Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина

На правах рукописи

Горбунова Ирина Анатольевна

ПОСТРОЕНИЕ ЭКСТРЕМАЛЬНЫХ БЕСПОВТОРНЫХ СЛОВ И ОЦЕНКА ИХ КОЛИЧЕСТВА

01.01.09 – Дискретная математика и математическая кибернетика

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
профессор, доктор физ.-мат. наук
А.М. Шур

Екатеринбург
2013

Оглавление

Введение

Изучение свойств символьных последовательностей (слов) и множеств таких последовательностей (формальных языков) составляет важное направление в современной дискретной математике, имеющее обширные приложения в различных разделах математики, компьютерных науках, биологии и других областях знания. Интерес к символьным последовательностям у математиков появился более ста лет назад. Тому были как «внутренние» причины (становление теории групп), так и «внешние», например, активное использование двоичного кода (азбука Морзе) для передачи сообщений посредством телеграфа и радио. Первым математиком, систематически изучавшим свойства символьных последовательностей, был норвежец А.Туэ, которого по праву считают основоположником комбинаторики слов.

Одной из центральных тем комбинаторики слов является изучение слов и языков, не содержащих *повторов* – идущих подряд одинаковых фрагментов. Понятие повтора легко проиллюстрировать и в естественных языках. Например, слово «мама» – это повтор, называемый *квадратом* (состоит из двух одинаковых фрагментов). Иногда само слово не является повтором, но повтор находится внутри него, как, например, в слове «банан» (фрагмент «ан» повторился 2 раза). Любой повтор характеризуется *экспонентой* – отношением длины слова к длине повторяющейся части. Так, слово «заноза» – это повтор с *экспонентой* $3/2$ (фрагмент «зано» повторился лишь до половины). Экспонента обобщает понятие показателя степени на дробные значения.[1]

Экспонента β называется *избегаемой* над данным алфавитом, если существует бесконечное слово (или, что эквивалентно, бесконечно много конечных слов) над данным алфавитом, не содержащее повторов с экспонентой, не меньшей β. Начало исследованиям в этой области также положил Туэ. Он доказал, что квадраты избегаемы над 3-буквенным алфавитом [50], а кубы и все дробные степени, большие 2, избегаемы над 2-буквенным алфавитом [51]. Результат Туэ для 2-буквенного алфавита неулучшаем, поскольку любое слово длины не меньше 4 в таком алфавите содержит квадрат. Однако для 3-буквенного алфавита найденное Туэ значение не было оптимальным: в 1972 году Ф.Дежан построила бесконечное слово над 3-буквенным

[1] Здесь «степень» относится к ассоциативной операции умножения (конкатенации) слов, т.е. (ма)^2 =мама.

алфавитом, которое не содержало повторов степени больше 7/4 [17].

Инфимум множества избегаемых экспонент для k-буквенного алфавита называется *границей повторяемости* и обозначается $\mathrm{RT}(k)$. Так, из работ Туэ следует, что $\mathrm{RT}(2) = 2$. Дежан доказала, что $\mathrm{RT}(3) = 7/4$, и высказала гипотезу о том, какие значения принимает $\mathrm{RT}(k)$ для $k \geq 4$ (см. Граничную теорему на стр. 11). Слова над k-буквенным алфавитом, не содержащие повторов с экспонентой больше $\mathrm{RT}(k)$, являются *экстремальными бесповторными словами*, они избегают любую экспоненту, избегаемую над данным алфавитом. Такие слова представляют наибольший интерес в классе бесповторных слов и требуют к себе особого внимания.

Предварительные сведения

В данном разделе приводятся основные определения комбинаторики слов, которые используются в последующих главах. Для более полной информации можно использовать, например, [7, 13, 28].

Слова

Под *алфавитом* Σ понимается непустое конечное множество, элементы которого называются *буквами* или *символами*. Через Σ_k будем обозначать произвольный k-буквенный алфавит.

Слово w – это конечная последовательность букв, т.е. $w = a_1 a_2 \cdots a_n$, где $a_1, a_2, \ldots, a_n \in \Sigma$, а $n = |w|$ – *длина* слова w. На слово w также можно смотреть как на функцию $w : \{1, \ldots, n\} \to \Sigma$. Числа из области определения этой функции называются *позициями*. Через $w[i]$, где $1 \leq i \leq |w|$, будем обозначать букву, стоящую на i-ой позиции в слове w. Слово длины 0 называется *пустым* и обозначается λ. Через $\Sigma(w)$ будем обозначать множество различных букв, входящих в слово w. Через Σ^* (Σ^+, Σ^n) обозначается множество всех слов (соответственно, всех непустых слов, всех слов длины n) над фиксированным алфавитом Σ.

Слово u называется *подсловом* (соответственно, *префиксом* или *суффиксом*) слова w, если w можно представить в виде $\bar{v}u\hat{v}$ (соответственно, $u\hat{v}$, $\bar{v}u$) для некоторых (возможно пустых) слов $\bar{v}$ и $\hat{v}$. Подслово (соответственно, префикс, суффикс) слова w называется *собственным*, если оно не совпадает с w. Для обозначения подслова слова w, начинающегося с i-ой позиции и заканчивающегося в j-ой позиции, будем использовать обозначение $w[i..j]$. *Реверсом* слова $w = a_1 a_2 \cdots a_n$ будем называть слово $\overleftarrow{w} = a_n \cdots a_2 a_1$.

Два слова w_1 и w_2 называются *сопряженными*, если найдутся такие непустые слова u и v, что $w_1 = uv$ и $w_2 = vu$ (иначе говоря, слова w_1 и w_2 получаются друг из друга циклическими сдвигами на определенное число позиций).

Периодом слова w называется любой период соответствующей функции. Минимальный период слова w будем обозначать $\operatorname{per}(w)$. *Экспонентой* слова w называется отношение длины слова к его минимальному периоду: $\exp(w) = |w|/\operatorname{per}(w)$. На экспоненту можно смотреть как на рациональный показатель степени. Например, слово $w = abcdabc$ имеет период 4 и $\exp(w) = 7/4$, т.е. $w = (abcd)^{7/4}$.

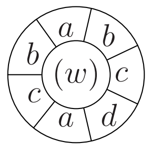
Помимо «обычных» мы будем использовать и другие виды слов.

ω-слово $\mathbf{w}$ – это бесконечное вправо слово, на которое можно смотреть как на функцию $\mathbf{w} : \mathbb{N} \to \Sigma$. *$Z$-слово* $\mathbf{w}$ – это бесконечное в обе стороны слово, т.е. $\mathbf{w} : \mathbb{Z} \to \Sigma$. Конечные подслова ω-слов и Z-слов являются «обычными» словами, следовательно, к ним применимы данные выше определения. Через Σ^ω (соответственно, Σ^Z) обозначается множество всех ω-слов (соответственно, всех Z-слов) над фиксированным алфавитом Σ.

Один из простейших способов построения бесконечных слов – это *метод итерации морфизма*. *Морфизмом* называется отображение $\phi : \Sigma^* \to \Delta^*$, где Σ и Δ – произвольные алфавиты, такое, что $\forall u, v \in \Sigma^*$ выполнено условие $\phi(uv) = \phi(u)\phi(v)$. Легко заметить, что любой морфизм однозначно определяется образами символов алфавита Σ. При этом, если длина каждого образа строго положительна, то морфизм называется *нестирающим*. Пусть $a \in \Sigma$, $\phi : \Sigma^* \to \Sigma^*$ – нестирающий морфизм и $\forall n \in \mathbb{N}_0$ $\phi^n(a)$ является собственным префиксом $\phi^{n+1}(a)$. Тогда можно определить предел последовательности $\left(\phi^n(a)\right)_{n \in \mathbb{N}_0}$ как ω-слово $\mathbf{w}_\phi$, у которого префикс длины $|\phi^n(a)|$ совпадает с $\phi^n(a)$ для любого $n \in \mathbb{N}_0$. В этом случае говорят, что ω-слово $\mathbf{w}_\phi$ получено итерацией морфизма ϕ. Полученное ω-слово $\mathbf{w}_\phi$ удовлетворяет равенству $\phi(\mathbf{w}_\phi) = \mathbf{w}_\phi$, поэтому его называют также *неподвижной точкой* морфизма ϕ.

Циклическое слово (w) получается, если у «обычного» слова w «склеить» начало и конец, т.е. на множестве позиций задать не линейный, а циклический порядок:

$$w = abcdacb$$



Циклическое слово (w) можно также ассоциировать с классом сопряженности (легко проверить, что сопряженность является отношением эквивалентности), в котором содержится слово w. Подсловами циклического слова (w) являются слово w и все его циклические сдвиги со всеми своими подсловами. Таким образом, подслова циклических слов – это «обычные» слова, длина которых не превосходит $|(w)| = |w|$.

Двумерное слово W размера $n \times q$ – это прямоугольная таблица размера $n \times q$ из символов алфавита Σ:

$$
\begin{array}{c|c|c|c|c|c|c|c|c|}
 & \multicolumn{8}{c}{q} \\
\cline{2-9}
 & b & a & b & a & b & a & b & a \\
\cline{2-9}
n & a & b & a & b & a & b & a & b \\
\cline{2-9}
 & b & a & b & a & b & a & b & a \\
\cline{2-9}
 & a & b & a & b & a & b & a & b \\
\cline{2-9}
\end{array}
$$

Подслова двумерного слова также являются двумерными словами (т.е. допускаются только «прямоугольные» подслова). В отличие от рассматриваемой в [21] модели двумерных слов, мы не будем использовать дополнительный символ для выделения границы двумерного слова. Через Σ^{**} обозначается множество всех двумерных слов над фиксированным алфавитом Σ.

Граница повторяемости

Пусть $\beta > 1$ – действительное число. Слово w $\big[\omega$-слово $\mathbf{w}$, Z-слово $\mathbf{w}$, циклическое слово $(w)\big]$ называется *β-свободным* (соответственно, *β^+-свободным*), если все его подслова имеют экспоненту меньше β (соответственно, не больше β). Например, слово $w = abcdabc$ является $(7/4)^+$-свободным, но не $7/4$-свободным (так как $\exp(w) = 7/4$), а слово $u = ababc$ является 3-свободным и даже 2^+-свободным, но не 2-свободным (так как содержит подслово $abab$ с экспонентой 2).

Экспонента β называется *избегаемой* над алфавитом Σ_k, если существует бесконечно много β-свободных слов над алфавитом Σ_k. В противном случае экспонента называется *неизбежной* над Σ_k. *Границей повторяемости* $\mathrm{RT}(k)$ называется действительное число, отделяющее избегаемые экспоненты над алфавитом Σ_k от неизбежных:

$$\mathrm{RT}(k) = \inf\{\beta \in \mathbb{R} \mid \beta - \text{ избегаемая экспонента над } \Sigma_k\} \tag{1}$$

Легко проверить, что сама экспонента $\mathrm{RT}(k)$ неизбежна над Σ_k, т.е. в алфавите Σ_k есть бесконечно много $\mathrm{RT}(k)^+$-свободных слов, но лишь конечное множество $\mathrm{RT}(k)$-свободных.

Границу повторяемости можно определить несколькими эквивалентными способами, заменив в формуле (1) условие «β – избегаемая экспонента над Σ_k» на одно из следующих условий:

(W) существует бесконечно много β-свободных слов над алфавитом Σ_k;

(I) существуют β-свободные слова над алфавитом Σ_k любой достаточно большой длины;

(S) существуют β-свободные слова над алфавитом Σ_k любой длины.

Однако при замене «обычных» слов на циклические эти три условия становятся попарно неэквивалентными, и соответствующие определения приобретают разный смысл. Действительно, если у нас есть β-свободное слово w над алфавитом Σ_k, то и любое его подслово является β-свободным. Но если взять циклическое β-свободное слово (w), то любое его подслово является β-свободным словом, но не β-свободным циклическим словом. Так слово $(abcdacb)$ – циклическое $(3/2)^+$-свободное слово, $bcdacb$ – его $(3/2)^+$-свободное подслово, однако $(bcdacb)$ не является циклическим $(3/2)^+$-свободным словом, так как содержит подслово bb с экспонентой 2. Поэтому при определении границы повторяемости для циклических слов мы получим три ее разновидности: *слабую* $\mathrm{CRT}_W(k)$, *среднюю* $\mathrm{CRT}_I(k)$ и *сильную* $\mathrm{CRT}_S(k)$, соответственно.

В нашей работе исследуется только сильная граница повторяемости для циклических слов, поэтому будем называть ее просто циклической границей повторяемости и обозначать $\mathrm{CRT}(k)$.

$\mathrm{RT}(k)^+$-свободные слова и $\mathrm{CRT}(k)^+$-свободные циклические слова являются «экстремальными» бесповторными словами и представляют особый интерес среди слов, избегающих повторы. $\mathrm{RT}(k)^+$-свободные слова [$\mathrm{RT}(k)^+$-свободные Z-слова] будем называть *граничными*.

Языки, антисловари, автоматы

Любое подмножество Σ^* (Σ^{**}) называется *языком* (соответственно, *двумерным языком*) над алфавитом Σ. Язык (двумерный язык) называется *факториальным*, если он замкнут относительно взятия подслов, т.е. вместе с любым словом содержит все его подслова, и *антифакториальным*, если никакое его слово не является подсловом другого слова.

Языки, состоящие из всех β-свободных (β^+-свободных) слов, будем называть β-свободными (β^+-свободными) языками над данным алфавитом или *языками ограниченной экспоненты*. Заметим, что такие языки являются факториальными. $\mathrm{RT}(k)^+$-свободные языки будем называть *граничными* и обозначать T_k.

Слово w (двумерное слово W) *запрещено* в языке (соответственно, в двумерном языке) L, если оно не является подсловом ни одного элемента L (иначе говорят, что язык L *избегает* слово w). Множество всех минимальных в отношении подсловного порядка запрещенных слов для языка (двумерного языка) L называется *антисловарем* языка L. Антисловарь всегда является антифакториальным языком. Антисловарь граничного языка T_k будем обозначать через A_k.

Конечным автоматом называется пятерка объектов $\mathcal{A} = (Q, \Sigma, \delta, S, T)$, где Q - конечное множество состояний с подмножествами начальных состояний S и терминальных состояний T, Σ - конечный алфавит, а $\delta \subset Q \times \Sigma \times Q$ - множество переходов. Конечный автомат обычно рассматривается как помеченный орграф с множеством вершин Q и множеством помеченных дуг δ. Автомат A *распознает слово* w, если в нем для некоторых $s \in S$ и $t \in T$ существует ориентированный (s,t)-маршрут, помеченный w. Множество всех слов, распознаваемых автоматом A, образует язык $L(A)$, *распознаваемый* этим автоматом. Языки, распознаваемые конечными автоматами, образуют в точности класс *регулярных* языков. Конечный автомат называется *детерминированным*, если в нем одно начальное состояние, и все переходы из любого фиксированного состояния имеют различные метки.

Факториальный язык является регулярным тогда и только тогда, когда регулярен его антисловарь. В частности, языки с конечным антисловарем образуют подкласс регулярных языков.

Бесконечные языки ограниченной экспоненты нерегулярны (в частности, антисловари A_k бесконечны). Действительно, достаточно длинные маршруты в автомате содержат циклы; заменив однократное прохождение цикла многократным, можно получить слово, распознаваемое автоматом, имеющее подслово сколь угодно большой экспоненты.

Минимальные запрещенные слова для граничного языка T_k при $k \geq 3$ имеют вид $u = ptp$, где $|pt|$ – это период слова u, а $\exp(u) = |u|/|pt| > \mathrm{RT}(k)$, причем экспонента любого собственного подслова слова u не превосходит $\mathrm{RT}(k)$. Если $|p| = m$, то такое запрещенное слово u будем называть *m-повтором*. Например, слово $abcdeab$ над алфавитом Σ_5 является 2-повтором.

Через $A_k^{(m)}$ будем обозначать подмножество в A_k, состоящее из всех r-повторов для любого $r \leq m$. Заметим, что множество $A_k^{(m)}$ является конечным для любого $m \in \mathbb{N}$. Кроме того, выполнена следующая цепочка вложений: $A_k^{(1)} \subseteq \cdots \subseteq A_k^{(m)} \subseteq \cdots \subset A_k$. Языки с антисловарями $A_k^{(m)}$ будем обозначать $T_k^{(m)}$. Таким образом, для каждого граничного языка мы получим приближающую его последовательность регулярных языков: $T_k \subset \cdots \subseteq T_k^{(m)} \subseteq \cdots \subseteq T_k^{(1)}$.

Количественные характеристики языков

Комбинаторной сложностью языка L называется функция $C_L(n)$, которая возвращает количество слов длины n в языке L. Комбинаторная сложность ω-слова или Z-слова есть комбинаторная сложность языка его подслов. Впервые комбинаторная сложность рассматривалась в [32]; с тех пор ее изучению было посвящено множество работ. О скорости роста комбинаторной сложности языка L можно судить по *индексу роста*, который задается формулой

$$\alpha(L) = \overline{\lim_{n\to\infty}} (C_L(n))^{1/n} \qquad (2)$$

Если $\alpha(L) > 1$, то это свидетельствует об экспоненциальном росте функции $C_L(n)$, значение $\alpha(L) = 1$ говорит о «медленном» (субэкспоненциальном) росте комбинаторной сложности, а $\alpha(L) = 0$ означает, что язык L конечен. Так как комбинаторная сложность принимает целые значения, случай $0 < \alpha(L) < 1$ очевидным образом невозможен.

