оценку

$$Q_*^{(n)}(t) = \sqrt{\frac{2b_n}{B_n} c_{n+1} \frac{1}{t^\alpha} (1 + o(1))} = \sqrt{\frac{2b_n}{B_n} k_{n+1} \cdot t^{-\frac{\alpha}{2}} (1 + o(1))}$$

При этом, полагая  $c_n = \sqrt{\frac{2b_n}{B_n} c_{n+1}}$ , мы остаёмся в рамках предположения индукции.

Учитывая (37), можем записать асимптотику  $Q_*^{(n)}(t)$  для произвольной группы  $n$ :

$$\begin{aligned} Q_*^{(n)}(t) &= \sqrt{\frac{2b_n}{B_n} \sqrt{\frac{2b_{n+1}}{B_{n+1}} \dots \sqrt{\frac{2b_{w-1}}{B_{w-1} B_w} \cdot t^{-\left(\frac{1}{2}\right)^{w-n}}}}} = \\ &= \prod_{k=n}^{w-1} \left(\frac{2b_k}{B_k}\right)^{\left(\frac{1}{2}\right)^{w-n+1}} \cdot \left(\frac{1}{B_w}\right)^{\left(\frac{1}{2}\right)^{w-n}} \cdot t^{-\left(\frac{1}{2}\right)^{w-n}} \end{aligned}$$

Учитывая (33), получаем

$$\begin{aligned} Q_i(t) &= c_n U_j^{(n)} t^{-\left(\frac{1}{2}\right)^{w-n}} \cdot (1 + o(1)) \\ P_i(t) &= \tilde{c}_n U_j^{(n)} t^{-1 - \left(\frac{1}{2}\right)^{w-n}} \cdot (1 + o(1)) \end{aligned}$$

где нетерминал  $A_i$  находится в последнем критическом классе цепочки или в одном из предшествующих классов,  $n$  — номер группы, в которую входит класс, содержащий  $A_i$ ,  $w$  — число групп, и

$$c_n = \prod_{k=n}^{w-1} \left(\frac{2b_k}{B_k}\right)^{\left(\frac{1}{2}\right)^{w-n+1}} \cdot \left(\frac{1}{B_w}\right)^{\left(\frac{1}{2}\right)^{w-n}}$$

### 3 Математические ожидания числа применений правила в деревьях вывода

Обозначим через  $q_{ij}^l(t, \tau)$  и  $\bar{q}_{ij}^l(t, \tau)$  случайные величины, равные числу применений правила  $r_{ij}$  в дереве вывода, соответственно, из  $D_l^t$  и  $D_l^{\leq t}$ , на ярусе  $\tau$ . Пусть также

$$S_{ij}^l(t) = \sum_{\tau=1}^{t-1} q_{ij}^l(t, \tau)$$

$$\bar{S}_{ij}^l(t) = \sum_{\tau=1}^{t-1} \bar{q}_{ij}^l(t, \tau)$$

и  $S_{ij}^l(t)$ ,  $\bar{S}_{ij}^l(t)$  — соответственно число применений правила  $r_{ij}$  в дереве из  $D_l^t$ ,  $D_l^{\leq t}$ . Для удобства записи положим

$$S_{ij}(t) = S_{ij}^l(t), \quad \bar{S}_{ij}(t) = \bar{S}_{ij}^l(t),$$

$$q_{ij}(t, \tau) = q_{ij}^l(t, \tau), \quad \bar{q}_{ij}(t, \tau) = \bar{q}_{ij}^l(t, \tau)$$

Рассмотрим математические ожидания некоторых из введённых величин. Обозначим

$$M_{ij}^l(t) = M[S_{ij}^l(t)], \quad \bar{M}_{ij}^l(t) = [\bar{S}_{ij}^l(t)].$$

Для нахождения величин  $\bar{M}_{ij}^l(t)$  и  $M_{ij}^l(t)$  будут использованы следующие три леммы.

**Лемма 2** [5] Пусть  $s, d$  — натуральные числа,  $m = (m_1, \dots, m_s)$  — вектор целых неотрицательных чисел,  $y = (y_1, \dots, y_s)$  — вектор, и  $\bar{m} = \sum_{j=1}^s m_j$ . Тогда

$$(1 - y_1)^{n_1} \dots (1 - y_s)^{n_s} = \sum_{\substack{\bar{m} \leq d \\ m \geq 0}} \binom{n_1}{m_1} \binom{n_2}{m_2} \dots \binom{n_s}{m_s} (-1)^{\bar{m}} y^m + R_d(n_1, \dots, n_s, y),$$

где  $y^m = y_1^{m_1} \dots y_s^{m_s}$ , и остаточный член представим в виде

$$R_d(n_1, \dots, n_s, y) = \sum_{\substack{\bar{m}=d \\ m \geq 0}} (-1)^d \varepsilon_m(n_1, \dots, n_s, y) y^m,$$

причём

$$0 \leq \varepsilon_m(n_1, \dots, n_s, y') \leq \varepsilon_m(n_1, \dots, n_s, y) \leq \binom{n_1}{m_1} \dots \binom{n_s}{m_s}$$

при  $0 \leq y_i \leq y'_i \leq 1$  ( $i = 1, \dots, s$ ).