Если язык L факториальный, то для его комбинаторной сложности выполняется неравенство $C_L(n_1+n_2) \leq C_L(n_1)C_L(n_2)$ для любых $n_1, n_2 \in \mathbb{N}$. А значит, можно воспользоваться следующей леммой, доказательство которой можно найти в [19, 53]:

Лемма Фекете. *Пусть $(a_n)_{n\geq 1}$ – последовательность действительных чисел такая, что для любых $n_1, n_2 \in \mathbb{N}$ выполнено неравенство $a_{n_1+n_2} \leq a_{n_1}+a_{n_2}$. Тогда предел $\lim\limits_{n\to\infty} \frac{a_n}{n}$ существует и равен $\inf\limits_{n\in\mathbb{N}} \frac{a_n}{n}$.*

Применяя данную лемму к последовательности $\big(\log C_L(n)\big)_{n\geq 1}$, получим существование предела $\lim\limits_{n\to\infty} \frac{\log C_L(n)}{n}$ (следовательно, и $\lim\limits_{n\to\infty} (C_L(n))^{1/n}$). А значит, индекс роста факториального языка можно определить следующим образом:

$$\alpha(L) = \lim_{n\to\infty} (C_L(n))^{1/n} \tag{3}$$

Аналогичные понятия можно ввести и для двумерных языков (подробнее см. [26]). Так, комбинаторную сложность двумерного языка L можно определить как функцию $C(n, q)$, возвращающую количество слов размера $n\times q$ в языке L. А индекс роста факториального двумерного языка – как двойной предел

$$\alpha(L) = \lim_{n\to\infty, q\to\infty} (C_L(n, q))^{1/nq} \tag{4}$$

Аналогично одномерному случаю, существование предела (4) следует из многомерного аналога леммы Фекете, доказанного в [10].

Хорошо известно, что индекс роста любого регулярного языка можно вычислить с помощью детерминированного конечного автомата, распознающего этот язык. Автомат должен удовлетворять такому условию: любая вершина принадлежит некоторому маршруту из начальной вершины в терминальную. При этом условии, индекс роста языка является максимальным положительным корнем характеристического многочлена матрицы смежности этого автомата (этот результат, по-видимому, является фольклорным; короткое доказательство можно найти в [46]). Эффективный как с теоретической, так и с практической точки зрения алгоритм для нахождения индекса роста регулярных языков описан в [45, 47]. Стоит отметить, что распознающий автомат с требуемыми свойствами для языков с конечными антисловарями быстро строится по алгоритму [14]. К сожалению, для двумерных языков не известно эффективных алгоритмов вычисления или даже оценки сверху индекса роста.

Обзор исследований по теме диссертации

Впервые понятие бесповторной символьной последовательности или бесповторного слова появилось в работах норвежского математика А.Туэ. В 1906 году он построил бесквадратное ω-слово над тернарным алфавитом [50]. А в 1912 году – бинарное ω-слово, избегающее кубы и все дробные степени, большие 2 [51]. Бесконечные слова Туэ строил с помощью метода итерации морфизма. Так, в случае бинарного алфавита он использовал морфизм

$$\theta \; : \; \begin{array}{l} 0 \mapsto 01 \\ 1 \mapsto 10 \end{array}$$

Морфизм θ имеет неподвижную точку $\mathbf{w}_\theta = \theta^\infty(0) = 0110100110010110\cdots$ (или $\bar{\mathbf{w}}_\theta = \theta^\infty(1) = 1001011001101001\cdots$, если начинать с 1), которая и является 2^+-свободным ω-словом.

К сожалению, обе работы Туэ были напечатаны в малоизвестном норвежском журнале, и не были известны специалистам до конца 1950-х годов. Поэтому многие из них были открыты и доказаны заново. Так, например, то же самое бинарное слово Морс построил в 1921 году [31], Эйве – в 1929 [18], Аршон – в 1937 [1] (на самом деле, еще в 1851 году в работе по теории чисел [37] Пруэ описал разбиение натурального ряда на два множества, неявно задававшее то же самое слово). Впоследствии ω-слово $\mathbf{w}_\theta$ получило название «слово Туэ-Морса».

Слово Туэ-Морса является граничным словом, в то время как все построенные Туэ бесквадратные слова над тернарным алфавитом граничными не являются. Это следует из результата Ф.Дежан 1972 года [17]. Дежан построила с помощью итерации морфизма тернарное $(7/4)^+$-свободное ω-слово и доказала, что никакое тернарное слово длины более 39 не является $(7/4)$-свободным, т.е. $\mathrm{RT}(3) = 7/4$. В той же работе Дежан доказала нижнюю оценку для избегаемости экспонент в k-буквенных алфавитах при $k \geq 4$ и предположила, что граница повторяемости совпадает с этой оценкой. Данное предположение получило название «гипотезы Дежан», а после завершения его доказательства – «граничной теоремы».

Граничная теорема (гипотеза Дежан; [11, 15–17, 30, 33, 36, 38, 51])**.** *Экспонента β избегаема над алфавитом Σ_k тогда и только тогда, когда $\beta > \mathrm{RT}(k)$, где $\mathrm{RT}(k)$ принимает следующие значения:*

k	2	3	4	5	6	$\cdots$	i	$\cdots$
$\mathrm{RT}(k)$	2	$7/4$	$7/5$	$5/4$	$6/5$	$\cdots$	$i/(i-1)$	$\cdots$

Доказательство данной гипотезы заняло целых 37 лет (см. рис. 1). Ситуация осложнилась тем, что методом итерации морфизма (как это делали Туэ и Дежан) нельзя построить $\mathrm{RT}(k)^+$-свободное ω-слово для $k \geq 4$ (см. [9]). Решение нашел

Пансьё. Он предложил идею бинарного кодирования граничных слов, о котором более подробно речь пойдет в пункте 1.1.1. Тем самым, он свел исходную задачу построения $\mathrm{RT}(k)^+$-свободного ω-слова над k-буквенным алфавитом к построению кодового ω-слова над бинарным алфавитом, которое можно было получить методом итерации морфизма. Пансьё удалось построить подходящий бинарный морфизм для случая $k = 4$ [36].

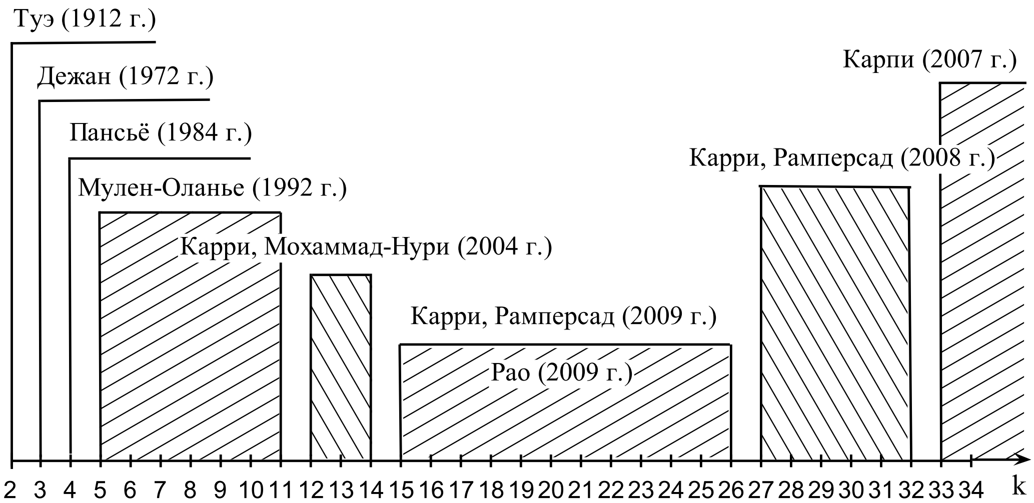


Рис. 1. *Доказательство гипотезы Дежан.*

Развивая идею Пансьё, Мулен-Оланье установил достаточные условия для того, чтобы слово, порожденное бинарным морфизмом, кодировало граничное слово. С помощью компьютера он подобрал бинарные морфизмы, удовлетворяющие найденным условиям, для $5 \leq k \leq 11$ [33].

Мохаммад-Нури и Карри, также прибегнув к помощи компьютера, построили бинарные морфизмы для $12 \leq k \leq 14$ [30]. Вместо проверки морфизмов по критерию Мулена-Оланье они использовали для генерации кодов граничных слов *слова Штурма* (ω-слово называется словом Штурма, если содержит ровно $n+1$ различное подслово длины n для любого $n \in \mathbb{N}$).

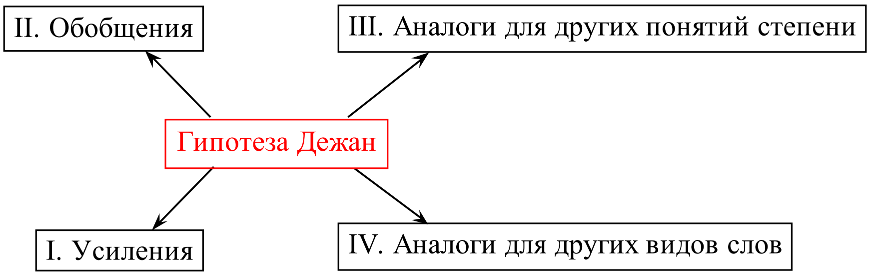
Прорыв в доказательстве совершил Карпи, ему удалось получить бескомпьютерное доказательство для $k \geq 33$ [11]. Его метод включал в себя работу с некоторым вспомогательным алфавитом: сначала строилось ω-слово над вспомогательным алфавитом Σ_m, где $m = \lfloor (k-3)/6 \rfloor$, которое с помощью специального морфизма переводилось в бинарное кодовое ω-слово некоторого граничного слова над алфавитом Σ_k. Карпи удалось найти универсальную конструкцию для вспомогательных слов при $m \geq 5$.

Прибегнув к помощи компьютера, Карри и Рамперсад смогли построить вспомогательные слова из конструкции Карпи для $m = 4$, покрыв тем самым случаи

$27 \le k \le 32$ [15]. А несколько модернизировав конструкцию Мулена-Оланье, Карри и Рамперсад смогли с помощью компьютера подобрать бинарные морфизмы для оставшегося интервала $15 \le k \le 26$ [16].

Примерно в то же время доказательство гипотезы завершил Рао [38], который искал кодовые бинарные слова в виде образов слова Туэ-Морса при подходящих морфизмах.

За долгие годы доказательства вокруг гипотезы Дежан возник целый круг «родственных» задач. Их можно условно разбить на четыре основные группы:



Опишем наиболее известные постановки задач и результаты для каждой из указанных групп.

I. В качестве усиления понятия границы повторяемости Бадкобех и Крошмор определили функцию

$$\mathrm{FRT}(k) = \inf\{\beta \in \mathbb{R} \mid \text{существует } \beta^{+}\text{-свободное } \omega\text{-слово над алфавитом } \Sigma_k, \text{ которое содержит конечное число подслов с экспонентой } \mathrm{RT}(k)\},$$

которая получила название «finite-repetition threshold». Из работы Кархумяки и Шаллита [23] известно, что если бинарное ω-слово является $(7/3)$-свободным, то оно содержит бесконечно много квадратов, а в работе [43] Шаллит построил бинарное $(7/3)^{+}$-свободное слово, содержащее всего 18 квадратов, откуда следует равенство $\mathrm{FRT}(2) = 7/3$. Равенство $\mathrm{FRT}(3) = \mathrm{RT}(3)$ доказано Бадкобех и Крошмором в работе [5], а в их совместной работе с Рао [6] были проанализированы ω-слова, полученные в доказательстве гипотезы Дежан при $k \ge 4$ и установлено, что они удовлетворяют требуемому более сильному условию, т.е. $\mathrm{FRT}(k) = \mathrm{RT}(k)$ при $k \ge 4$.

Все граничные языки бесконечны. Но насколько «большими» они могут быть? Из [40] известно, что граничный язык над 2-буквенным алфавитом «небольшой», количество слов в нем с увеличением длины растет полиномиально. Для остальных алфавитов существует *экспоненциальная гипотеза* (происхождение которой считается фольклорным), согласно которой, начиная с 3-буквенного алфавита, граничные языки растут экспоненциально. Для $k = 3, 4$ экспоненциальную гипотезу доказал Ошем [35], для $5 \le k \le 10$ – Колпаков и Рао [25].

Любое граничное слово длины больше k обязательно содержит подслова вида $a_1a_2\cdots a_{k-1}a_1$. В работе [52] Туневым и Шуром для $k \geq 5$ было введено понятие *чистого* граничного слова, т.е. граничного слова, *все* подслова которого с экспонентой $\mathrm{RT}(k)$ имеют вид $a_1a_2\cdots a_{k-1}a_1$; тем самым, усиливается рассмотренное выше требование конечности числа подслов экспоненты $\mathrm{RT}(k)$. Естественным образом возникает вопрос: конечно или бесконечно множество чистых граничных слов над алфавитом Σ_k? Для 5-буквенного алфавита Бадкобех, Крошмор и Рао [6] доказали, что это множество бесконечно, построив граничное ω-слово, содержащее всего 60 подслов с экспонентой $5/4$, период каждого из которых равен 4. Тунев и Шур для любого нечетного $7 \leq k \leq 101$ доказали более сильный результат [52]: множество всех чистых граничных слов над алфавитом Σ_k не только бесконечно, но и имеет экспоненциальный рост. Как следствие, этот результат доказывает экспоненциальную гипотезу для любого нечетного $7 \leq k \leq 101$.

II. Ограничения можно накладывать не только на экспоненту слов, но и на длину повторяющейся части. В [22] Илие, Ошем и Шаллит ввели понятие *обобщенной* границы повторяемости $\mathrm{RT}(k,l)$, которая учитывает только повторы с длиной повторяющейся части $\geq l$. Если параметр l равен единице, то обобщенная граница повторяемости совпадает с обычной. На данный момент точных значений обобщенной границы повторяемости известно мало. Фиоренци, Ошем и Васле получили наилучшие верхние оценки для $\mathrm{RT}(k,l)$ [20], а Румянцев – наилучшие нижние оценки [41].

III. До этого момента мы считали, что «повторяющиеся» фрагменты слова должны полностью совпадать. Но можно рассматривать так называемые *абелевы* степени, когда повторяющиеся фрагменты являются *анаграммами* друг друга, т.е. совпадают лишь набором (мультимножеством) входящих в них букв. Например, в слове «ротатор» фрагменты «рот» и «тор» равны в абелевом смысле, поэтому минимальный абелев период этого слова равен 4, а абелева экспонента – 7/4, хотя обычная – 7/6. Самсонов и Шур в [42] предложили абелев аналог гипотезы Дежан, согласно которому (сильная) абелева граница повторяемости $\mathrm{ART}(k)$ предположительно принимает следующие значения:

k	2	3	4	5	$\cdots$	i	$\cdots$
$\mathrm{ART}(k)$	$11/3$	2	$9/5$	$3/2$	$\cdots$	$(i-2)/(i-3)$	$\cdots$

При этом в [42] доказано, что значения в таблице являются оценками снизу для $\mathrm{ART}(k)$.

Понятие t-*абелевой* степени служит промежуточным звеном между обычными и абелевыми степенями. Слово $u_1u_2\cdots u_n$ называется t-абелевой n-степенью, если в словах $u_1, u_2, \cdots, u_n$ совпадает количество вхождений любых подслов длины $\leq t$. Нетрудно заметить, что при $t = 1$ определение t-абелевой степени совпадает с целой абелевой степенью, а с ростом t становится все более похожим на обычную степень. Керянен доказал, что абелевы (а значит, и t-абелевы для любых $t \geq 1$) квадраты избегаемы над 4-буквенным алфавитом [24]; над 3-буквенным алфавитом неизбежны 2-абелевы квадраты и есть гипотеза о том, что 3-абелевы квадраты избегаемы [27]; над бинарным алфавитом избегаемы 3-абелевы кубы [29] и предполагается избегаемость 2-абелевых кубов (в то же время абелевы кубы неизбежны, см. выше про $\mathrm{ART}(k)$).

IV. Повторы можно искать не только в «обычных» словах, но и в *циклических*, у которых склеены начало и конец. Тогда, оставив прежнее определение повтора, можно получить аналог границы повторяемости для циклических слов (точнее целых три аналога: для слабой, средней и сильной циклической границы повторяемости). Значения всех трех видов циклической границы повторяемости для бинарного алфавита установили Аберкане и Карри [3, 4], а для тернарного алфавита – Шур [48]:

k	$\mathrm{CRT}_W(k)$	$\mathrm{CRT}_I(k)$	$\mathrm{CRT}_S(k)$
2	2	$7/3$	$5/2$
3	$7/4$	$7/4$	2

Кроме того, для слабой и средней циклической границы повторяемости Карри выдвинул гипотезу, что, начиная с 3-буквенного алфавита, их значения совпадают с «обычной» границей повторяемости, т.е. $\mathrm{CRT}_W(k) = \mathrm{CRT}_I(k) = \mathrm{RT}(k)$ для любого $k \geq 3$.

Обзор диссертации

Объект исследования и цели диссертации

Объектом исследования являются экстремальные бесповторные слова. Причем помимо «обычных» слов, на позициях которых задан линейный порядок, рассматриваются также циклические слова. В качестве вспомогательных объектов возникают двумерные слова, а также цилиндрические слова, имеющие как линейную, так и двумерную структуру.

Основной целью является развитие теории бесповторных слов в двух основных направлениях: структурный и количественный анализ экстремальных бесповторных слов и нахождение границы повторяемости для циклических слов.

Основные проблемы, рассматриваемые в диссертации:

1. Описать строение экстремальных бесповторных слов и выявить структурные закономерности, общие для всех алфавитов.

2. Оценить количество экстремальных бесповторных слов заданной длины.

3. Найти значение циклической границы повторяемости.

Основные методы

Для доказательства результатов, представленных в диссертации, в основном используются методы комбинаторики слов, основанные на свойствах периодических слов, кодировании, циклических и двумерных словах. Также используется метод регулярных приближений и комбинаторные алгоритмы для вычисления индекса роста регулярных языков (модификации А.М. Шура). Практически все численные результаты получены с помощью компьютера.

Структура диссертации и основные результаты

Диссертация состоит из введения, двух глав, списка литературы и приложения. Общий объем диссертации составляет 107 страниц. Библиографический список содержит 60 наименований.

Во введении определяются основные используемые понятия, а также приводится краткий обзор исследований в области бесповторных последовательностей.