**Лемма 3** Пусть  $A(t)$  — последовательность матриц размером  $k \times k$ , и  $A(t) \rightarrow A$  при  $t \rightarrow \infty$ , причём  $A > 0$ , и её перронов корень  $r = 1$ . Пусть  $b(t) = bt^\alpha(1 + o(1))$  — последовательность векторов длины  $k$ , где  $b \geq 0$ ,  $b \neq 0$  и  $\alpha$  — действительное число. Тогда для последовательности векторов  $x(t)$  при  $t = 1, 2, \dots$ , определяемой

рекуррентным соотношением  $x(t) = b(t) + A(t)x(t-1)$  при  $t \rightarrow \infty$  справедливо соотношение

$$\frac{x_i(t)}{vx(t)} \rightarrow u_i,$$

при условии что  $x(t_0) > 0$  для некоторого номера  $t_0$ , где  $u, v > 0$  — соответственно правый и левый собственные векторы матрицы  $A$  при нормировке  $vu = 1$ .

Доказательство леммы принадлежит А. Борисову.

**Лемма 4** Пусть последовательность  $x_t$ ,  $x_t > 0$  при любом  $t \geq 0$ , удовлетворяет рекуррентному соотношению

$$x_{t+1} = \alpha t^\alpha (1 + \varepsilon_1(t)) + (1 - bt^\beta (1 + \varepsilon_2(t)))x_t,$$

где  $\beta < 0$ ,  $b > 0$ , и  $\varepsilon_1(t), \varepsilon_2(t) = o(1)$  при  $t \rightarrow \infty$ . Тогда верны следующие асимптотические равенства:

$$(1) \quad x_t = \frac{\alpha t^{\alpha+1}}{\alpha + 1} (1 + o(1)) \quad \text{при} \quad \beta < -1, \alpha \geq 0 \quad (39)$$

$$(2) \quad x_t = \frac{\alpha t^{\alpha+1}}{\alpha + b + 1} (1 + o(1)) \quad \text{при} \quad \beta = -1, \alpha > -1 \quad (40)$$

$$(3) \quad x_t = \frac{\alpha t^{\alpha-\beta}}{b} (1 + o(1)) \quad \text{при} \quad -1 < \beta < 0 \quad (41)$$

Доказательство леммы принадлежит А. Борисову.

Вначале рассмотрим  $\overline{M}_{ij}^q(t)$ . Пусть  $p(\cdot)$  — вероятность дерева  $d$  в грамматике  $G$ . Рассмотрим множество  $D_{ql}^{\leq t}$  деревьев из  $D_q^{\leq t}$ , первый ярус которых получен применением правила  $r_{ql}$  к корню дерева. Пусть

$$\overline{P}_{ql}^{ij}(t) = \sum_{d \in D_{ql}^{\leq t}} p(d) q_{ij}(d),$$

где  $q_{ij}(d)$  — число применений правила  $r_{ij}$  в дереве  $d$ , и  $\overline{P}_{ql}^{ij}(t)$  — вклад деревьев из  $D_{ql}^{\leq t}$  в матожидание  $\overline{M}_{ij}^q(t)$ . Для краткости, обозначим  $\overline{P}_{ql} = \overline{P}_{ql}^{ij}$ . Тогда

$$\overline{M}_{ij}^q(t) = \sum_{l=1}^{n_q} \overline{P}_{ql}(t). \quad (42)$$

Рассмотрим величину  $\overline{P}_{ql}(t)$ . Пусть

$$q_{ij}(d) = q_{ij}^{(1)}(d) + q_{ij}^{(2)}(d),$$

где  $q_{ij}^{(1)}(d)$  — число применений правила  $r_{ij}$  в дереве  $d$  на первом его ярусе, а  $q_{ij}^{(2)}(d)$  — на остальных ярусах. Тогда

$$\overline{P}_{ql}(t) = \sum_{d \in D_{ql}^{\leq t}} p(d) q_{ij}(d) = \sum_{d \in D_{ql}^{\leq t}} p(d) q_{ij}^{(1)}(d) + \sum_{d \in D_{ql}^{\leq t}} p(d) q_{ij}^{(2)}(d) = \overline{P}_{ql}^{(1)} + \overline{P}_{ql}^{(2)}(t)$$

Очевидно,  $q_{ij}^{(1)}(d) = \delta_i^q \delta_j^l$  (где  $\delta$  — символ Кронекера). Тогда

$$\bar{P}_{ql}^{(1)}(t) = \delta_i^q \delta_j^l \frac{p_{ij} Q_{s_{ij}}(t-1)}{1 - Q_q(t)},$$

где  $Q_X(t)$  — вероятность наборов деревьев вывода высоты не превосходящей  $t-1$ , набор корней которых задан характеристическим вектором  $X \in \mathbb{N}^k$ .

Обозначим также  $\delta^i(n) = (\delta_k^i)|_{i=\overline{1,n}} \in 0, 1^n$ .

Вероятность дерева  $p(d)$  при  $d \in D_{ql}^{\leq t}$  можно выразить как

$$p(d) = \frac{p_{ql}}{1 - Q_q(t)} p_1(d) p_2(d) \dots p_{\bar{s}_{ql}}(d),$$

где  $p_j(d)$  — вероятность поддерева  $d$  с корнем в  $j$ -м узле первого яруса. Тогда

$$\bar{P}_{ql}^{(2)}(t) = \frac{p_{ql}}{1 - Q_q(t)} \sum_{d \in D_{ql}^{\leq t}} \prod_{n=1}^{\bar{s}_{ql}} p_n(d) \sum_{m=1}^{\bar{s}_{ql}} q_{ij}^{\prime m}(d),$$

где  $q_{ij}^{\prime m}(d)$  — число применений правила  $r_{ij}$  в поддереве дерева  $d$  с корнем в  $m$ -том нетерминале первого яруса.

Выделим в  $d$  поддеревья  $d_1, d_2, \dots, d_{\bar{s}_{ql}}$ , где  $d_j$  — поддерево с корнем в  $j$ -м узле первого яруса дерева  $d$ . Тогда

$$\begin{aligned} \bar{P}_{ql}^{(2)}(t) &= \frac{p_{ql}}{1 - Q_q(t)} \sum_{m=1}^{\bar{s}_{ql}} \sum_{d \in D_{ql}^{\leq t}} \left( \prod_{n=1}^{\bar{s}_{ql}} p_n(d) \right) q_{ij}^{\prime m}(d) = \\ &= \frac{p_{ql}}{1 - Q_q(t)} \sum_{m=1}^{\bar{s}_{ql}} \sum_{d_j: j \neq m} p_1(d) \dots p_{m-1}(d_{m-1}) p_{m+1}(d_{m+1}) \dots p_{\bar{s}_{ql}}(d_{\bar{s}_{ql}}) q_{ij}^{\prime m}(d) = \\ &= \frac{p_{ql}}{1 - Q_q(t)} \sum_{m=1}^{\bar{s}_{ql}} Q_{s_{ql}-\delta^m} q_{ij}(d_m) = \frac{p_{ql}}{1 - Q_q(t)} \sum_{m=1}^k s_{ql}^m \bar{M}_{ij}^m(t-1) Q_{s_{ql}-\delta^m}(t-1) \end{aligned}$$

Зная  $\bar{P}_{ql}(t) = \bar{P}_{ql}^{(1)}(t) + \bar{P}_{ql}^{(2)}(t)$ , получаем

$$\bar{M}_{ij}^q(t) = \frac{1}{1 - Q_q(t)} \left( \delta_i^q p_{ij} Q_{s_{ij}}(t-1) + \sum_{l=1}^{n_q} p_{ql} \sum_{m=1}^k s_{ql}^m \bar{M}_{ij}^m(t-1) Q_{s_{ql}-\delta^m}(t-1) \right)$$

Обозначая

$$\bar{M}_{ij}^{\prime q}(t) = \bar{M}_{ij}^q(t)(1 - Q_q(t)),$$

имеем

$$\bar{M}_{ij}^{\prime q}(t) = \delta_i^q p_{ij} Q_{s_{ij}}(t-1) + \sum_{l=1}^{n_q} p_{ql} \sum_{m=1}^k s_{ql}^m \bar{M}_{ij}^{\prime m}(t-1) Q_{s_{ql}-\delta^m}(t-1) \quad (43)$$

Рекуррентное соотношение (43) является опорной точкой для вычисления  $\bar{M}_{ij}^q(t)$ . Получим аналогичное уравнение для  $M_{ij}^q(t)$ .

Аналогично (42)

$$M_{ij}^q(t) = \sum_{l=1}^{n_q} P_{ql}(t),$$

где  $P_{ql}(t)$  — вклад деревьев из  $D_{ql}^t$  в матожидание  $M_{ij}^q(t)$ . Положим  $P_{ql}(t) = P_{ql}^{(1)}(t) + P_{ql}^{(2)}(t)$ , аналогично тому, как это сделано для  $\bar{P}_{ql}(t)$ . При этом

$$P_{ql}^{(1)}(t) = \delta_i^q \delta_j^l \frac{p_{ij} R_{s_{ij}}(t-1)}{P_q(t)},$$

где  $R_X(t)$  — вероятность наборов деревьев из  $D^{\leq t}$ , набор корней которых задан характеристическим вектором  $X$ , и высота хотя бы одного из которых достигает  $t-1$ .  $P_{ql}^{(2)}(t)$  можно представить в виде

$$P_{ql}^{(2)}(t) = \sum_{m=1}^{\bar{s}_{ql}} P_{ql}^{(2)m}(t),$$

где  $P_{ql}^{(2)m}(t)$  — вклад деревьев с  $m$ -м корнем на первом ярусе в  $M_{ij}^q(t)$ .