Первая глава посвящена «обычным» экстремальным бесповторным словам и состоит из двух разделов. В первом разделе приводятся два способа кодирования граничных слов: бинарное (Пансьё) и цилиндрическое (Шур), описание которых можно найти в пунктах 1.1.1 и 1.1.2, соответственно. Идея цилиндрического кодирования граничных слов оказалась весьма плодотворной и позволила выявить закономерности в строении граничных слов, общие для всех алфавитов. Основным результатом этого раздела является полное описание m-повторов для всех $m \leq 7$ (см. теоремы 1.1,1.2,1.3 и 1.4). Описание m-повторов универсально для всех алфавитов, содержащих не менее $2m-2$ букв и представляет из себя конечный список Q_m двумерных слов, каждое из которых является «маркером» цилиндрического представления m-повтора. Отсутствие таких маркеров является необходимым условием того, что слово является граничным.

Во втором разделе найденные множества Q_m используются для построения регулярных приближений граничных языков и получения верхних оценок для их индексов роста (см. таблицу 1.1). Основываясь на результатах вычислительных экспериментов, нами было сделано предположение, что равномерные m-повторы не только одинаково устроены для всех алфавитов, но и привносят практически одинаковый вклад в индексы роста граничных языков над различными алфавитами. Таким образом, была сформулирована гипотеза *о существовании пределов* $\lim\limits_{k\to\infty} \alpha(T_k^{(m)}) = \alpha^{(m)}$ *для любого* $m \geq 3$ (см. гипотезу 1.1). Стоит отметить, что в случае $m = 2$ последовательность $\big(\alpha(T_k^{(m)})\big)_{k\geq 5}$ является стационарной, т.е. $\alpha(T_k^{(2)}) = \alpha^{(2)} \approx 1,324718$ для любого $k \geq 5$. В ключевом с точки зрения оценки индексов роста граничных языков случае $m = 3$ удалось доказать не только существование данного предела ($\alpha^{(3)} \approx 1.242096777$), но и то, что он *совпадает с индексом роста двумерного языка*, заданного множеством запрещенных подслов Q_3, полученным в первом разделе для цилиндрических представлений. Соответствующий результат (теорема 1.6) является основным в данном разделе и позволяет усилить гипотезу 1.1 (см. замечание 1.15). Кроме того, сравнивая вклад 3-, 6- и 7-повторов, можно предположить, что с ростом m влияние m-повторов на индекс роста граничных языков убывает. Следовательно, можно сформулировать гипотезу, что *индексы роста граничных языков мало отличаются от индексов роста уже исследованных приближений*, т.е. $\lim\limits_{k\to\infty} \alpha(T_k) = \alpha \approx 1.242$ (см. гипотезу 1.2).

Вторая глава посвящена циклическим экстремальным бесповторным словам. Получен и практически полностью доказан *аналог гипотезы Дежан для циклических слов* (см. гипотезу 2.1). Нижняя оценка для циклической границы повторяемости получена для любого $k \geq 4$ (см. предложение 2.1). Для получения верхних оценок была предложена общая конструкция бесповторных циклических слов, корректность которой обеспечивается граничной теоремой. Детали реализации данной конструкции существенно различаются для разных алфавитов, поэтому пришлось отдельно рассмотреть случаи $k = 4, 5, 6, 7, 9$ и нечетные $k \geq 11$ (см. теоремы 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, соответственно). В результате было доказано, что $\mathrm{CRT}(k) = \frac{\lceil k/2\rceil+1}{\lceil k/2\rceil}$ для любого $k \geq 6$, а также установлено, что $3/2 \leq \mathrm{CRT}(4) \leq 7/4$ и $4/3 \leq \mathrm{CRT}(5) \leq 3/2$ (гипотеза 2.1 утверждает, что в обоих случаях точной является нижняя оценка).

Апробация результатов работы

Результаты диссертации были представлены на 12-й международной конференции по теоретическим компьютерным наукам (Монс, 2008), 8-й международной конференции по комбинаторике слов (Прага, 2011), втором российско-финском симпозиуме по дискретной математике (Турку, 2012), а также на семинарах «Алгебраические системы» и «Дискретная математика» (УрГУ/УрФУ, 2008-2012).

Ссылки на результаты диссертации присутствуют в работах других авторов: [6, 8, 16, 25, 34].

Публикации

Основные результаты диссертации опубликованы в работах [54–60], среди которых три работы [56, 58, 60] опубликованы в рецензируемых изданиях из списка, рекомендованного ВАК. В совместных работах [59, 60] А.М. Шуром было введено цилиндрическое кодирование и доказан ряд общих свойств цилиндрических представлений граничных слов, а автором было получено описание структуры цилиндрических представлений равномерных m-повторов для $3 \leq m \leq 6$ и выполнена экспериментальная часть. В совместных работах [57, 58] А.М. Шуру принадлежат постановки задач, а также усовершенствование ряда предложенных автором доказательств и вспомогательных конструкций; доказательство основных результатов принадлежит автору.

Глава 1

Граничные слова

В данной главе речь будет идти о $\mathrm{RT}(k)^+$-свободных словах над алфавитом Σ_k, где $k \geq 5$. Случаи $k = 3$, $k = 4$ выбиваются из общей формулы $\mathrm{RT}(k) = k/(k-1)$, и структура граничных слов для них во многом специфична и нехарактерна для больших алфавитов. Поэтому мы исключили их из рассмотрения. Что касается случая $k = 2$, то для него структура граничных слов хорошо изучена (см., в частности, [12, 40, 49]).

1.1. Структура граничных слов

Целью данного раздела является поиск общих закономерностей в строении граничных слов над разными алфавитами. Вполне естественно для решения этой задачи использовать некоторую кодировку граничных слов, которая позволит работать над одним алфавитом.

1.1.1. Кодировка Пансьё

Пытаясь доказать гипотезу Дежан, Пансьё в [36] предложил идею бинарного кодирования $(k-1)/(k-2)$-свободных слов (граничные слова над любым алфавитом являются $(k-1)/(k-2)$-свободными). Конструкция, описанная Пансьё, основывается на том, что любое $(k-1)/(k-2)$-свободное слово над алфавитом Σ_k удовлетворяет двум условиям:

(1) в любом подслове длины $k-1$ все буквы различны;

(2) двум соседним вхождениям одной буквы предшествуют разные символы.

Любое слово (Z-слово), удовлетворяющее условиям (1) и (2), будем называть *словом Пансьё*. Предположим, что в слове Пансьё $w \in \Sigma_k^*$ встречается символ a, которому предшествует цепочка $a_1 \ldots a_{k-1}$. В силу (1) все буквы $a_1, \ldots, a_{k-1}$ различны, более того, в подслове $a_2 \ldots a_{k-1}a$ также не должно быть повторяющихся символов. Значит возможны всего два варианта: либо $a = a_1$, либо $a = a_k$, где $a_k \notin \{a_1, \ldots, a_{k-1}\}$. Будем символ a кодировать нулем в первом случае и единицей во втором (кодировать начинаем с k-го символа в слове), тогда слову w можно поставить в соответствие *характеристическое* слово $\mathrm{Bin}(w)$ над бинарным алфавитом $B = \{0, 1\}$, построенное по следующему правилу:

$$\mathrm{Bin}(w)[i] = \begin{cases} 0 & \text{, если } w[k{+}i{-}1] = w[i]; \\ 1 & \text{, если } w[k{+}i{-}1] \notin \{w[i], \ldots, w[k{+}i{-}2]\}. \end{cases}$$

Так, например, для некоторого $w \in \Sigma_5^*$:

$$\begin{array}{rcl} w & \to & a\,b\,c\,d\,a\,e\,c\,b\,d\,e\,a\,b\ldots \\ \mathrm{Bin}(w) & \to & \qquad\quad 0\,1\,0\,1\,1\,0\,1\,0\ldots \end{array}$$

Отметим, что каждое характеристическое слово длины $n{-}k{+}1$ определяет единственное с точностью до переименования букв слово длины n с заданными свойствами (таким образом, вместо $k!$ слов над алфавитом Σ_k можно рассматривать одно слово над бинарным алфавитом).

Сформулируем ряд важных свойств слов Пансьё и их бинарных кодов.

Замечание 1.1. *Множество всех слов Пансьё над алфавитом Σ_k совпадает с языком* $T_k^{(2)}$. Действительно, условие (1) обеспечивает отсутствие 1-повторов, а условие (2) – отсутствие 2-повторов.

Замечание 1.2. *Расстояние между двумя соседними вхождениями одной буквы в слове Пансьё $w \in \Sigma_k^*$ может принимать только три значения: $k{-}1$, k и $k{+}1$.* Действительно, пусть $a_1 a_2 \cdots a_{k-1}$ – подслово слова w. Найдем расстояние до следующего вхождения буквы a_1. Для этого достаточно рассмотреть всевозможные варианты продолжения этого префикса без нарушения правил (1) и (2):

$$a_1\, a_2\, a_3 \cdots a_{k-1} \left| \begin{array}{l} a_1 \cdots \\ a_k\ a_1\ \cdots \\ a_k\ a_2\ \ a_1\ \cdots \end{array} \right.$$

Замечание 1.3. *Если w – слово Пансьё, то* $\mathrm{Bin}(w)$ *не содержит подслов* 00 *и* 111. Действительно, предположим, что это не так. Тогда получим противоречие со свойством (2):

$$\begin{array}{cc} \qquad\qquad\quad 0\ \ 0 & \qquad\qquad\quad 1\ \ 1\ \ 1 \\ a_1\, a_2\, a_3 \ldots a_{k-1}\, a_1\, a_2 & a_1\, a_2\, a_3 \ldots a_{k-1}\, a_k\, a_1\, a_2 \end{array}$$

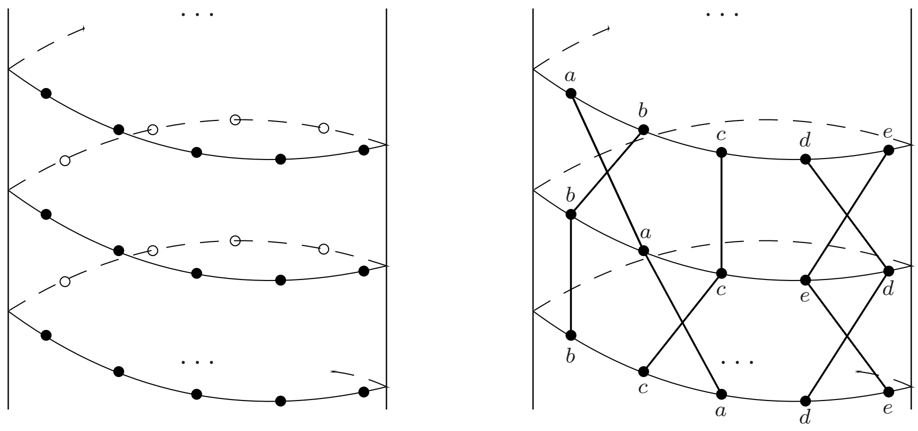
Замечание 1.4. *Непосредственно из замечания 1.3 следует, что любое характеристическое слово разбивается на блоки* 1 *и* 01 *(с возможно нецелыми блоками в начале и/или конце слова), причем блок* 1 *может появляться только в окружении блоков* 01.

Таким образом, при переходе к характеристическим словам нам сразу удалось найти нечто общее между граничными словами над разными алфавитами (см. замечания 1.3 и 1.4). А именно: для множества бинарных кодов слов языка $T_k^{(2)}$ антисловарем является множество $\{00, 111\}$ независимо от значения k. Все прочие структурные закономерности непосредственно зависят от мощности исходного алфавита, что заставляет нас продолжить поиски более подходящего способа кодирования граничных слов.

1.1.2. Цилиндрическое представление

Начиная с этого пункта и до конца раздела, речь будет идти о Z-словах Пансьё.

Продолжая идею кодирования, можно рассмотреть графическое представление слов Пансьё, которое было предложено А.М. Шуром и впервые описано в работе [59]. Представим, что у нас есть некоторый бесконечный цилиндр, а слово – это нить с нанизанными на неё буквами. Тогда слово можно «наматывать» на цилиндр так, чтобы количество букв в одном витке равнялось мощности алфавита k (см. рис. 1.1 a)). Теперь соединим одинаковые символы в соседних витках отрезками. По замечанию 1.2 соседние вхождения одного символа можно соединить тремя способами (считаем, что отрезок идет от нижней буквы к верхней): ↕ (буква осталась на той же позиции), ↗ (сместилась на одну позицию вправо) и ↖ (сместилась на одну позицию влево) (см. рис. 1.1 b)), которые эквивалентны расстояниям k, $k-1$ и $k+1$, соответственно.



a) Наматываем слово на цилиндр ($k = 9$) b) Соединяем одинаковые буквы

Рис. 1.1. *Цилиндрическое представление слов.*

Полученное графическое изображение из отрезков будем называть *цилиндрическим представлением* слова $\mathbf{w}$ и обозначать через $\mathrm{Cyl}(\mathbf{w})$. «Траекторию» движения каждой буквы в цилиндрическом представлении будем называть её *следом*. Таким образом, след буквы – это ломаная, составленная из отрезков, соответствующих перемещениям буквы между витками. Если говорить о следе буквы в цилиндрическом представлении некоторого подслова u слова $\mathbf{w}$, то это будет ломаная, начинающаяся в первом вхождении данной буквы в u и заканчивающаяся в последнем. Например, след буквы a в подслове $abcde \cdots bcade$ Z-слова Пансьё, изображенном на рис. 1.1, имеет вид $\backslash\backslash$.

Перечислим ряд свойств цилиндрических представлений слов Пансьё.

Замечание 1.5. *Пусть* $\mathbf{w}$ *–* Z*-слово Пансьё, тогда двум соседним буквам в* $\mathbf{w}$ *могут соответствовать только следующие пары отрезков в* $\mathrm{Cyl}(\mathbf{w})$*:* $\vert\,/$, $/\,\backslash$, $\backslash\,\vert$ *и* $\backslash\,/$. *Пары* $\vert\,\vert$, $/\,/$ *и* $\backslash\,\backslash$ *невозможны, так как приводят к нарушению правила* (2)*, а в остальных случаях происходит «наложение» букв (разные буквы переходят на одно и то же место). Следовательно, цилиндрическое слово* Z*-слова Пансьё является комбинацией палочек* $\vert$ *и крестиков* $/\,\backslash = \times$*, причем палочка может появляться только в окружении крестиков (т.е. на самом деле можно говорить о блоках* $\vert\,\times$ *и* $\times$*).*

Замечание 1.6. *Легко заметить соответствие между блоком* 1 *в характеристическом слове и блоком* $\vert$ *в цилиндрическом представлении, а также между блоками* 01 *и* $\times$. Действительно 0 в характеристическом слове говорит о том, что предыдущее вхождение соответствующей буквы находится на расстоянии $k{-}1$ от данного, а значит согласно замечанию 1.5 в цилиндрическом представлении эта буква будет закодирована $/$. Появление 1 после 0 говорит о том, что между двумя вхождениями соответствующей ей буквы находится подслово ограниченное двумя вхождениями буквы, соответствующей 0, т.е. расстояние будет равно $k{+}1$, а в цилиндрическом коде получим $\backslash$. И наконец, 1, идущая после 1, эквивалентна появлению буквы на расстоянии k от предыдущего ее вхождения, что дает нам $\vert$ в цилиндрическом представлении.

Таким образом, если для некоторого Z-слова Пансьё $\mathbf{w}$ мы «размотаем» его цилиндрическое представление $\mathrm{Cyl}(\mathbf{w})$ и вытянем его в одну линию, то получим в точности перекодированное (согласно найденному в замечании 1.6 соответствию) характеристическое слово $\mathrm{Bin}(\mathbf{w})$. Например, для некоторого фрагмента Z-слова Пансьё $\mathbf{w}$ над алфавитом Σ_5:

$$
\begin{aligned}
\mathbf{w} &= \cdots abcde\,ba\ c\ ed\ b\ ca \cdots \\
\mathrm{Bin}(\mathbf{w}) &= \qquad\ \cdots 01\ 1\ 01\ 1\ 01 \cdots \\
\mathrm{Cyl}(\mathbf{w}) &= \qquad\ \cdots \times\ \vert\ \times\ \vert\ \times \cdots
\end{aligned}
$$

Для множества цилиндрических представлений (как и для множества характеристических слов) слов Пансьё выполняются общие закономерности, которые выражаются в одном и том же не зависящем от k множестве запрещенных объектов (следует из замечания 1.5). Однако за счет двумерной структуры цилиндрическое представление позволяет выявить и другие общие закономерности, о чем будет изложено далее.

1.1.3. Равномерные m-повторы

Среди m-повторов можно выделить два класса: *короткие* ($m < k$) и *длинные* или *ядерные* ($m \geq k$) (такое разделение было впервые введено в [33]). Подробнее о длинных повторах можно найти в [11, 33], далее речь пойдет только о коротких повторах. Пусть ptp – короткий m-повтор, тогда все буквы в слове p различны в силу условия (1). Более того, за счет выполнения условия $m < k$ слово p всегда целиком укладывается в один виток на цилиндре. Короткий m-повтор ptp называется *равномерным*, если каждая буква $a \in \Sigma(p)$ встречается в ptp одинаковое число раз. Следующие четыре свойства m-повторов будут весьма полезны в дальнейшем (их доказательство принадлежит А.М. Шуру и может быть найдено в [59, 60]).