Обозначим через  $S_1$  вклад в  $P_{ql}^{(2)m}(t)$  наборов деревьев, в которых ярус  $t$  достигается деревом с корнем в  $m$ -м нетерминале первого яруса. Очевидно,

$$S_1 = \frac{(1 - Q_{z_m}(t-1)) Q_{s_{ql}-\delta z_m}(t-1) M_{ij}^{z_m}(t-1)}{P_q(t)},$$

где  $z_m$  —  $m$ -й нетерминал первого яруса.

Пусть  $S_2$  — вклад наборов, где ярус  $t$  достигается через другие деревья. Тогда

$$S_2 = \frac{(1 - Q_{z_m}(t-1)) R_{s_{ql}-\delta m}(t-1) \bar{M}_{ij}^{z_m}(t-1)}{P_q(t)}$$

В результате, для  $M_{ij}^q(t)$  получаем

$$\begin{aligned} M_{ij}^q(t) &= \sum_{l=1}^{n_q} \left( P_{ql}^{(1)}(t) + \sum_{m=1}^{\bar{s}_{ql}} P_{ql}^{(2)m}(t) \right) = \\ &= \frac{1}{P_q(t)} [\delta_i^q p_{ij} R_{s_{ij}}(t-1) + \sum_{l=1}^{n_q} p_{ql} \sum_{m=1}^k (P_m(t-1) Q_{s_{ql}-\delta m}(t-1) M_{ij}^m(t-1) + \\ &\quad + (1 - Q_m(t-1)) R_{s_{ql}-\delta m}(t-1) \bar{M}_{ij}^m(t-1))] \end{aligned}$$

Из леммы (2) следуют выражения для  $Q_X(t)$  и  $R_X(t)$

$$\begin{aligned} Q_X(t) &= \prod_{i=1}^k (1 - Q_i(t))^{x_i} = 1 - \sum_{i=1}^k x_i Q_i(t) + \Theta \left( \sum_{1 \leq i, j \leq k} x_i x_j Q_i(t) Q_j(t) \right) \\ R_X(t) &= Q_X(t) - Q_X(t-1) = \sum_{i=1}^k x_i P_i(t) + \Theta \left( \sum_{1 \leq i, j \leq k} x_i x_j Q_i(t) Q_j(t) \right) \end{aligned} \tag{44}$$

Теперь приступим к вычислению  $\overline{M}_{ij}'^q(t)$  и  $M_{ij}'^q(t)$ .

Пусть вначале  $A_q$  и  $A_i$  принадлежат классам, находящимся в одной группе  $\mathcal{M}_n$ . Тогда  $\overline{M}_{ij}'^{(\alpha)}(t) = 0$  для всех  $A_\alpha \in K \in \mathcal{M}_m$ , таких что  $m > n$ . Для  $\overline{M}_{ij}'^{(n)}(t)$  получаем:

$$\overline{M}_{ij}'^q(t) = \delta_i^q p_{ij} Q_{s_{ij}}(t-1) + \sum_{l=1}^{n_q} p_{ql} \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} s_{ql}^\alpha \overline{M}_{ij}'^\alpha(t-1) Q_{s_{ql}-\delta^\alpha}(t-1)$$

Подставляя выражение (44) для  $Q_X(t)$ , получаем

$$\begin{aligned} \overline{M}_{ij}'^q(t) &= \delta_i^q p_{ij} (1 + o(1)) \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha \overline{M}_{ij}'^\alpha(t-1) - \\ &- \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha \overline{M}_{ij}'^\alpha(t-1) \sum_{\beta: A_\beta \in K \in \mathcal{M}_n} (s_{ql}^\beta - \delta_\beta^\alpha) Q_\beta(t-1) (1 + o(1)) \end{aligned} \quad (45)$$

Непосредственной проверкой устанавливается, что

$$\begin{aligned} a_\alpha^q &= \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha \\ b_{\alpha\beta}^q &= \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha (s_{ql}^\beta - \delta_\beta^\alpha) \end{aligned} \quad (46)$$