Лемма 1.1. *Пусть $ptp \in \Sigma_k^*$ – m-повтор такой, что $m < (k{+}3)/2$. Тогда данный повтор является равномерным, а именно, каждая буква $a \in \Sigma(p)$ входит в слово ptp ровно m раз.*

Лемма 1.2. *Пусть $ptp \in \Sigma_k^*$ – равномерный m-повтор. Тогда каждая буква $a \in \Sigma$ входит в слово pt ровно $m{-}1$ раз.*

Следствие 1.1. *Если $ptp \in \Sigma_k^*$ – равномерный m-повтор, то при наматывании слова ptp на цилиндр оба вхождения слова p окажутся одно под другим, причем длина следа (т.е. количество отрезков в нем) любой буквы $a \in \Sigma(p)$ будет равна $m{-}1$, а количество отрезков $\nearrow$ и $\searrow$ в нем одинаково.*

Утверждение следствия 1.1 можно обратить, получая следующее предложение:

Предложение 1.1. *Пусть $m \geq 3$ – фиксированное число. Тогда для любого натурального $k > 2m{-}3$ Z-слова Пансьё $\mathbf{w}_k^{(m)} \in \Sigma_k^Z$, избегающие r-повторы для всех $r \leq m$, имеют идентичное описание на языке цилиндрических представлений. А именно существует конечное множество Q_m фрагментов размера $O(m){\times}O(m)$ такое, что слово является подсловом Z-слова Пансьё $\mathbf{w}_k^{(m)}$ тогда и только тогда, когда его цилиндрическое представление не содержит фрагментов из Q_m.*

Действительно, если из цилиндрического представления равномерного m-повтора ptp «вырезать» фрагмент размера $O(m){\times}O(m)$, ограниченный сверху и снизу словом

p, то независимо от мощности исходного алфавита он должен состоять из m следов длины $m-1$, которые начинаются и заканчиваются в одинаковых позициях. Так, задача построения цилиндрического представления 3-повтора (а из леммы 1.1 следует, что все они равномерны при любом $k \geq 5$) сводится к нахождению подходящих следов длины 2, которые бы соединили буквы a, b и c в верхнем и нижнем ряду на рис. 1.2.

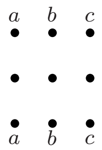


Рис. 1.2. *3-повторы.*

Чтобы буква через два шага вернулась на ту же позицию, ее след должен иметь один из трех видов: , , . на рис. 1.3 показаны всевозможные комбинации этих следов, не противоречащие замечанию 1.5.

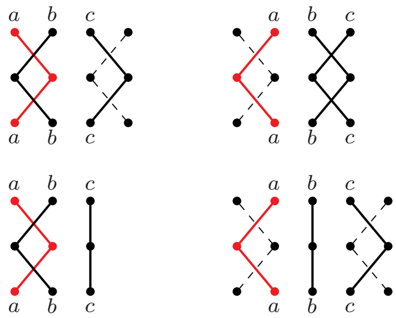


Рис. 1.3. *3-повторы. Пунктиром показаны следы соседних букв, которые однозначно достраиваются по имеющимся следам.*

Таким образом, основными целями наших исследований в данном направлении являются:

- нахождение множеств Q_m для малых значений m;
- поиск закономерностей в структуре множеств Q_m для различных значений m, если таковые имеются;
- использование найденных множеств Q_m для получения численных оценок для индексов роста соответствующих граничных языков.

1.1.4. 3-повторы

Описание множества Q_3 можно легко получить, исходя из представленных на рис. 1.3 фрагментов.

Теорема 1.1. *Для Z-слова Пансьё $\mathbf{w} \in \Sigma_k^Z$ следующие условия эквивалентны:*

1) $\mathbf{w}$ *содержит* 3*-повтор;*

2) цилиндрическое представление $\mathrm{Cyl}(\mathbf{w})$ *содержит* *или* *;*

3) характеристическое слово $\mathrm{Bin}(\mathbf{w})$ *содержит подслова вида* $1101\underbrace{\cdots}_{k-6}1011$ *или* $0101\underbrace{\cdots}_{k-5}1010$ *(для* $k = 5$ *первое слово имеет вид* 1101011*).*

Доказательство. 1) $\Rightarrow$ 2). Пусть 3-повтор $u = abc \cdots abc$ является подсловом в $\mathbf{w}$. Тогда $\mathrm{Cyl}(\mathbf{w})$ содержит один из фрагментов, изображенных на рис. 1.3. Так как $\mathrm{Cyl}(\mathbf{w})$ является Z-словом, то u имеет продолжение в $\mathbf{w}$ как влево, так и вправо. Тогда следы букв a, b и c могут быть однозначно дополнены следами соседних букв, которые показаны на рис. 1.3 пунктиром. В каждом случае $\mathrm{Cyl}(\mathbf{w})$ обязательно содержит или .

2) $\Rightarrow$ 1). Заметим, что если в $\mathbf{w}$ встречается фрагмент , то слева и справа от него обязательно находятся столбики из крестиков (см. замечание 1.5). Значит мы всегда можем найти три буквы, без ограничения общности a, b и c, которые конструкциями и возвращаются на свои места. Это означает, что $\mathbf{w}$ содержит подслово u вида $abc\underbrace{\cdots}_{2k-3}abc$, которое по определению является 3-повтором.

2) $\Leftrightarrow$ 3). Согласно замечанию 1.6 появление в $\mathrm{Cyl}(\mathbf{w})$ означает, что бинарное представление некоторого подслова из $\mathbf{w}$ имеет вид $11\underbrace{\cdots}_{k-2}11$ [1] (и наоборот, подслово $11\underbrace{\cdots}_{k-2}11$ в $\mathrm{Bin}(\mathbf{w})$ соответствуют в цилиндрическом представлении). По замечанию 1.4 в бинарном коде слова Пансьё перед 11 всегда стоит 10, а после – 01. В результате получаем требуемое подслово $1101\underbrace{\cdots}_{k-6}1011$. Аналогично, появление в $\mathrm{Cyl}(\mathbf{w})$ эквивалентно $0101\underbrace{\cdots}_{k-4}0101$ в $\mathrm{Bin}(\mathbf{w})$. А так как перед 0 может быть только 1, то получаем слово $0101\underbrace{\cdots}_{k-5}10101$, которое содержит искомое подслово. $\square$

1.1.5. 4- и 5-повторы

Следующая лемма будет неоднократно использоваться в доказательстве последующих результатов.

Лемма 1.3. *Пусть* $ptp \in \Sigma_k^*$ *– равномерный* m*-повтор;* $a \in \Sigma(p)$*. Тогда след любой буквы* $b \neq a$ *пересекает след буквы* a *четное число раз.*

[1]Еще одна единица нужна, чтобы подчеркнуть, что перед нами блок 1, а не часть блока 01.

Доказательство. Предположим обратное. Если первое и последнее вхождения буквы a занимают позиции i и $i+(m-1)k$, а буквы b, соответственно, позиции j_1 и $j_2+(m-1)k$, где $j_1 < i < j_2$ или $j_2 < i < j_1$, то несложно заметить, что b не может принадлежать p (после переходов b оказывается по другую сторону от a, а в слове p порядок букв сохраняется). Тогда $|p| < |j_2-j_1|$, но $|j_2-j_1| \leq m-1$, так как буква b встречается в ptp ровно $m-1$ раз по лемме 1.2 и каждый раз может смещаться влево или вправо не более чем на 1. Получаем $|p| < m$, что противоречит определению m-повтора. □

Замечание 1.7. *Пусть след S длины $m-1$ подходит для равномерного m-повтора, т.е. входит в состав фрагмента $Q \in Q_m$ включающего m следов длины $m-1$, возвращающих буквы на те же позиции. Тогда следы, симметричные S относительно вертикальной или горизонтальной оси, также подходят для равномерных m-повторов.* Действительно, они будут входить в состав фрагментов, симметричных Q относительно вертикальной или горизонтальной оси, соответственно.

Теперь перейдем к основному результату данного пункта.

Теорема 1.2. *Не существует равномерных 4- и 5-повторов.*

Доказательство. Предположим, что существует равномерный 4-повтор ptp. Тогда для следа буквы $a \in \Sigma(p)$ есть всего два варианта с точностью до симметрии, которые изображены на рис. 1.4 (выделены жирными линиями). Но в случае a) след пересекает соседние следы всего по одному разу (см. рис. 1.4 a)), что по лемме 1.3 невозможно. А в случае b) соседний справа след однозначно достраивается до симметричного, но вот следующий за ним не возвращает букву на место (нельзя соединить точки b_3 и c_3 на рис. 1.4 b)). Следовательно, слева достроить след также не получится (нельзя попасть из точки b_1 в точку c_1). Таким образом не удалось найти подходящие следы для 4 последовательных букв в слове p, что доказывает отсутствие равномерных 4-повторов.

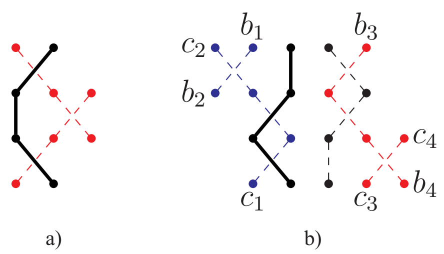


Рис. 1.4. *4-повторы.*

Теперь предположим, что существует равномерный 5-повтор ptp. Возможные с точностью до симметрии варианты следа буквы $a \in \Sigma(p)$ представлены на рис. 1.5

(выделены жирными линиями). В случаях $a)$ и $b)$ след пересекает соседние нечетное число раз, что невозможно по лемме 1.3. В случае $c)$ соседний слева след однозначно достраивается до симметричного, а справа достроить не удается (нельзя попасть из точки b_1 в точку c_1, см. рис. 1.5 c)). Значит в силу симметрии левая сторона также не может быть дальше достроена. В остальных случаях требуется соединить точки b_1 и c_1; b_2 и c_2 (см. рис. 1.5 d),e)), но тогда появятся конструкции или , которые по теореме 1.1 эквивалентны наличию 3-повтора в слове ptp, что противоречит определению 5-повтора.

□

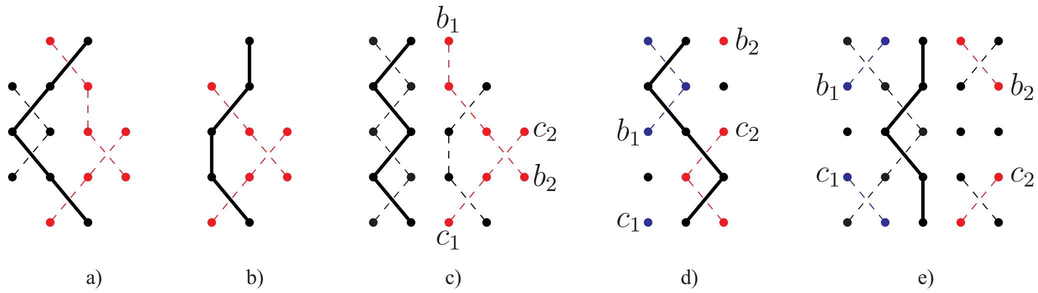


Рис. 1.5. *5-повторы.*

1.1.6. 6-повторы

Лемма 1.4. *Пусть ptp – равномерный 6-повтор. Тогда обязательно найдется буква $a \in \Sigma(p)$, след которой в цилиндрическом представлении слова ptp начинается или заканчивается на* .

Доказательство. В $\mathrm{Cyl}(ptp)$ 6 букв должны вернуться на свои места. Предположим, что все 6 следов начинаются и заканчиваются не на . Тогда имеет место один из случаев, изображенных на рис. 1.6.

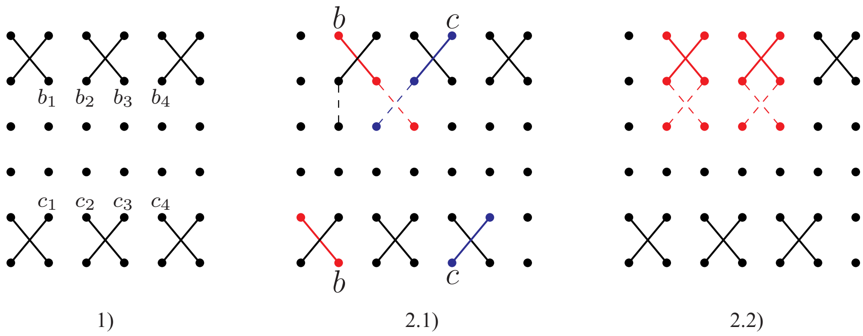


Рис. 1.6. *Невозможные фрагменты для 6-повторов.*

Первый случай, когда каждый след начинается и заканчивается разными отрезками, невозможен, так как мы должны соединить точки b_1, c_1; b_2, c_2; b_3, c_3; b_4, c_4, а это приведет к появлению равномерного 4-повтора, что невозможно по теореме 1.2. Теперь рассмотрим случай, когда каждый след начинается и заканчивается на одинаковые отрезки. Например, пусть след начинается и заканчивается на ↗. Тогда по следствию 1.1 в него должно войти еще два отрезка ↘ и один ↕. Если второй отрезок в следе равен ↕, то это испортит соседние следы, так как в них окажется три одинаковых отрезка (следы букв b и c на рис. 1.6). Если же вторые символы всех следов не равны ↕, то получим фрагмент в $\mathrm{Cyl}(ptp)$, а значит и 3-повтор в слове ptp по теореме 1.1, что невозможно по условию. $\square$

На рис. 1.7 изображено 5 различных допустимых следов для равномерных 6-повторов. Обозначим их $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}$, как показано на рисунке, а их зеркальные отражения относительно вертикальной оси обозначим, соответственно, $\mathbf{1}'$, $\mathbf{2}'$, $\mathbf{3}'$, $\mathbf{4}'$, $\mathbf{5}'$. Фрагмент, состоящий из нескольких идущих подряд следов, будем обозначать последовательностью номеров этих следов. Например, фрагмент на рис. 1.7 обозначается как $\mathbf{12345}$.

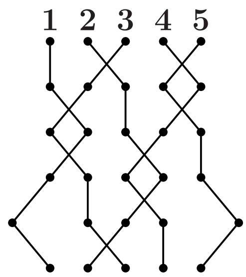


Рис. 1.7. *Допустимые следы для 6-повторов.*

Теорема 1.3. *Пусть $k \geq 10$. Тогда слово Пансьё $\mathbf{w} \in \Sigma_k^Z$, избегающее 3-повторы, содержит 6-повтор тогда и только тогда, когда $\mathrm{Cyl}(\mathbf{w})$ содержит один из фрагментов:* $\mathbf{123451}$, $\mathbf{234512}$, $\mathbf{345123}$, $\mathbf{451234}$, $\mathbf{512345}$, $\mathbf{1'5'4'3'2'1'}$, $\mathbf{2'1'5'4'3'2'}$, $\mathbf{3'2'1'5'4'3'}$, $\mathbf{4'3'2'1'5'4'}$, $\mathbf{5'4'3'2'1'5'}$.

Доказательство. По лемме 1.1, начиная с $k = 10$, все 6-повторы (а стало быть и все m-повторы с $m < 6$) над алфавитом Σ_k являются равномерными. Поскольку 6-повтор по определению является минимальным запрещенным словом, то условие отсутствия 3-повторов является излишним при доказательстве необходимости, однако оно играет ключевую роль при доказательстве достаточности. Равномерные 4- и 5-повторы в слове $\mathbf{w}$ отсутствуют по теореме 1.2.

$(\Rightarrow)$. Пусть $a_1a_2a_3a_4a_5a_6 \cdots a_1a_2a_3a_4a_5a_6$ – равномерный 6-повтор. Все возможные следы (без учета симметричных и полученных циклическим сдвигом [2]) букв $a_1, a_2, a_3, a_4, a_5, a_6$ представлены на рис. 1.8.

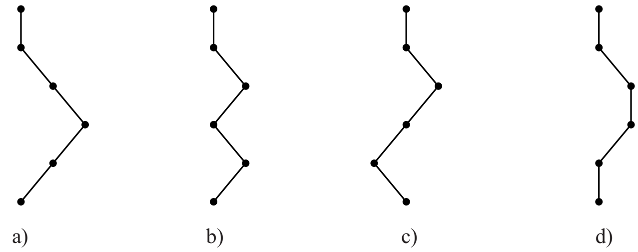


Рис. 1.8. *Варианты следов для 6-повторов.*

Согласно замечанию 1.7 можно не рассматривать следы, симметричные данным, а по лемме 1.4 можно не рассматривать циклические сдвиги следов $a)-c)$ (циклические сдвиги следа $d)$ содержат фрагмент , соответствующий наличию 3-повтора, поэтому и так не участвуют в рассмотрении). В случае $c)$ перед нами в точности след $\mathbf{1}$, всевозможные фрагменты, построенные с его помощью (включая симметричный ему след $\mathbf{1}'$), перечислены во втором пункте условия теоремы. Проверим, что следы $a)$, $b)$ и $d)$ нельзя достроить до нужного фрагмента.

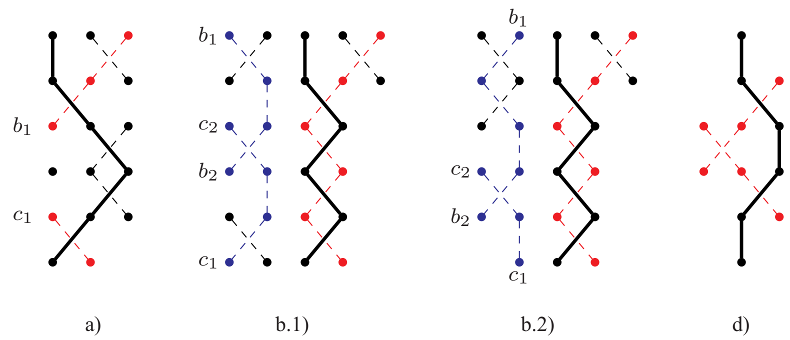


Рис. 1.9. *Варианты следов для 6-повторов.*

В случае $a)$ по лемме 1.3 необходимо соединить точки b_1 и c_1, но тогда появится одна из конструкций или , что по теореме 1.1 эквивалентно наличию 3-повтора, что невозможно по условию. В случае $b)$ получаем, что справа находится недопустимый для 6-повтора след. А слева достроить можно двумя способами, причем в каждом случае мы не можем попасть из точки b_1 в c_1, что также не дает возможность получения фрагмента, возвращающего 6 букв на те же позиции. След $d)$ имеет по одному пересечению с соседними следами, что по лемме 1.3 недопустимо для 6-повтора.