Заменяя соответствующие выражения в (45), а также подставляя асимптотику для  $Q^{(n)}(t)$ , получаем

$$\begin{aligned} \overline{M}_{ij}'^q(t) &= \delta_i^q p_{ij} (1 + o(1)) + \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} a_\alpha^q \overline{M}_{ij}'^\alpha(t-1) - \\ &- c_n t^{-\left(\frac{1}{2}\right)^{w-n}} \sum_{\alpha, \beta} b_{\alpha\beta}^q U_\beta \overline{M}_{ij}'^\alpha(t-1) (1 + o(1)) \end{aligned} \quad (47)$$

Применяя лемму (3) для вектора  $(\overline{M}_{ij}'^{q_1}(t), \overline{M}_{ij}'^{q_2}(t), \dots, \overline{M}_{ij}'^{q_\gamma}(t))$ , где  $A_{q_1}, A_{q_2}, \dots, A_{q_\gamma}$  — нетерминалы классов группы  $\mathcal{M}_n$ , имеем

$$\overline{M}_{ij}'^q(t) = U_q \sum_{l: A_l \in K \in \mathcal{M}_n} V_l \overline{M}_{ij}'^l(t) = U_q M_*^{(n)}(t).$$

Домножая (47) на  $V^{(n)}$  слева, получаем

$$\delta \overline{M}_*^{(n)}(t) = V_i p_{ij} (1 + o(1)) - c_n t^{-\left(\frac{1}{2}\right)^{w-n}} \sum_{q, \alpha, \beta} V_q b_{\alpha\beta}^q U_\alpha U_\beta = V_i p_{ij} - c_n t^{-\left(\frac{1}{2}\right)^{w-n}} B_n$$

Нетрудно видеть, что величина  $\overline{M}_*^{(n)}(t)$  удовлетворяет условиям леммы (4). Применяя её, получаем

$$\begin{aligned} \overline{M}_*^{(n)}(t) &= \left( \frac{V_i p_{ij}}{c_n B_n + 1} \right) \cdot t \cdot (1 + o(1)), & \text{если } n = w \\ \overline{M}_*^{(n)}(t) &= \left( \frac{V_i p_{ij}}{c_n B_n} \right) \cdot t^{-\left(\frac{1}{2}\right)^{w-n}} \cdot (1 + o(1)), & \text{если } n < w \end{aligned}$$

Пусть теперь классы, содержащие  $A_q$  и  $A_i$ , находятся в различных группах. Тогда

$$\overline{M}_{ij}'^q(t) = \delta_i^q p_{ij} Q_{s_{ij}}(t-1) + \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n \cup \mathcal{M}_{n+1}} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha \overline{M}_{ij}'^\alpha(t-1) Q_{s_{ql}-\delta^\alpha}(t-1)$$

Учитывая  $Q^{(n+1)}(t) = o(Q^{(n)}(t))$ , получаем

$$\begin{aligned} \overline{M}_{ij}'^q(t) &= O(p_{ij}) + \\ &+ \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha \overline{M}'^\alpha_{ij}(t-1) \left( 1 - \sum_{\beta: A_\beta \in K \in \mathcal{M}_n} (s_{ql}^\alpha - \delta_\beta^\alpha) Q_\beta(t-1) \right) (1 + o(1)) + \\ &+ \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_{n+1}} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha \overline{M}'^\alpha_{ij}(t-1) (1 + o(1)) \end{aligned}$$

Положим  $\overline{M}_{ij}'^\alpha(t-1) = \overline{d}'_{n+1} t^{\gamma(n+1)} (1 + o(1))$ , что выполняется для  $n+1 = w$ . Подставляя это выражение, а также (46), получаем

$$\begin{aligned} \overline{M}_{ij}'^q(t) &= \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} a_\alpha^q \overline{M}_{ij}'^\alpha(t-1) - c_n t^{-\left(\frac{1}{2}\right)^{w-n}} \sum_{\alpha, \beta} b_{\alpha\beta}^q U_\beta \overline{M}_{ij}'^\alpha(t-1) (1 + o(1)) + \\ &+ \overline{d}'_{n+1} t^{\gamma(n+1)} \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_{n+1}} a_\alpha^q U_\alpha (1 + o(1)) \end{aligned}$$

Домножая на  $V_q$ , имеем

$$\begin{aligned} \delta \overline{M}_*^{(n)}(t) &= \overline{d}'_{n+1} \left( \sum_{\substack{q: A_q \in K \in \mathcal{M}_n \\ \alpha: A_\alpha \in K \in \mathcal{M}_{n+1}}} V_q a_\alpha^q U_\alpha \right) \cdot t^{\gamma(n+1)} - \\ &- c_n t^{-\left(\frac{1}{2}\right)^{w-n}} \cdot \left( \sum_{q, \alpha, \beta} V_q b_{\alpha\beta}^q U_\alpha U_\beta \right) \cdot \overline{M}_*^{(n)}(t-1) (1 + o(1)) = \\ &= \overline{d}'_{n+1} b_{n+1} t^{\gamma(n+1)} (1 + o(1)) - c_n B_n t^{-\left(\frac{1}{2}\right)^{w-n}} \overline{M}_*^{(n)}(t-1) (1 + o(1)) \end{aligned}$$