[2] Подразумевается, что на след можно смотреть как на вертикально записанное слово над алфавитом $\Delta = \{\ ,\ ,\ \}$

($\Leftarrow$). Наличие перечисленных конструкций в цилиндрическом представлении слова $\mathbf{w}$ свидетельствует о наличие у него подслова u вида

$$a_1a_2a_3a_4a_5a_6\underbrace{\cdots}_{5k-6}a_1a_2a_3a_4a_5a_6,$$

которое по определению является 6-повтором. Действительно, $\exp(u){=}(5k{+}6)/(5k){>}k/(k{-}1)$ и оно не содержит других повторов меньшей длины, так как 1-, 2-, 3-повторы запрещены по условию, равномерные 4- и 5-повторы не существуют по теореме 1.2, а неравномерных 4- и 5-повторов над алфавитом Σ_k для $k \geq 10$ нет по лемме 1.1. $\square$

Замечание 1.8. *Аналогично случаю $m = 3$ можно получить описание характеристических слов 6-повторов, используя соответствие, найденное в замечании 1.6. Тогда $\mathrm{Bin}(\mathbf{w})$, где $\mathbf{w}$ из условия теоремы 1.3, содержит подслово одного из следующих видов (перечислены в том же порядке, что и их цилиндрические аналоги из пункта 2 теоремы 1.3):*

$$11010110\underbrace{\cdots}_{k-7}0110101\underbrace{\cdots}_{k-7}0101101\underbrace{\cdots}_{k-8}0110101\underbrace{\cdots}_{k-7}0101101,$$
$$0101101\underbrace{\cdots}_{k-8}01101011\underbrace{\cdots}_{k-8}01011010\underbrace{\cdots}_{k-8}110101101\underbrace{\cdots}_{k-8}011010,$$
$$0101101\underbrace{\cdots}_{k-8}01101011\underbrace{\cdots}_{k-8}01011010\underbrace{\cdots}_{k-8}110101101\underbrace{\cdots}_{k-8}0110101,$$
$$0110101\underbrace{\cdots}_{k-7}0101101\underbrace{\cdots}_{k-8}0110101\underbrace{\cdots}_{k-7}0101101\underbrace{\cdots}_{k-7}1101011,$$
$$0110101\underbrace{\cdots}_{k-7}0101101\underbrace{\cdots}_{k-7}1101011\underbrace{\cdots}_{k-7}10110101\underbrace{\cdots}_{k-7}010110,$$
$$1101011\underbrace{\cdots}_{k-7}0101101\underbrace{\cdots}_{k-7}0110101\underbrace{\cdots}_{k-7}10101101\underbrace{\cdots}_{k-7}011010,$$
$$0110101\underbrace{\cdots}_{k-8}110101101\underbrace{\cdots}_{k-9}010110101\underbrace{\cdots}_{k-9}01101011\underbrace{\cdots}_{k-7}0101101,$$
$$0110101\underbrace{\cdots}_{k-8}110101101\underbrace{\cdots}_{k-9}010110101\underbrace{\cdots}_{k-9}01101011\underbrace{\cdots}_{k-7}010110,$$
$$0101101\underbrace{\cdots}_{k-7}0110101\underbrace{\cdots}_{k-7}10101101\underbrace{\cdots}_{k-7}0110101\underbrace{\cdots}_{k-8}1101011,$$
$$0101101\underbrace{\cdots}_{k-7}011010\underbrace{\cdots}_{k-7}1101011\underbrace{\cdots}_{k-7}0101101\underbrace{\cdots}_{k-7}0110101.$$

1.1.7. 7-повторы

Лемма 1.5. *Пусть ptp – равномерный 7-повтор. Тогда обязательно найдется буква $a \in \Sigma(p)$, след которой в цилиндрическом представлении слова ptp либо начинается (заканчивается) на $\updownarrow$, либо начинается и заканчивается одинаковыми наклонными отрезками, а на втором (предпоследнем) месте у него стоит $\updownarrow$.*

Доказательство. В $\mathrm{Cyl}(ptp)$ 7 букв должны вернуться на свои места. Предположим, что все 7 следов начинаются и заканчиваются не на $\updownarrow$. Тогда выполняется один из случаев, изображенных на рис. 1.10.

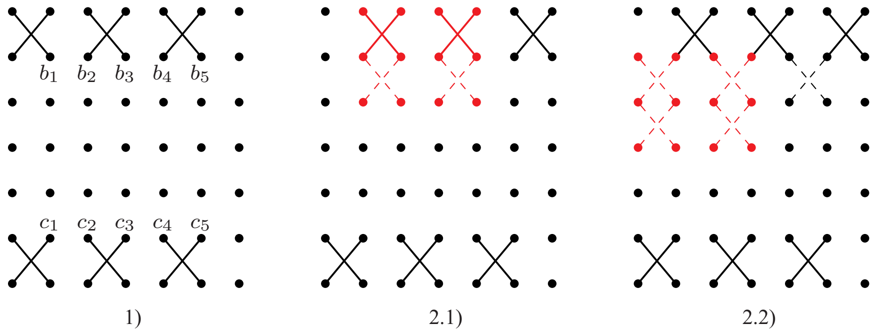


Рис. 1.10. *Невозможные фрагменты для 7-повторов.*

Первый случай, когда каждый след начинается и заканчивается разными отрезками, невозможен, так как мы должны соединить точки b_1, c_1; b_2, c_2; b_3, c_3; b_4, c_4; b_5, c_5, а это приведет к появлению равномерного 5-повтора, что невозможно по теореме 1.2. Теперь рассмотрим случай, когда каждый след начинается и заканчивается на одинаковые наклонные отрезки. Предположим, что второй отрезок каждого из 7 следов не равен . Если второй отрезок в следе не совпадает с первым, то получим фрагмент (см. рис. 1.10 случай 2.1)) в $\mathrm{Cyl}(ptp)$, а значит и 3-повтор в слове ptp по теореме 1.1, что невозможно поскольку 7-повтор по определению является минимальным запрещенным словом. Если же второй отрезок в следе совпадает с первым (а значит и с последним) и равен б.о.о. , то по следствию 1.1 остальные три отрезка должны быть равны . И снова получим фрагмент (см. рис. 1.10 случай 2.2)) в $\mathrm{Cyl}(ptp)$, что невозможно. $\square$

На рис. 1.11 показаны возможные варианты следов для равномерного 7-повтора, обозначенные $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}$, а также симметричные им – $\mathbf{1}', \mathbf{2}', \mathbf{3}', \mathbf{4}'$.

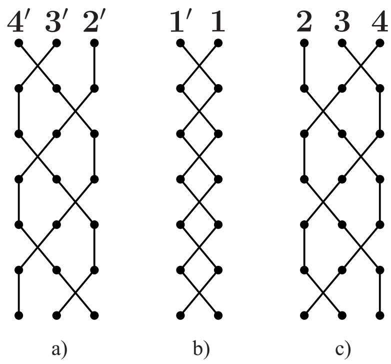


Рис. 1.11. *Возможные следы для 7-повторов.*

Теорема 1.4. *Пусть $k \geq 12$. Тогда слово Пансьё $\mathbf{w} \in \Sigma_k^Z$, избегающее 3- и 6-повторы, содержит 7-повтор тогда и только тогда, когда $\mathrm{Cyl}(\mathbf{w})$ содержит один из фрагментов:*

$\mathbf{2342342}$, $\mathbf{3423423}$, $\mathbf{4234234}$, $\mathbf{4'3'2'4'3'2'4'}$, $\mathbf{3'2'4'3'2'4'3'}$, $\mathbf{2'4'3'2'4'3'2'}$, $\mathbf{1'123423}$, $\mathbf{1'14'3'2'4'3'}$, $\mathbf{2341'123}$, $\mathbf{341'1234}$, $\mathbf{41'12342}$, $\mathbf{2341'14'3'}$, $\mathbf{341'14'3'2'}$, $\mathbf{41'14'3'2'4'}$, $\mathbf{4'3'2'1'14'3'}$, $\mathbf{3'2'1'14'3'2'}$, $\mathbf{2'1'14'3'2'4'}$, $\mathbf{4'3'2'1'123}$, $\mathbf{3'2'1'1234}$, $\mathbf{2'1'12342}$, $\mathbf{2342341'}$, $\mathbf{342341'1}$, $\mathbf{42341'12}$, $\mathbf{42341'14'}$, $\mathbf{4'3'2'4'3'2'1'}$, $\mathbf{3'2'4'3'2'1'1}$, $\mathbf{2'4'3'2'1'14'}$, $\mathbf{2'4'3'2'1'12}$, $\mathbf{1'12341'1}$, $\mathbf{1'14'3'2'1'1}$, $\mathbf{41'12341'}$, $\mathbf{41'14'3'2'1'}$, $\mathbf{2'1'14'3'2'1'}$, $\mathbf{2'1'12341'}$.

Доказательство. Отметим, что по лемме 1.1, начиная с $k = 12$, все 7-повторы (а стало быть и все m-повторы с $m < 7$) над алфавитом Σ_k являются равномерными.

($\Rightarrow$). Пусть $a_1a_2a_3a_4a_5a_6a_7 \cdots a_1a_2a_3a_4a_5a_6a_7$ – равномерный 7-повтор. Возможные следы (без учета симметричных и полученных циклическим сдвигом) букв $a_1, a_2, a_3, a_4, a_5, a_6, a_7$ представлены на рис. 1.12.

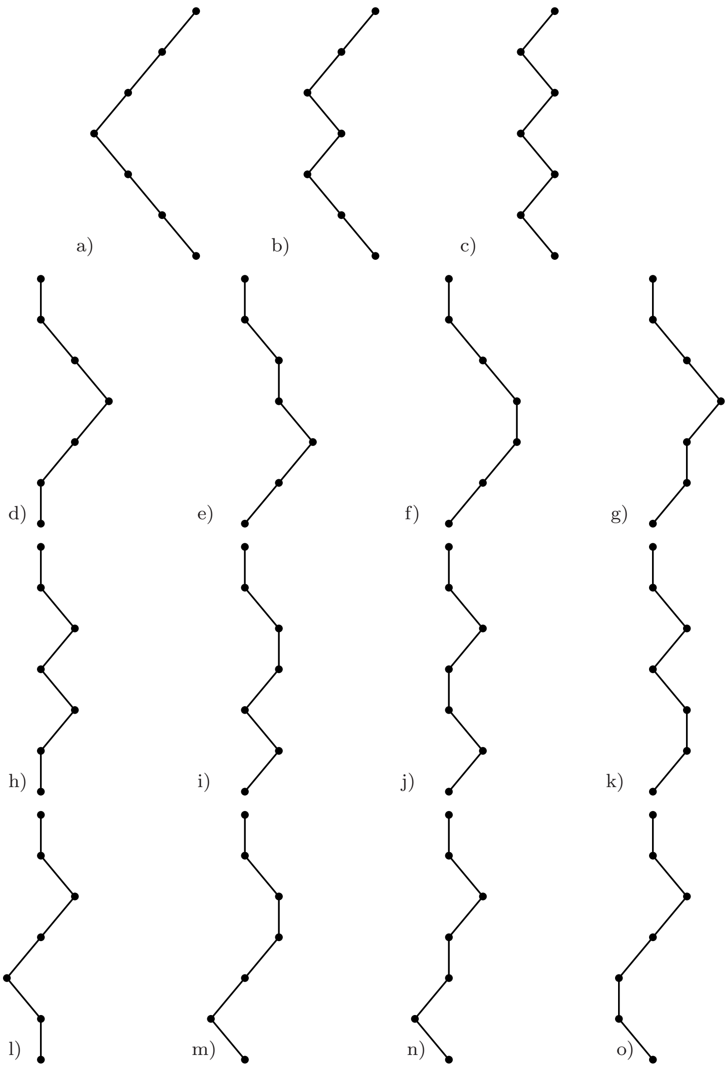


Рис. 1.12. *Варианты следов для 7-повторов.*

В силу замечания 1.7 можно не рассматривать симметричные случаи, а по лемме 1.5 достаточно рассмотреть только случаи $d) - o)$ и циклические сдвиги следов

$e) - g), m), o)$, начинающиеся и заканчивающиеся одинаковыми наклонными отрезками с $\vert$ на втором (или предпоследнем) месте. Отметим, что $f)$ – это в точности след **2**, а его подходящий циклический сдвиг – след **3**. Всевозможные фрагменты, подходящие для 7-повтора, с участием этих и симметричных им следов описаны в пункте 2 условия теоремы. Проверим, что остальные следы на эту роль не подходят.

Следы $d), i), k), m), o)$ имеют нечетное число пересечений с соседними следами (см. рис. 1.13), поэтому не подходят по лемме 1.3.

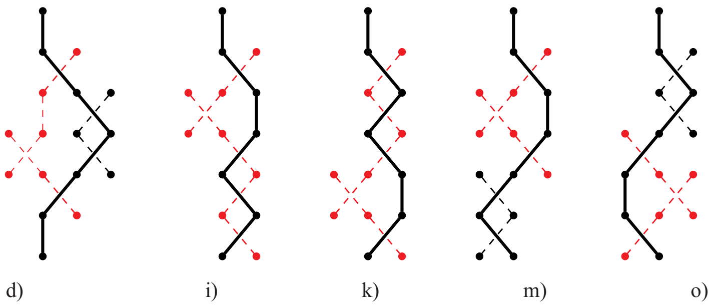


Рис. 1.13. *Варианты следов для 7-повторов.*

Рассмотрим следы $e)$ и $g)$ (см. рис. 1.14). Соседний справа от них след единственным образом можно продолжить так, чтобы он имел четное число пересечений с исходным следом, но тогда он не возвращает букву на место (точки b_3 и c_3 нельзя соединить), а соседний слева след можно достроить двумя способами, но тогда соседний с ним след будет неправильным (точки b_1 и c_1 нельзя соединить).

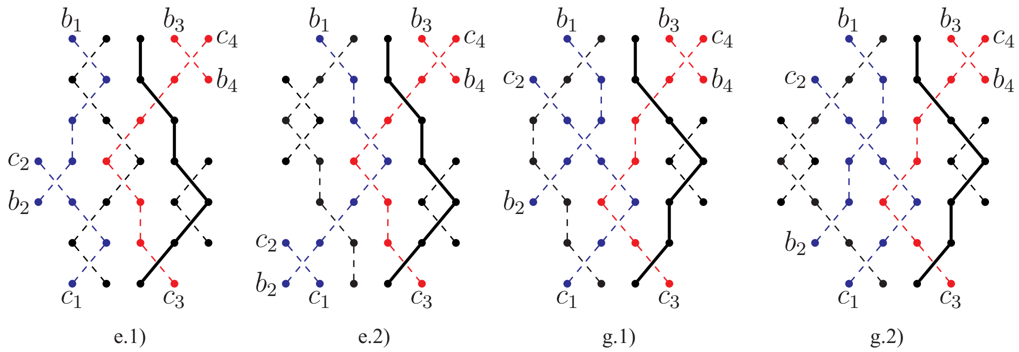


Рис. 1.14. *Варианты следов для 7-повторов.*

В случаях $h)$ и $j)$ (см. рис. 1.15), соседние слева и справа следы являются неподходящими (не получается соединить точки b_1 и c_1; b_3 и c_3). Как и в случаях $l)$ и $n)$ (см. рис. 1.15), невозможно соединить точки b_1 и c_1; b_2 и c_2 так, чтобы не получить фрагмент соответствующий наличию 3-повтора в слове.

Все следы, начинающиеся на $\vert$, разобраны. Теперь займемся следами, которые начинаются и заканчиваются на один и тот же наклонный отрезок, б.о.о. на $\diagdown$, а на втором месте у них стоит $\vert$. Таких следов всего два с точностью до симметрии (см. рис. 1.16 случаи e'), f')). След $f')$ – это в точности след 3, являющийся циклическим

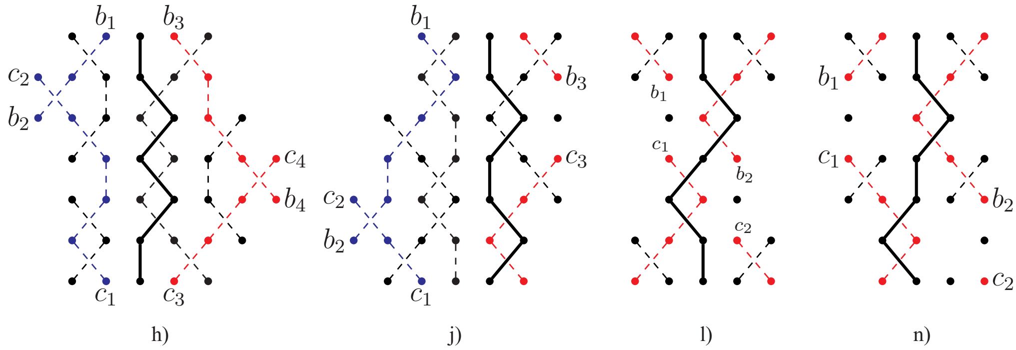


Рис. 1.15. *Варианты следов для 7-повторов.*

сдвигом следа f, а след e') – это с точностью до симметрии циклический сдвиг следов e), g), m) и o). Все возможные фрагменты с участием следа f') перечислены в пункте 2 условия теоремы. А след e') имеет нечетное число пересечений с соседними следами, поэтому не подходит по лемме 1.3.

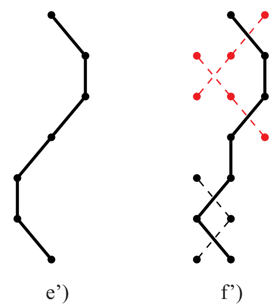


Рис. 1.16. *Варианты следов для 7-повторов.*

($\Leftarrow$). Пусть в $\mathrm{Cyl}(\mathbf{w})$ встретился один из перечисленных фрагментов. Это эквивалентно тому, что слово $\mathbf{w}$ имеет подслово u вида

$$a_1a_2a_3a_4a_5a_6a_7\underbrace{\cdots}_{6k-7}a_1a_2a_3a_4a_5a_6a_7,$$

которое по определению является 7-повтором. Действительно, $\exp(u){=}(6k{+}7)/(6k){>}k/(k{-}1)$ и оно не содержит других повторов меньшей длины, так как 1-, 2-, 3- и 6-повторы запрещены по условию, равномерные 4- и 5-повторы не существуют по теореме 1.2, а неравномерных 4- и 5-повторов над алфавитом Σ_k для $k \geq 12$ нет по лемме 1.1. $\square$

Замечание 1.9. *Для слова* $\mathbf{w}$ *из условия теоремы 1.4* $\mathrm{Bin}(\mathbf{w})$ *содержит подслово одного из следующих видов:*

$$11011011\underbrace{\cdots}_{k-7}01101101\underbrace{\cdots}_{k-9}1101101101\underbrace{\cdots}_{k-9}011011011\underbrace{\cdots}_{k-10}1101101101\underbrace{\cdots}_{k-9}0110110,$$

$$01101101\underbrace{\cdots}_{k-9}011011011\underbrace{\cdots}_{k-10}1101101101\underbrace{\cdots}_{k-9}011011011\underbrace{\cdots}_{k-10}1101101101\underbrace{\cdots}_{k-9}01101101,$$

$$01101101\underbrace{\cdots}_{k-9}011011011\underbrace{\cdots}_{k-10}1101101101\underbrace{\cdots}_{k-9}011011011\underbrace{\cdots}_{k-8}01101101\underbrace{\cdots}_{k-8}11011011,$$

$$01101101 \underbrace{\cdots}_{k-9} 1101101101 \underbrace{\cdots}_{k-9} 011011011 \underbrace{\cdots}_{k-10} 1101101101 \underbrace{\cdots}_{k-9} 01101101 \underbrace{\cdots}_{k-9} 11011011,$$

$$01101101 \underbrace{\cdots}_{k-9} 1101101101 \underbrace{\cdots}_{k-9} 011011011 \underbrace{\cdots}_{k-10} 1101101101 \underbrace{\cdots}_{k-9} 011011011 \underbrace{\cdots}_{k-8} 0110110,$$

$$11011011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 011011011 \underbrace{\cdots}_{k-10} 1101101101 \underbrace{\cdots}_{k-9} 011011011 \underbrace{\cdots}_{k-8} 01101101,$$

$$01101101 \underbrace{\cdots}_{k-8} 01011011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-8} 01011011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-8} 0101101,$$

$$0101101 \underbrace{\cdots}_{k-7} 01101101 \underbrace{\cdots}_{k-8} 01011011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-8} 01011011 \underbrace{\cdots}_{k-8} 0110110,$$

$$110101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-9} 110101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-9} 110101101 \underbrace{\cdots}_{k-8} 0110101,$$

$$0101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-9} 110101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-9} 110101101 \underbrace{\cdots}_{k-8} 01101011,$$

$$01011011 \underbrace{\cdots}_{k-9} 0110101101 \underbrace{\cdots}_{k-11} 110101101101 \underbrace{\cdots}_{k-11} 01101011011 \underbrace{\cdots}_{k-10} 0101101101 \underbrace{\cdots}_{k-10} 11010110,$$

$$11010101 \underbrace{\cdots}_{k-7} 01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 0110110,$$

$$0101011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 01101101,$$

$$010101101 \underbrace{\cdots}_{k-10} 0110110110 \underbrace{\cdots}_{k-11} 110101011011 \underbrace{\cdots}_{k-11} 01101101101 \underbrace{\cdots}_{k-10} 010101101 \underbrace{\cdots}_{k-9} 11011011,$$

$$0110101 \underbrace{\cdots}_{k-8} 110101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-9} 110101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-9} 11010110,$$

$$01101011 \underbrace{\cdots}_{k-9} 110101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-9} 110101101 \underbrace{\cdots}_{k-8} 01101011 \underbrace{\cdots}_{k-7} 0101101,$$

$$110101101 \underbrace{\cdots}_{k-9} 0101101101 \underbrace{\cdots}_{k-11} 01101011011 \underbrace{\cdots}_{k-12} 110101101101 \underbrace{\cdots}_{k-11} 0110101101 \underbrace{\cdots}_{k-9} 01011011,$$

$$01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 11010101,$$

$$01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110101011 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-7} 0101011,$$

$$11011011 \underbrace{\cdots}_{k-8} 010101101 \underbrace{\cdots}_{k-10} 0110110110 \underbrace{\cdots}_{k-11} 110101011011 \underbrace{\cdots}_{k-11} 01101101101 \underbrace{\cdots}_{k-10} 01010110,$$

$$110110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 0110110,$$

$$0110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 01101101,$$

$$01101011 \underbrace{\cdots}_{k-9} 0110110101 \underbrace{\cdots}_{k-11} 110110101101 \underbrace{\cdots}_{k-11} 01101101011 \underbrace{\cdots}_{k-10} 0110101101 \underbrace{\cdots}_{k-10} 11011010,$$

$$011010101 \underbrace{\cdots}_{k-10} 0110110110 \underbrace{\cdots}_{k-11} 110110101011 \underbrace{\cdots}_{k-11} 01101101101 \underbrace{\cdots}_{k-10} 011010101 \underbrace{\cdots}_{k-9} 11011011,$$

$$01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 11011010,$$

$$01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-9} 110110101 \underbrace{\cdots}_{k-8} 01101101 \underbrace{\cdots}_{k-7} 0110101,$$

$$110110101 \underbrace{\cdots}_{k-9} 0110101101 \underbrace{\cdots}_{k-11} 01101101011 \underbrace{\cdots}_{k-12} 110110101101 \underbrace{\cdots}_{k-11} 0110110101 \underbrace{\cdots}_{k-9} 01101011,$$

$$11011011 \underbrace{\cdots}_{k-8} 011010101 \underbrace{\cdots}_{k-10} 0110110110 \underbrace{\cdots}_{k-11} 110110101011 \underbrace{\cdots}_{k-11} 01101101101 \underbrace{\cdots}_{k-10} 01101010,$$

$$0110101 \underbrace{\cdots}_{k-7} 0101101 \underbrace{\cdots}_{k-7} 0110101 \underbrace{\cdots}_{k-7} 0101101 \underbrace{\cdots}_{k-7} 0110101 \underbrace{\cdots}_{k-7} 0101101,$$

$$0101101 \underbrace{\cdots}_{k-7} 0110101 \underbrace{\cdots}_{k-7} 0101101 \underbrace{\cdots}_{k-7} 0110101 \underbrace{\cdots}_{k-7} 0101101 \underbrace{\cdots}_{k-7} 0110101,$$

$$010110101\underbrace{\cdots}_{k-10}011010110\underbrace{\cdots}_{k-10}11010110101\underbrace{\cdots}_{k-10}0110101101\underbrace{\cdots}_{k-9}010110101\underbrace{\cdots}_{k-9}11010110,$$
$$010101101\underbrace{\cdots}_{k-10}011011010\underbrace{\cdots}_{k-10}11010101101\underbrace{\cdots}_{k-10}0110110101\underbrace{\cdots}_{k-9}010101101\underbrace{\cdots}_{k-9}11011010,$$
$$110101101\underbrace{\cdots}_{k-9}010110101\underbrace{\cdots}_{k-10}011010110\underbrace{\cdots}_{k-10}11010110101\underbrace{\cdots}_{k-10}0110101101\underbrace{\cdots}_{k-9}01011010,$$
$$110110101\underbrace{\cdots}_{k-9}010101101\underbrace{\cdots}_{k-10}011011010\underbrace{\cdots}_{k-10}11010101101\underbrace{\cdots}_{k-10}0110110101\underbrace{\cdots}_{k-9}01010110.$$

Полученное описание множеств Q_3–Q_7 (см. теоремы 1.1, 1.2, 1.3 и 1.4) свидетельствует о том, что между строением множеств Q_m для разных значений m не прослеживается никакой закономерности. Поэтому не представляется возможным предложить некую общую схему для их построения. К тому же с ростом m мощность множества Q_m, а также сложность входящих в него структур, быстро возрастает, что делает нецелесообразным получение описания структуры m-повторов для $m > 7$. Однако естественным образом возникает вопрос, как сказывается «одинаковость» в строении равномерных m-повторов на количественных характеристиках граничных языков.

1.2. Оценка количества граничных слов

Как уже упоминалось ранее, для оценки количества слов длины n в языке L используется комбинаторная сложность $C_L(n)$. Но сама по себе эта величина не очень информативна. Лучше оценивать не количество слов, а скорость их прироста с увеличением длины, т.е. вычислять индекс роста $\alpha(L)$. Но граничные языки не являются регулярными, поэтому для получения верхних оценок для их индексов роста мы будем использовать последовательность индексов роста языков $T_k^{(m)}$ (подробнее о методе регулярных приближений см. в [44]).

1.2.1. Верхние оценки для индексов роста граничных языков

Обозначим через $\hat{T}_k^{(m)}$ – множество всех подслов Z-слов $\mathbf{w}_k^{(m)} \in \Sigma_k^Z$, не содержащих r-повторы для $r \le m$. Из [46] известно, что индексы роста языков $\hat{T}_k^{(m)}$ и $T_k^{(m)}$ совпадают. При переходе от слов к их бинарным кодам комбинаторная сложность уменьшается в $k!$ раз, при этом индекс роста не изменится. Обозначим через $P_k^{(m)}$ множество всех бинарных кодов слов языка $\hat{T}_k^{(m)}$. Тогда справедливо следующее соотношение между индексами роста описанных языков:

$$\alpha(P_k^{(m)}) = \alpha(\hat{T}_k^{(m)}) = \alpha(T_k^{(m)}) \xrightarrow[m\to\infty]{} \alpha(T_k) \tag{1.1}$$

Получим верхние оценки для индексов роста граничных языков, используя стандартный алгоритм, который применяется не к самим граничным языкам, а к соответствующим бинарным языкам. Как происходит вычисление индекса роста для языков

$P_k^{(m)}$ продемонстрируем на примере $m = 2$. В этом случае для любого k антисловарь состоит из 2 слов $\{00, 111\}$. Построим автомат, распознающий язык $P_k^{(2)}$:

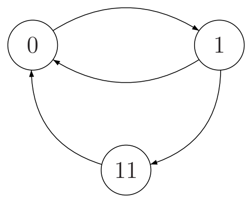


Рис. 1.17. *Автомат, распознающий язык $P_k^{(2)}$.*

Как известно, количество слов длины n в языке равно количеству маршрутов длины n в распознающем этот язык автомате. Используя этот факт, легко вычислить значение индекса роста по автомату, распознающему данный язык (например, составив для числа путей рекуррентное соотношение $C(n) = C(n-2) + C(n-3)$). Так для $m = 2$ получаем, что для любого $k \geq 5$ $\alpha(T_k^{(2)}) \approx 1.324718$.

Для построения антисловарей языков $P_k^{(m)}$ для $m = 3, 6, 7$ будем использовать полученное в теореме 1.1 и замечаниях 1.8 и 1.9 описание структуры характеристических слов запрещенных повторов в языках $\hat{T}_k^{(3)}$ ($k \geq 5$), $\hat{T}_k^{(6)}$ ($k \geq 10$) и $\hat{T}_k^{(7)}$ ($k \geq 12$), соответственно (ограничения на k вытекают из леммы 1.1). Затем по алгоритму из работы [14] для каждого языка строим распознающий его автомат. А по автомату вычисляем индекс роста с любой наперед заданной степенью точности, используя алгоритм 1 из работы [47]. В таблице 1.1 приведены полученные значения индексов роста языков $T_k^{(3)}$, $T_k^{(6)}$ и $T_k^{(7)}$ и связь между ними.

Анализируя данные значения, можно сделать несколько выводов:

1. значения индексов роста языков $T_k^{(3)}$ с ростом k быстро стабилизируются, что позволяет предположить существование предела $\lim\limits_{k\to\infty} \alpha(T_k^{(3)})$ и его примерное значение 1.242096777;

2. величина $\delta_{36} = \alpha(T_k^{(6)}) - \alpha(T_k^{(3)})$, описывающая вклад 6-повторов в индекс роста языка T_k, с ростом k стабилизируется в районе значения $4.3 \cdot 10^{-5}$, что значительно меньше вклада 3-повторов;

3. величина $\delta_{67} = \alpha(T_k^{(7)}) - \alpha(T_k^{(6)})$, описывающая вклад 7-повторов в индекс роста языка T_k, с ростом k стабилизируется в районе значения $3.6 \cdot 10^{-6}$, что позволяет предположить, что с ростом m влияние m-повторов на индекс роста языка не зависит от мощности алфавита и быстро убывает;

Естественным образом напрашивается следующая гипотеза:

Гипотеза 1.1. *Для любого $m \geq 3$ существует предел $\lim\limits_{k\to\infty} \alpha(T_k^{(m)}) = \alpha^{(m)}$.*

Таблица 1.1. *Верхние оценки индексов роста граничных языков;* $\delta_{36} = \alpha(T_k^{(6)}) - \alpha(T_k^{(3)})$, $\delta_{67} = \alpha(T_k^{(7)}) - \alpha(T_k^{(6)})$.

k	$\alpha(T_k^{(3)})$	$\alpha(T_k^{(6)})$	δ_{36}	$\alpha(T_k^{(7)})$	δ_{67}
5	1,2149529259				
6	1,2525850156				
7	1,2419872870				
8	1,2365057034				
9	1,2468689328				
10	1,2393804408	1,2393569373	$2{,}35\cdot10^{-5}$		
11	1,2426566183	1,2426131641	$4{,}35\cdot10^{-5}$		
12	1,2429286087	1,2428815613	$4{,}70\cdot10^{-5}$	1,2428793902	$2{,}17\cdot10^{-6}$
13	1,2409226614	1,2408787655	$4{,}39\cdot10^{-5}$	1,2408729371	$5{,}83\cdot10^{-6}$
14	1,2428289279	1,2427804533	$4{,}85\cdot10^{-5}$	1,2427767435	$3{,}71\cdot10^{-6}$
15	1,2418774954	1,2418379960	$3{,}95\cdot10^{-5}$	1,2418340134	$3{,}98\cdot10^{-6}$
16	1,2420000807	1,2419572202	$4{,}29\cdot10^{-5}$	1,2419536019	$3{,}62\cdot10^{-6}$
17	1,2423240470	1,2422793681	$4{,}47\cdot10^{-5}$	1,2422759823	$3{,}39\cdot10^{-6}$
18	1,2418973834	1,2418553246	$4{,}21\cdot10^{-5}$	1,2418514866	$3{,}84\cdot10^{-6}$
19	1,2421895750	1,2421451742	$4{,}44\cdot10^{-5}$	1,2421416402	$3{,}53\cdot10^{-6}$
20	1,2420949436	1,2420520022	$4{,}29\cdot10^{-5}$	1,2420483333	$3{,}67\cdot10^{-6}$
21	1,2420552103	1,2420123323	$4{,}29\cdot10^{-5}$		
22	1,2421449456	1,2421012197	$4{,}37\cdot10^{-5}$		
...	...				
58	1,2420967776				
59	1,2420967762				
60	1,2420967771				

Так, например, $\alpha^{(3)} \approx 1.242096777$. Причем раз вклад 6- и 7-повторов невелик по сравнению с 3-повторами и предположительно с ростом m влияние m-повторов убывает, можно сделать предположение, что индекс роста граничных языков T_k не сильно отличается от индекса роста их приближений $T_k^{(3)}$.

Гипотеза 1.2. *Последовательность* $\big(\alpha(T_k)\big)_{k\geq 5}$ *индексов роста граничных языков стремится к числу* $\alpha \approx 1.242$ *при* $k \to \infty$.

Отметим, что гипотеза 1.2 является усилением экспоненциальной гипотезы, обсуждавшейся на стр. 13.

1.2.2. Развёртки цилиндров

Теперь попробуем использовать для вычисления индекса роста найденное соответствие между m-повторами и наличием определенных двумерных конструкций в цилиндрических представлениях. Как уже упоминалось ранее, индексы роста языков $T_k^{(m)}$ и $\hat{T}_k^{(m)}$ совпадают. Обозначим через $\mathrm{Cyl}_k^{(m)}$ множество цилиндрических представлений слов из $\hat{T}_k^{(m)}$.

Отметим, что мы не случайно работаем с подсловами Z-слова, т.е. с языком $\hat{T}_k^{(m)}$. Представим, что у нас есть некоторое слово w. Построение его цилиндрического представления начинается с буквы $w[k+1] = a$ (буквы из верхнего витка не с чем соединять). По замечанию 1.2 либо это первое появление буквы a в слове w, либо она уже появлялась на 1 или 2 позиции. В двух последних случаях нам ничего не мешает закодировать букву a отрезками $|$ или $/$, соответственно. А вот с первым случаем дело обстоит сложнее. Если w – «обычное» слово, то в его цилиндрическом коде появится «дырка». Но если w является подсловом Z-слова, то оно имеет продолжение влево любой длины. Причем начинается любое продолжение с буквы a (дальше находится она не может по замечанию 1.2). С учетом этого мы кодируем букву $w[k+1]$ отрезком $\backslash$.

Рассмотрим произвольное слово $W \in \mathrm{Cyl}_k^{(m)}$ длины l. Количество слов длины l задается значением комбинаторной сложности $C_{\mathrm{Cyl}_k^{(m)}}(l)$. Теперь выделим множество слов в языке $\mathrm{Cyl}_k^{(m)}$, которые имеют длину nk, где $n \in \mathbb{N}$. Цилиндрические слова с целым числом витков можно представить в виде двумерных слов. Для этого мы должны цилиндрическое слово «разрезать» по образующей цилиндра как показано на рис. 1.18 и развернуть на плоскости. Тогда получим прямоугольную таблицу с символами алфавита $\Delta = \{|, /, \backslash\}$, которую будем называть *развёрткой цилиндра*.

Пусть $w \in \hat{T}_k^{(m)}$ такое, что $|w| = (n+1)k$. Тогда его развёртка цилиндра имеет размер $n \times k$, а в i-ой строке на j-ом месте в ней стоит отрезок, соединяющий букву $w[ik+j] = a$ с ее предыдущим вхождением. Например, для слова $w = abcdaecbdeacdba$ над алфавитом Σ_5 развёртка цилиндра имеет вид:

$\backslash$	$/$	$\backslash$	$\vert$	$/$
$\backslash$	$\vert$	$/$	$\backslash$	$/$

Найденные в теоремах 1.1, 1.3 и 1.4 закономерности переносятся и на развёртки цилиндров, только с небольшими изменениями.