Рассматривать случай  $n = w$  не имеет смысла, поэтому в выражении  $t^{-\left(\frac{1}{2}\right)^{w-n}}$  показатель всегда будет больше  $-1$ . Учитывая это, и применяя лемму (4), получаем

$$\overline{M}_*^{(n)}(t) = \frac{\overline{d}'_{n+1} b_{n+1} t^{\gamma(n+1) + \left(\frac{1}{2}\right)^{w-n}}}{c_n B_n}$$

Отсюда,

$$\overline{M}_*^{(n)}(t) = \prod_{j=n}^{h-1} \left( \frac{b_{j+1}}{c_j B_j} \right) \cdot \left( \frac{V_i p_{ij}}{c_h B_h + \delta_h^\alpha} \right) \cdot t^{\left(\left(\frac{1}{2}\right)^{\alpha-h-1} - \left(\frac{1}{2}\right)^{\alpha-n}\right)}$$

Подставляя (3) и  $\overline{M}_{ij}'^q(t) = \overline{M}_{ij}^q(t)(1 - Q_q(t))$ , получаем

$$\overline{M}_*^{(n)}(t) = \frac{U_q}{1 - Q_q(t)} \prod_{j=n}^{h-1} \left( \frac{b_{j+1}}{c_j B_j} \right) \cdot \left( \frac{V_i p_{ij}}{c_h B_h + \delta_h^\alpha} \right) \cdot t^{\left(\left(\frac{1}{2}\right)^{\alpha-h-1} - \left(\frac{1}{2}\right)^{\alpha-n}\right)}$$

Перейдём к вычислению  $M'_{ij}(t)$ . Вначале пусть нетерминалы  $A_q$  и  $A_i$  принадлежат классам из одной группы  $\mathcal{M}_n$ . Полагая  $M'_{ij}(t) = M'_{ij}(t)P_q(t)$ , получаем

$$\begin{aligned} M'_{ij}(t) = & O(t^{-1-(\frac{1}{2})}) + \sum_{\alpha: Q_\alpha \in K \in \mathcal{M}} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha M'_{ij}{}^\alpha(t-1) - \\ & - \sum_{\alpha, \beta: A_\alpha, A_\beta \text{ в группе } \mathcal{M}_n} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha (s_{ql}^\beta - \delta_\beta^\alpha) Q_n(t-1) M'_{ij}{}^\alpha(t-1) + \\ & + \sum_{\alpha, \beta} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha (s_{ql}^\beta - \delta_\beta^\alpha) P_n(t-1) \overline{M}'_{ij}{}^\alpha(t-1) \end{aligned}$$

Подставляя выражение (46) для первых и вторых моментов, получаем

$$\begin{aligned} M'_{ij}{}^q(t) = & \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} q_\alpha^q M'_{ij}{}^\alpha(t-1) - c_n t^{-(\frac{1}{2})^{w-n}} \sum_{\alpha, \beta} b_{\alpha\beta}^q U_\beta M'_{ij}{}^\alpha(t-1) (1 + o(1)) + \\ & + \tilde{c}_n t^{-1-(\frac{1}{2})^{w-n}} \sum_{\alpha, \beta} b_{\alpha\beta}^q U_\beta \overline{M}'_{ij}{}^\alpha(t-1) (1 + o(1)) \end{aligned}$$

Подставляя выражение для  $\overline{M}'_{ij}{}^\alpha(t-1)$ , имеем

$$\begin{aligned} M'_{ij}{}^q(t) = & \sum_{\alpha} a_\alpha^q M'_{ij}{}^\alpha(t-1) - c_n t^{-(\frac{1}{2})} \sum_{\alpha, \beta} b_{\alpha\beta}^q U_\beta M'_{ij}{}^\alpha(t-1) (1 + o(1)) + \\ & + \tilde{c}_n d_n t^{-1} \sum_{\alpha, \beta} b_{\alpha\beta}^q U_\alpha U_\beta (1 + o(1)) \quad (48) \end{aligned}$$

Применяя лемму (3), получаем

$$\begin{aligned} M'_{ij}(t) &= U_q M_*^{(n)}(t) (1 + o(1)) \\ M_*^{(n)}(t) &= \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} V_\alpha M'_{ij}{}^\alpha(t) \end{aligned}$$

Домножая (48) на  $V^{(n)}$ , получаем

$$\delta V_*^{(n)}(t) = \tilde{c}_n d_n B_n t^{-1} - c_n t^{-(\frac{1}{2})^{w-n}} B_n V_*^{(n)}(t) (t-1) (1 + o(1))$$

Применяя лемму (4), получаем в результате

$$M_*^{(n)}(t) = \begin{cases} \tilde{c}_n \overline{d}'_n B_n (1 + o(1)), & \text{при } \alpha = n \\ \frac{\tilde{c}_n \overline{d}'_n}{c_n} t^{-1-(\frac{1}{2})^{w-n}} (1 + o(1)), & \text{при } \alpha > n \end{cases}$$

Пусть теперь  $A_q$  и  $A_i$  находятся в классах, принадлежащих разным группам  $\mathcal{M}_n$  и  $\mathcal{M}_h$  ( $h > m$ ). Тогда

$$\begin{aligned} M'_{ij}{}^q(t) = & \delta_i^q p_{ij} R_{s_{ij}}(t-1) + \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n \cup \mathcal{M}_{n+1}} \sum_{l=1}^{n_q} p_{ql} s_{ql}^\alpha [Q_{s_{ql}}(t-1) M'_{ij}{}^\alpha(t-1) + \\ & + (1 - Q_\alpha(t-1)) R_{s_{ql}-\delta^\alpha}(t-1) \overline{M}'_{ij}{}^\alpha(t-1)] \end{aligned}$$



откуда

$$M'_{ij}{}^q(t) = \sum_{\alpha: A_\alpha \in K \in \mathcal{M}_n} a_\alpha^q M'_{ij}{}^\alpha(t-1) - \sum_{\alpha, \beta \text{ в группе } \mathcal{M}_n} b_{\alpha\beta}^q Q_\beta(t-1) M'_{ij}{}^\alpha(t-1)(1+o(1)) + \\ + \sum_{\alpha, \beta \text{ в группе } \mathcal{M}_n} b_{\alpha\beta}^q P_\beta(t-1) \bar{M}_{ij}{}^\alpha(t-1)(1+o(1))$$

Домножая на  $V^{(n)}$ , имеем

$$\delta M_*^{(n)}(t) = \tilde{c}_n \bar{d}_n B_n t^{-1 - (\frac{1}{2})^{w-n} + (\frac{1}{2})^{w-h} (2 - (\frac{1}{2})^{h-n})} (1+o(1)) - \\ - c_n B_n t^{-(\frac{1}{2})^{w-n}} M_*^{(n)}(t-1)(1+o(1))$$

Поскольку  $n < w$ , показатель степени в выражении  $t^{-(\frac{1}{2})^{w-n}}$  всегда больше  $-1$ , и по лемме (4) получаем

$$M_*^{(n)}(t) = \frac{\tilde{c}_n \bar{d}_n}{c_n} t^{-1 + (\frac{1}{2})^{w-h} (2 - (\frac{1}{2})^{h-n})}$$

Объединяя результаты для  $n < w$  и  $n = w$ , получаем

$$M_*^{(n)}(t) = \frac{\tilde{c}_n \bar{d}_n B_n}{\delta_w^n (c_n B_n - 1) + 1} \cdot t^{-1 + (\frac{1}{2})^{w-h} (2 - (\frac{1}{2})^{h-n})} (1+o(1))$$

Откуда

$$M_*^{(n)}(t) = \frac{U_q}{P_q(t)} \frac{\tilde{c}_n \bar{d}_n B_n}{\delta_w^n (c_n B_n - 1) + 1} \cdot t^{-1 + (\frac{1}{2})^{w-h} (2 - (\frac{1}{2})^{h-n})} (1+o(1))$$

## 4 Энтропия

Пусть  $L^t$  — множество слов языка  $L_G$ , порождаемых деревьями вывода из  $D^t$ . Будем рассматривать грамматики с однозначным выводом.

По определению, энтропия языка  $L^t$  есть

$$H(L^t) = - \sum_{\alpha \in L^t} p_t(\alpha) \log p_t(\alpha),$$

где  $p_t(\alpha) = p(\alpha : \alpha \in L^t) = p(\alpha)/p(L^t)$ . Используя это выражение для  $p_t(\alpha)$ , получаем

$$H(L^t) = - \frac{1}{P(L^t)} \sum_{\alpha \in L^t} p_t(\alpha) (\log p(\alpha) - \log P(L^t)) = \\ = \frac{\log P(L^t)}{P(L^t)} \sum_{\alpha \in L^t} p(\alpha) - \frac{1}{P(L^t)} \sum_{\alpha \in L^t} p_t(\alpha) \log p(\alpha) = \\ = \log P(L^t) - \frac{1}{P(L^t)} \sum_{\alpha \in L^t} p(\alpha) \log p(\alpha)$$

Выразим вероятность слова  $\alpha$  через вероятности правил вывода  $r_{ij}$ . Поскольку рассматривается грамматика с однозначным выводом, каждому слову  $\alpha$  из  $L^t$  соответствует единственное дерево  $d(\alpha)$  из  $D^t$  и единственный левый вывод  $\omega_l(\alpha) = (r_{i_1, j_1}, r_{i_2, j_2}, \dots, r_{i_s, j_s})$ . Получаем

$$p(\alpha) = p(r_{i_1, j_1}) \cdot \dots \cdot p(r_{i_s, j_s}) = \prod_{i=1}^k \prod_{j=1}^{n_i} p_{ij}^{q_{ij}(\alpha)},$$