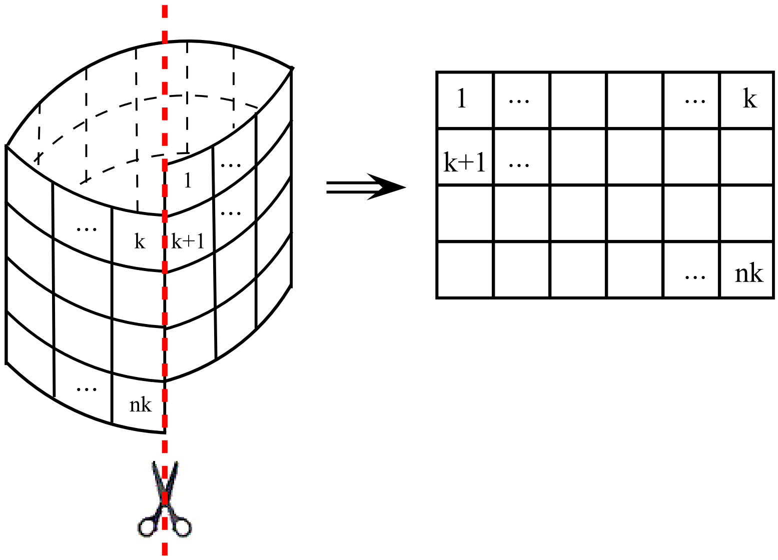


Рис. 1.18. *Переход к развёрткам цилиндров.*

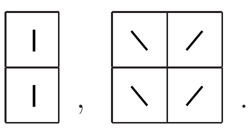
1.2.3. Случай m=3

Остановимся подробнее на случае $m = 3$. Наш выбор основан не только на том, что цилиндрическое представление 3-повторов имеет более простую структуру. Согласно результатам пункта 1.2.1, языки $T_k^{(3)}$ являются наиболее «ценным» приближением языков T_k. Действительно:

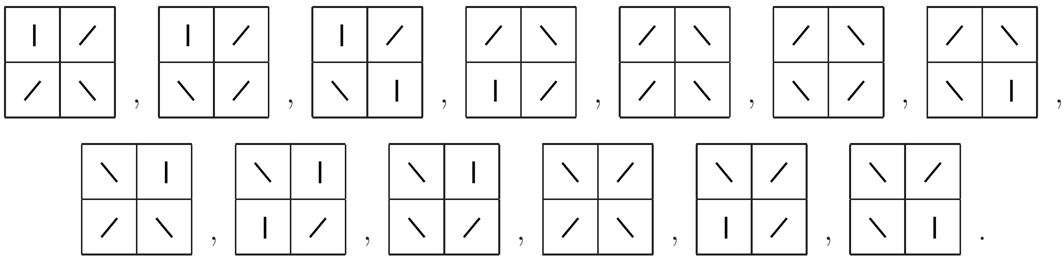
- не существует равномерных 4- и 5-повторов;
- вклад 6- и 7-повторов, как показывает таблица 1.1, мал по сравнению с вкладом 3-повторов;
- с ростом m влияние m-повторов на индекс роста предположительно быстро убывает.

Подмножество языка $\mathrm{Cyl}_k^{(3)}$, состоящее из слов с целым числом витков, обозначим Cyl'_k. Причем на элементы языка Cyl'_k будем смотреть сразу как на развёртки цилиндров. Количество слов в Cyl'_k длины nk (т.е. количество развёрток цилиндров размера $n{\times}k$) обозначим $C'(n,k)$.

Замечание 1.10. *Пусть* $w \in \hat{T}_k^{(3)}$ *таково, что* $\mathrm{Cyl}(w) \in \mathrm{Cyl}'_k$. *Из теоремы 1.1 следует, что наличие 3-повтора в слове w эквивалентно появлению в его развёртке цилиндра следующих подслов:*



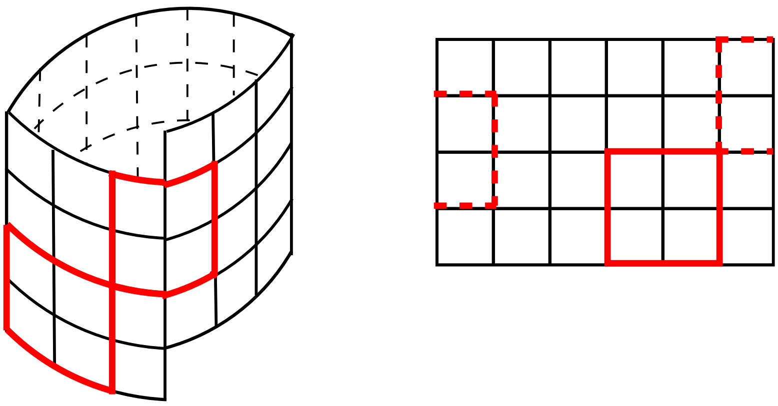
Объединяя замечания 1.10 и 1.5, получим полный список запрещенных подслов для развёрток цилиндров языка Cyl'_k. Отметим, что каждый запрет целиком укладывается в блок 2×2. Таким образом, разрешенных блоков размера 2×2 остается всего 13 из $3^4 = 81$ (состоящее из них множество обозначим Q'_3):



Определим двумерный язык $D \subseteq \Delta^{**}$ следующим образом: двумерное слово W принадлежит языку D тогда и только тогда, когда любое его подслово размера 2×2 попадает в множество Q'_3. В [21] языки, определяемые подобным образом, т.е. списком разрешенных подслов заданного размера, названы *локальными*. Через D_k обозначим множество слов языка D ширины k.

Замечание 1.11. *Любое слово W языка* Cyl'_k *принадлежит множеству* D_k. Действительно, любое слово в Cyl'_k имеет ширину k и по замечаниям 1.10 и 1.5 не содержит подслов размера 2×2 не из множества Q'_3.

Однако обратное утверждение неверно. Дело в том, что наличие разрешенных подслов размера 2×2 у двумерного слова проверяется только внутри самого слова, а у развёрток цилиндров этому условию удовлетворяет также стык – место бывшего «разреза» цилиндра (см. рис. 1.19).



a) Развёртка цилиндра

b) Двумерное слово

Рис. 1.19. *Отличие развёрток цилиндров от двумерных слов.*

1.2.4. Индекс роста двумерного языка

Язык D является факториальным, так как если слово $W \in D$ не содержит блоков не из множества Q'_3, то и любое его подслово тем более. Следовательно, можно вычислить индекс роста языка D как двойной предел: $\alpha(D) = \lim\limits_{n,k\to\infty} (C_D(n,k))^{1/nk}$. В дальнейшем комбинаторную сложность языка D будем обозначать через $C(n,k)$.

Эффективного алгоритма для вычисления индекса роста двумерного языка не известно. Но любая подпоследовательность последовательности $\left\{(C(n,k))^{1/nk}\right\}_5^\infty$ сходится к тому же пределу $\alpha(D)$ (см. [2]). Можно рассмотреть, например, диагональную подпоследовательность $\left\{(C(n,n))^{1/n^2}\right\}_5^\infty$, в которой фигурирует всего одна переменная. Для нахождения пределов иногда бывает полезна следующая теорема, доказательство которой можно найти в [2].

Теорема 1.5 (Теорема Штольца)**.** *Пусть $(a_n)_{n\geq 1}$ и $(b_n)_{n\geq 1}$ – две числовые последовательности, причем $(b_n)_{n\geq 1}$ неограничена и строго возрастает с некоторого момента. Тогда*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{(a_{n+1}-a_n)}{(b_{n+1}-b_n)}$$

при условии, что предел справа существует.

Если прологарифмировать выбранную диагональную последовательность и положить $\frac{a_n}{b_n} = \frac{\log C(n,n)}{n^2}$, то, дважды применяя теорему Штольца, получим:

$$\log \alpha(D) = \lim_{n\to\infty} \frac{\log C(n,n)}{n^2} = \lim_{n\to\infty} \frac{\log C(n,n) - \log C(n-1,n-1)}{2n-1} =$$
$$= \lim_{n\to\infty} \frac{\log C(n,n) + \log C(n-2,n-2) - 2\log C(n-1,n-1)}{2}$$

После потенцирования получим три последовательности, которые (при существовании пределов двух последних) сходятся к одному пределу $\alpha(D)$. При помощи компьютера нам удалось сосчитать количество двумерных слов размера $n\times n$ в языке D до $n = 36$ включительно и вычислить несколько первых членов этих последовательностей (см. таблицу 1.2). Судя по полученным результатам первые две последовательности являются монотонными и сходятся очень медленно (вторая чуть быстрее), а вот третья последовательность осциллирует вокруг значения ≈ 1.242097, очень похожего на предсказанное ранее значение $\alpha^{(3)}$.

Так как существование двух последних пределов формально не доказано, к предложенному методу нельзя относиться как к способу вычисления индекса роста двумерного локального языка. Однако, полученные численные оценки являются сильным свидетельством в пользу того, что значение $\alpha(D)$ действительно очень близко к $\alpha^{(3)}$. В качестве альтернативного способа оценки индекса роста языка D можно рассмотреть двумерные слова специальной формы, изображенной на рис. 1.20, и оценить, как

Таблица 1.2. *Оценки индекса роста двумерного языка D.*

n	$(C(n,n))^{1/n^2}$	$\left(\frac{C(n,n)}{C(n-1,n-1)}\right)^{1/(2n-1)}$	$\frac{(C(n,n)C(n-2,n-2))^{1/2}}{C(n-1,n-1)}$
3	1.627251	1.438233	1.191687
4	1.525034	1.402991	1.318617
5	1.464419	1.362547	1.229958
...	...	...	...
31	1.275217	1.258792	1.242096
32	1.274167	1.258259	1.242097
33	1.273181	1.257758	1.242094
34	1.272255	1.257288	1.242097
35	1.271382	1.256845	1.242098
36	1.270558	1.256427	1.242096

меняется количество таких слов при добавлении одной клетки к последнему столбцу. Полученные результаты подтверждают найденную раннее оценку. Так, например, для $n = 100, r = 34, l = 50$ добавление клетки дает приращение числа двумерных слов примерно в 1.24209673 раза.

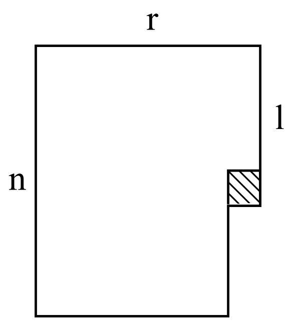


Рис. 1.20. *Вычисление индекса роста двумерного языка.*

1.2.5. Автоматы $\mathcal{A}_k$ и $\mathcal{B}_k$

У языков Cyl'_k и D_k много общего в структуре слов. Но можно ли как-то связать их численные характеристики?

На язык D_k можно смотреть как на одномерный регулярный язык над алфавитом Δ^k с индексом роста $\alpha(D_k)^k$. Тогда индекс роста языка D_k над алфавитом Δ вполне естественно определить как $\alpha(D_k) = \lim\limits_{n\to\infty}(C(n,k))^{1/nk}$, причем этот предел существует. Следовательно, $\alpha(D) = \lim\limits_{k\to\infty}\alpha(D_k) = \lim\limits_{k\to\infty}\lim\limits_{n\to\infty}(C(n,k))^{1/nk}$, так как последний предел является повторным пределом существующего двойного (про связь между двойными и повторными пределами можно найти в [2]).

Для D_k как языка над алфавитом Δ^k можно построить распознающий его автомат $\mathcal{A}_k$:

(A1) вершинами автомата $\mathcal{A}_k$ являются слова размера $1{\times}k$ языка D_k (они совпадают со словами длины k языка Cyl'_k);

(A2) ребро $u \to v$ существует тогда и только тогда, когда слово $\begin{smallmatrix} u \\ v \end{smallmatrix}$ размера $2 \times k$ принадлежит языку D_k, такое ребро будем помечать словом v;

(A3) каждая вершина является как начальной, так и заключительной.

Построенный автомат является так называемым *однозначным недетерминированным* автоматом (начальная вершина, конечная вершина и метка однозначно определяют маршрут в автомате). Для таких автоматов функция, возвращающая количество маршрутов длины n, имеет тот же индекс роста, что и комбинаторная сложность распознаваемого языка (см. [47]). Обозначим через $P_k(u,n)$ количество путей длины n в автомате $\mathcal{A}_k$, начинающихся в вершине u. Тогда $P_k(n) = \sum P_k(u,n)$ – количество всех маршрутов длины n в автомате $\mathcal{A}_k$. А значит $P_k(n) = C(n{+}1, k)$.

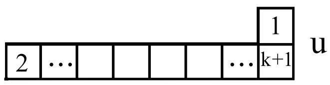
На язык $\mathrm{Cyl}_k^{(3)}$ можно смотреть как на одномерный регулярный язык «размотанных» цилиндрических кодов слов языка $\hat{T}_k^{(3)}$. Причем, так как количества слов и их цилиндрических кодов отличаются в $k!$ раз, то индексы роста этих языков совпадают:

$$\alpha(\mathrm{Cyl}_k^{(3)}) = \lim_{n\to\infty} (C_{\mathrm{Cyl}_k^{(3)}}(n))^{1/n} = \lim_{n\to\infty} (C_{\hat{T}_k^{(3)}}(n))^{1/n} = \alpha(\hat{T}_k^{(3)}) = \alpha(T_k^{(3)}) = \alpha^{(3)} \quad (1.2)$$

Последовательность $\left\{(C'(n,k))^{1/nk}\right\}_5^\infty$ является подпоследовательностью сходящейся последовательности $\left\{(C_{\mathrm{Cyl}_k^{(3)}}(n))^{1/n}\right\}_5^\infty$, а значит сходится к тому же пределу:

$$\lim_{n\to\infty} (C'(n,k))^{1/nk} = \alpha(\mathrm{Cyl}_k^{(3)}) = \alpha^{(3)} \quad (1.3)$$

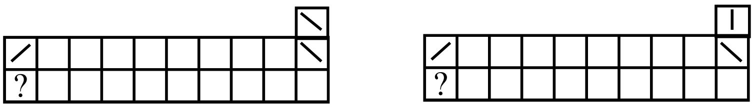
Для языка $\mathrm{Cyl}_k^{(3)}$ построим так называемый граф Рози порядка $k{+}1$ (см. [39]); обозначим его через $\mathcal{R}_k$. Вершинами этого графа являются слова языка $\mathrm{Cyl}_k^{(3)}$ длины $k{+}1$:



Ребро $u \to v$ существует тогда и только тогда, когда найдется слово $w \in \mathrm{Cyl}_k^{(3)}$ длины $k{+}2$ таково, что u является его префиксом, а v – суффиксом. Каждое ребро $u \to v$ будем помечать последней буквой слова v, тогда автомат $\mathcal{R}_k$ будет распознавать язык $\mathrm{Cyl}_k^{(3)}$ и может быть использован для вычисления его индекса роста.

Теперь рассмотрим автомат $\mathcal{R}_k^k$. Вершины у него те же, что и у $\mathcal{R}_k$, а ребро $u \to v$ существует тогда и только тогда, когда найдется слово $w \in \mathrm{Cyl}_k^{(3)}$ длины $2k{+}1$ такое, что u является его префиксом, а v – суффиксом. Меткой такого ребра служит наибольший собственный суффикс слова v. Отметим, что можно вычислить все слова

v, в которые из u выходит ребро, зная только последние k символов слова u. Исключение составляет лишь случай, когда суффикс длины k слова u имеет вид как на рис. 1.21. В случае $b)$ суффикс слова v может начинаться с или , тогда как в случае $a)$ только с , чтобы избежать появления запрещенного фрагмента .



a) b)

Рис. 1.21. *Вычисление индекса роста двумерного языка.*

Построим автомат $\mathcal{B}_k$ следующим образом:

(B1) вершинами автомата $\mathcal{B}_k$ являются слова длины k языка Cyl'_k (т.е. наибольшие собственные суффиксы вершин автомата $\mathcal{R}_k^k$);

(B2) ребро $u \to v$ существует тогда и только тогда, когда в $\mathcal{R}_k^k$ существует ребро $au \to bv$ для некоторых $a, b \in \Delta$, с единственным исключением: если u имеет вид как на рис. 1.21, то v начинается ;

(B3) каждая вершина является как начальной, так и заключительной.

Обозначим через $P'_k(u, n)$ количество путей длины n в автомате $\mathcal{B}_k$, начинающихся в вершине u. Тогда $P'_k(n) = \sum P'_k(u, n)$ – количество всех путей длины n в автомате $\mathcal{B}_k$. Причем $P'_k(n) \leq C'(n{+}1, k)$ за счет того, что мы выкинули из рассмотрения слова языка Cyl'_k, имеющие на стыке фрагмент .

Замечание 1.12. *Если смотреть на $\mathcal{A}_k$ и $\mathcal{B}_k$ как на орграфы, то легко заметить, что $\mathcal{B}_k$ является суграфом $\mathcal{A}_k$.* Действительно, множества вершин у них совпадают, а любой маршрут в $\mathcal{B}_k$ определяет некоторое слово языка Cyl'_k, которое по замечанию 1.11 является словом языка D_k, а значит аналогичный маршрут существует и в $\mathcal{A}_k$.

Следствие 1.2. *Для любого $n \in \mathbb{N}$ и $k \geq 5$ справедливо следующее соотношение: $P'_k(n) \leq C'(n{+}1, k) \leq C(n{+}1, k) = P_k(n)$.*

1.2.6. Свойства автоматов $\mathcal{A}_k$ и $\mathcal{B}_k$

Пусть $k > 11$. Две вершины u и v (автомата $\mathcal{A}_k$ или $\mathcal{B}_k$) будем называть подобными и обозначать $u \sim v$, если у них совпадают суффиксы длины $k{-}11 > 0$. Легко проверить, что отношение подобия на множестве всех вершин автомата $\mathcal{A}_k$ ($\mathcal{B}_k$) является отношением эквивалентности.