где  $q_{ij}(\alpha)$  — число применений правила  $r_{ij}$  при выводе слова  $\alpha$  (учитывая единственность дерева вывода, это число определяется единственным образом). Тогда

$$\sum_{\alpha \in L^t} p(\alpha) \log p(\alpha) = \sum_{\alpha \in L^t} p(\alpha) \sum_{i=1}^k \sum_{j=1}^{n_i} q_{ij}(\alpha) \log p_{ij} = \sum_{i=1}^k \sum_{j=1}^{n_i} \log p_{ij} \sum_{\alpha \in L^t} q_{ij}(\alpha) p(\alpha)$$

Пользуясь определением  $M(S_{ij}(t))$ , получаем

$$\sum_{\alpha \in L^t} p(\alpha) \log p(\alpha) = \sum_{i=1}^k \sum_{j=1}^{n_i} \log p_{ij} M(S_{ij}(t)) P(L^t)$$

Отсюда

$$H(L^t) = \log P(L^t) - \sum_{i=1}^k \sum_{j=1}^{n_i} M(S_{ij}(t)) \log p_{ij} (1 + o(1))$$

По определению,  $P(L^t) = P_1(t) = O(t^{-1 - (\frac{1}{2})^{w-1}})$ , и  $\log P(L^t) = O(\log t)$ . Подставляя выражение для  $M(S_{ij}(t)) = M_{ij}(t)$ , получаем

$$H(L^t) = \sum_{i=1}^k \sum_{j=1}^{n_i} H(R_i) d_i t^2 (1 + o(1)),$$

где  $H(R_i) = -\sum_{j=1}^{n_i} p_{ij} \log p_{ij}$  — энтропия множества  $R_i$  правил вывода. Асимптотика  $t^2$  задаётся величиной  $M_{ij}^q(t)$  для последнего критического класса и классов, следующих за ними.

Сформулируем теорему:

**Теорема 2** Энтропия языка  $L^t$ , состоящего из слов, порождаемых в разложимой стохастической КС-грамматике вида «цепочки» с однозначным выводом деревьями высоты  $t$ , выражается формулой

$$H(L^t) \sum_{i \in I} \sum_{j=1}^{n_i} d_i H(R_i) \cdot t^2,$$

где  $d_i > 0$ ,  $H(R_i) = -\sum_{j=1}^{n_i} p_{ij} \log p_{ij}$  — энтропия множества  $R_i$  правил вывода с нетерминалов  $A_i$  в левой части, и  $I$  — множество индексов нетерминалов, содержащихся в последнем критическом классе, а также классах, следующих за ним.

## 5 Заключение

В результате проведённого исследования были изучены основные вероятностные характеристики грамматик заданного класса. Полученные асимптотические оценки позволяют непосредственно перейти к построению алгоритма асимптотически оптимального кодирования для рассматриваемого класса языков сообщений, а также существенно упрощают исследование этой задачи для КС-языков в общем случае.

## Список литературы

- [1] **Шеннон К.** Математическая теория связи. М.: ИЛ, 1963
- [2] **Марков А. А.** Введение в теорию кодирования. М.: Наука, 1982
- [3] **Фу К.** Структурные методы в распознавании образов. М.: Мир, 1977
- [4] **Ахо А., Ульман Дж.** Теория синтаксического анализа, перевода и компиляции. Том 1. М.: Мир, 1978
- [5] **Севастьянов Б. А.** Ветвящиеся процессы. — М.: Наука, 1971 — 436 с.
- [6] **Гантмахер Ф. Р.** Теория матриц. — 5-е изд., — М.: ФИЗМАТЛИТ, 2010
- [7] **Жильцова Л. П.** О матрице первых моментов разложимой стохастической КС-грамматики. УЧЁНЫЕ ЗАПИСКИ КАЗАНСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА, Том 151, кн. 2, 2009
- [8] **Жильцова Л. П.** Закономерности применения правил грамматики в выводах слов стохастического контекстно-свободного языка // Математические вопросы кибернетики. Выр. 9. М.: Наука, 2000. С. 100-126.
- [9] **Жильцова Л. П.** О нижней оценке стоимости кодирования и асимптотически оптимальном кодировании стохастического контекстно-свободного языка // Дискретный анализ и исследование операций. Серия 1, т. 8, №3. Новосибирск: Издательство Института математики СО РАН, 2001. С. 26-45.
- [10] **Борисов А. Е.** Закономерности в словах стохастических контекстно-свободных языков, порождённых грамматиками с двумя классами нетерминальных символов. Вопросы экономного кодирования.