Замечание 1.13. *Отношение подобия разбивает множество всех вершин автомата $\mathcal{A}_k$ ($\mathcal{B}_k$) на классы эквивалентности, мощность которых не зависит от k и не превышает константы $N = 28$.* Действительно, самыми «большими» являются классы, в которые входят вершины с суффиксами, начинающимися на ⟋. А их мощность равна количеству слов в языке $\mathrm{Cyl}_k^{(3)}$ длины 11, заканчивающихся на | и ⟍, что в точности равно 28.

Лемма 1.6. *Пусть $k > 12$. Тогда для любой вершины $u = u_1 \cdots u_k$ и для любого символа $a \in \Delta$ при условии, что $u_{12} \neq$ | или $a \neq$ |, существует ребро $u \to x$ в $\mathcal{B}_k$ такое, что 12-й символ в x равен a.*

Доказательство. Пусть $x = x_1 \cdots x_k$. Сначала покажем, что, если условие леммы выполняется для некоторого i-ого символа в слове x, то оно также будет справедливо и для каждого j-ого символа, где $i \leq j \leq k$. Достаточно проверить справедливость данного утверждения для $j = i+1$. Действительно, слово $\genfrac{}{}{0pt}{}{u}{x}$ состоит из блоков размера 2×2, входящих в множество Q_3', а значит каждая буква в нем зависит только от трех соседних с ней символов. В частности, x_{i+1} зависит от u_i, u_{i+1} и x_i. Поэтому, доказав для $i+1$-ого символа, по индукции можно дойти до любого j-ого символа. Для $u_i u_{i+1}$ существует 4 возможных варианта. на рис. 1.22 показано, что для любого из этих 4 вариантов, если x_i может быть любым символом алфавита Δ, удовлетворяющим условию леммы, то x_{i+1} тоже. Для этого просто перебираем возможные варианты для $x_i x_{i+1}$ так, чтобы блок $\genfrac{}{}{0pt}{}{u_i\ u_{i+1}}{x_i\ x_{i+1}}$ принадлежал множеству Q_3', а x_{i+1} принимал все допустимые по условию значения.

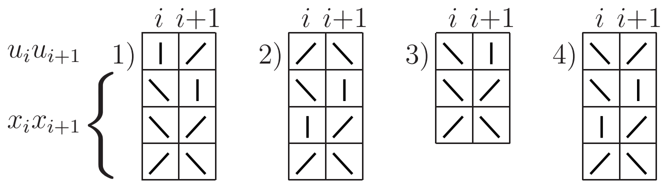


Рис. 1.22. *Возможные варианты для* $\genfrac{}{}{0pt}{}{u_i\ u_{i+1}}{x_i\ x_{i+1}}$.

Для каждой вершины u автомата $\mathcal{B}_k$ обозначим через i_u минимальный номер, обладающий свойством, что для любого символа $a \in \Delta$, кроме случая $a = u_{i_u} =$ |, существует вершина x такая, что в $\mathcal{B}_k$ есть дуга $u \to x$ и $x_{i_u} = a$. Если для каждой вершины u значение i_u будет не больше 12, то лемма будет доказана. Отметим, что нам необязательно перебирать все вершины u, достаточно рассмотреть лишь их всевозможные префиксы длины не больше 12, так как x_{i_u}-ый символ зависит только от u_{i_u-1}, u_{i_u} и x_{i_u-1}. И еще одно важное замечание: в цилиндрическом слове символ u_k предшествует символу x_1, а значит непосредственно влияет на его значение. Чтобы избежать рассмотрения символа u_k (и не делать наши рассуждения зависящими от k)

будем искать подходящий номер для каждого значения x_1 (разумеется кроме случая $u_1 = x_1 = \vert$), а максимум из них и будет искомым значением i_u.

на рис. 1.23 показаны возможные префиксы для слова u, начинающиеся с $\vert$ (случаи $1) - 3)$) и $\diagup$ (случаи $4) - 11)$). Для каждого префикса слова u и для любого допустимого x_1 строим возможные префиксы слова x пока не найдем позицию, удовлетворяющую условию леммы. Так, например, в случае $2)$ для $u^{(2)} = \vert \diagup \diagdown \diagup \diagdown \vert \cdots$ нужно рассмотреть $x_1 = \diagup$ и $x_1 = \diagdown$. В первом подслучае получим позицию номер 5, во втором – 6. Следовательно, $i_{u^{(2)}} = 6$. В случае $3)$ для слова $u^{(3)} = \vert \diagup \diagdown \diagup \diagdown \diagup \diagdown \cdots$ и $x_1 = \diagup$ искомая позиция также равна 5. За счет того, что первые пять символов в словах $u^{(2)}$ и $u^{(3)}$ совпадают, этот случай полностью повторяет случай $2)$ при $x_1 = \diagup$, поэтому на рис. 1.23 он не изображен. Аналогично в случае $6)$ при $x_1 = \diagup$ или $x_1 = \diagdown$ мы ссылаемся на случай $5)$; в случаях $8)$ и $9)$ при $x_1 = \vert$ или $x_1 = \diagdown$ ссылаемся на случай $7)$; а в случае $11)$ при $x_1 = \vert$ или $x_1 = \diagdown$ ссылаемся на случай $10)$.

Максимум среди i_u в случаях $1) - 11)$ равен 11, он достигается в случае $9)$. Если слово u начинается с $\diagdown$, то его подслово $u_2 \cdots u_{i_u}$ попадает в один из случаев $1) - 11)$. Следовательно, максимальное i_u в этом случае увеличится не более, чем на единицу. Отсюда для любой вершины u $i_u \leq 12$, что доказывает лемму. $\square$

Лемма 1.7. *Пусть $u \sim v$ и $u \to x$ – ребро в $\mathcal{A}_k$. Тогда существует ребро $v \to y$ в $\mathcal{B}_k$ такое, что $x \sim y$.*

Доказательство. Пусть $u = u_1 \cdots u_k$, $v = v_1 \cdots v_k$ и $x = x_1 \cdots x_k$, причем по условию $u_{12} \cdots u_k = v_{12} \cdots v_k$. Заметим, что если мы знаем только $u_{12} \cdots u_k$ и x_{12}, то можем восстановить все возможные варианты для $x_{13} \cdots x_k$ независимо от $u_1 \cdots u_{11}$ и $x_1 \cdots x_{11}$ (об этом уже говорилось в доказательстве леммы 1.6).

Теперь рассмотрим множество вершин $y = y_1 \cdots y_k$ таких, что в $\mathcal{B}_k$ существует ребро $v \to y$ и $y_{12} = x_{12}$. По лемме 1.6 данное множество не пусто. Так как $u_{12} \cdots u_k = v_{12} \cdots v_k$ и $x_{12} = y_{12}$, то множество возможных вариантов для $x_{13} \cdots x_k$ и $y_{13} \cdots y_k$ совпадают, а значит можно выбрать y, у которого $y_{13} \cdots y_k = x_{13} \cdots x_k$. Тогда получим, что $v \to y$ – ребро в $\mathcal{B}_k$ и $y_{12} \cdots y_k = x_{12} \cdots x_k$, т.е. y – искомая вершина. $\square$

1.2.7. Связь между индексами роста граничных языков и языка D

Договоримся через $\deg^+_{\mathcal{A}_k}(u)$ обозначать полустепень исхода вершины u в автомате $\mathcal{A}_k$.

Теорема 1.6. *Предел $\lim\limits_{k\to\infty} \alpha(T_k^{(3)})$ существует и равен индексу роста двумерного языка D, т.е. $\alpha^{(3)} = \alpha(D)$.*

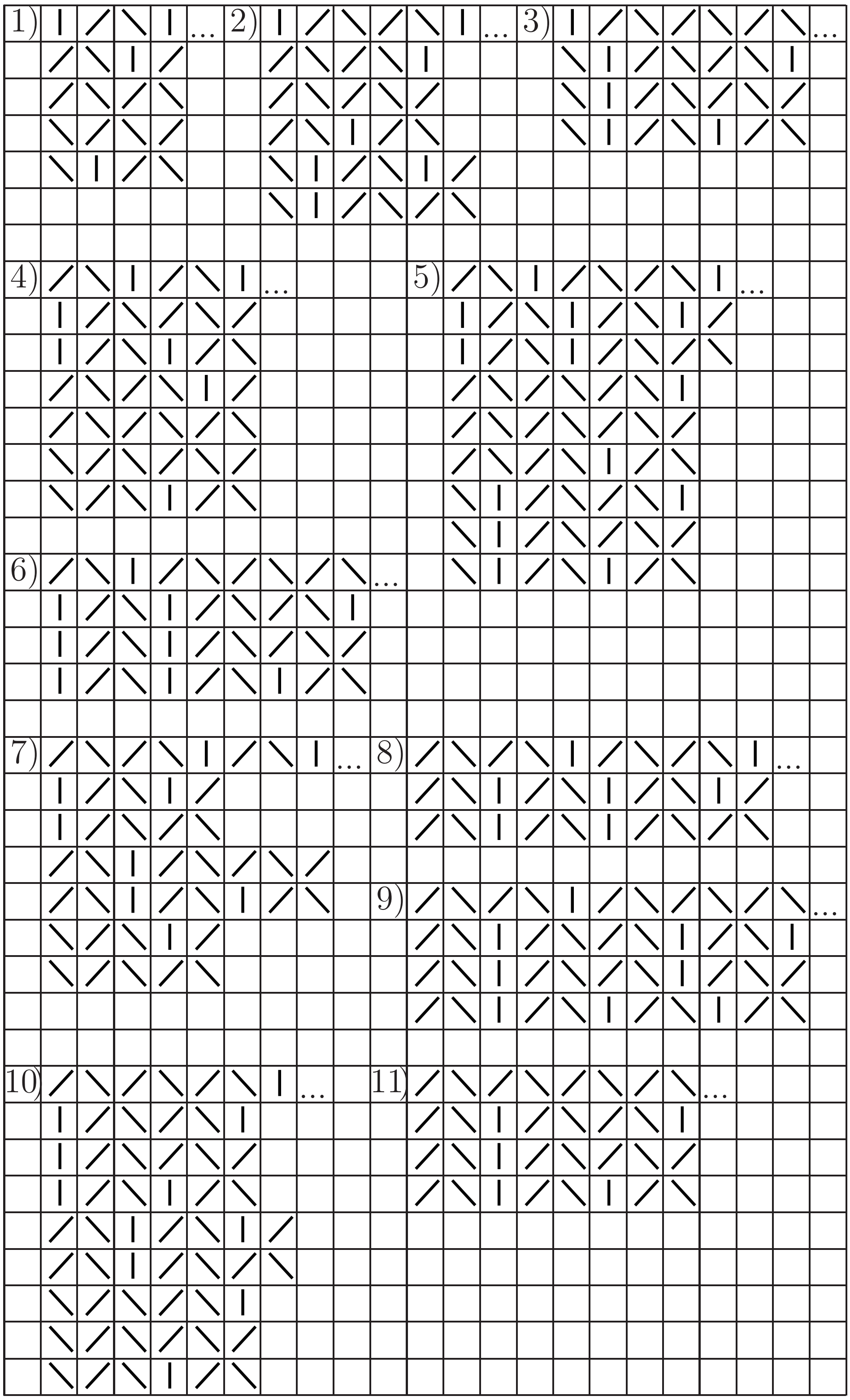


Рис. 1.23. *Возможные префиксы слов u и x. Показаны 11 префиксов для слова u (первая строка в каждом случае) и для каждого из них возможные префиксы слова x до искомой позиции, удовлетворяющей условию леммы 1.6.*

Доказательство. Напомним сначала то, что нам уже известно об индексах роста вышеупомянутых языков:

$$\alpha(T_k^{(3)}) = \alpha(\hat{T}_k^{(3)}) = \alpha(\mathrm{Cyl}_k^{(3)}) = \lim_{n\to\infty}(C'(n,k))^{1/nk}$$

$$\alpha(D) = \lim_{k\to\infty}\alpha(D_k) = \lim_{k\to\infty}\lim_{n\to\infty}(C(n,k))^{1/nk}$$

Таким образом, необходимо доказать, что предел $\lim\limits_{k\to\infty}\lim\limits_{n\to\infty}(C'(n,k))^{1/nk}$ существует и равен пределу $\lim\limits_{k\to\infty}\lim\limits_{n\to\infty}(C(n,k))^{1/nk}$. Идея доказательства заключается в том, чтобы получить двухсторонние оценки для отношения $\big(C'(n,k)/C(n,k)\big)^{1/nk}$. В силу следствия 1.2:

$$\frac{P_k'(n)}{P_k(n)} \le \frac{C'(n+1,k)}{C(n+1,k)} \le 1$$

Значит осталось получить нижнюю оценку для отношения $P_k'(n)/P_k(n)$. Для этого обратимся к автоматам $\mathcal{A}_k$ и $\mathcal{B}_k$.

Зафиксируем произвольную вершину u (вместе с ней фиксируем $k = |u|$) и построим для нее дерево $A_k(u)$ по следующим правилам:

- метками вершин дерева $A_k(u)$ являются вершины автомата $\mathcal{A}_k$;
- вершина u является меткой корня дерева $A_k(u)$;
- каждая вершина в $A_k(u)$ с меткой v имеет ровно $\deg^+_{\mathcal{A}_k}(v)$ детей, метками которых являются все вершины автомата $\mathcal{A}_k$, в которые идет дуга из вершины v.

Таким образом существует биекция между множеством вершин n-ого уровня в дереве $A_k(u)$ и множеством путей длины n в автомате $\mathcal{A}_k$, начинающихся в вершине u. А значит n-ый уровень в дереве $A_k(u)$ содержит в точности $P_k(u,n)$ вершин. Заменяя $\mathcal{A}_k$ на $\mathcal{B}_k$, аналогично построим для вершины u дерево $B_k(u)$, на n-ом уровне которого содержится $P_k'(u,n)$ вершин.

Индуктивно применяя лемму 1.7, получим, что метка каждой вершины уровня n в дереве $A_k(u)$ подобна метке некоторой вершины уровня n в дереве $B_k(u)$. Через $\mathsf{parent}(s)$ обозначим родителя вершины s. Построим отображение $\mu : A_k(u) \to B_k(u)$ по следующим правилам:

- образ корня дерева $A_k(u)$ равен корню дерева $B_k(u)$;
- если s – вершина n-ого уровня дерева $A_k(u)$, помеченная словом x, то $\mu(s)$ – это вершина уровня n дерева $B_k(u)$, помеченная словом $y \sim x$;
- $\mu(\mathsf{parent}(s)) = \mathsf{parent}(\mu(s))$.

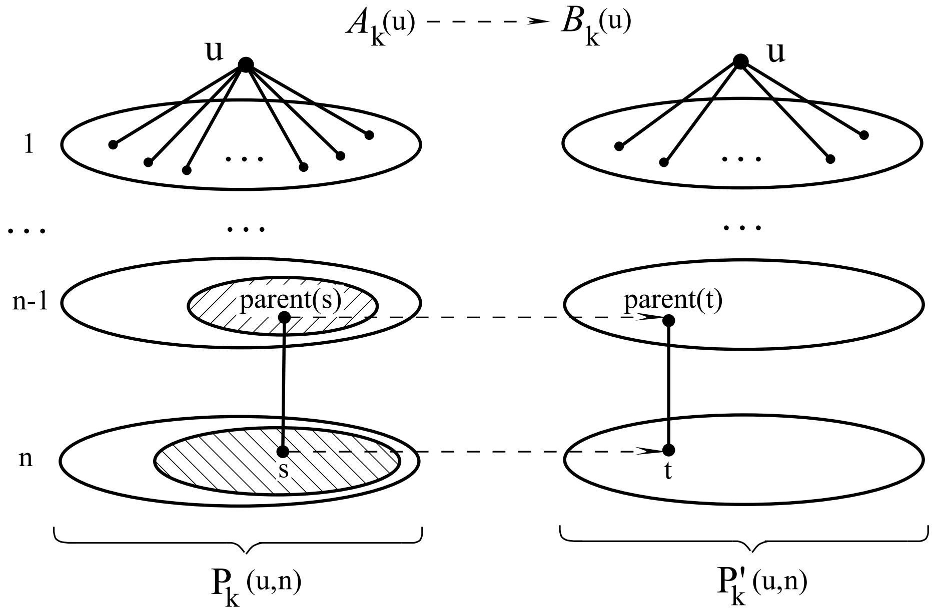


Рис. 1.24. *Доказательство теоремы 1.6. Пунктиром показано действие отображения μ.*

Существование такого отображения непосредственно вытекает из леммы 1.7 и структуры деревьев $A_k(u)$ и $B_k(u)$.

Рассмотрим произвольную вершину t уровня n в дереве $B_k(u)$ и оценим мощность множества $\mu^{-1}(t)$. Предположим, что $|\mu^{-1}(\mathsf{parent}(t))| = K$. Пусть s – вершина уровня n дерева $A_k(u)$ такая, что $\mu(s) = t$. Тогда согласно определению отображения μ $\mathsf{parent}(s) \in \mu^{-1}(\mathsf{parent}(t))$ (см. рис. 1.24). Так как все дети вершины $\mathsf{parent}(s)$ различны, то по замечанию 1.13 не более N из них могут служить прообразами вершины t относительно отображения μ. Таким образом $|\mu^{-1}(t)| \leq KN$. Если повторить эти рассуждения по индукции, то для корня получим $|\mu^{-1}(u)| = 1$, а для вершины $t : |\mu^{-1}(t)| \leq N^n$. Суммируя по всем вершинам уровня n, получим $P_k(u,n) \leq N^n P'_k(u,n)$. Объединяя результаты для всех вершин u, в итоге имеем: $P_k(n) \leq N^n P'_k(n)$.

Теперь перейдем снова к оценке отношения комбинаторных сложностей:

$$\frac{1}{N^n} \leq \frac{P'_k(n)}{P_k(n)} \leq \frac{C'(n+1,k)}{C(n+1,k)} \leq 1\,,$$
$$\left(\frac{1}{N^n}\right)^{1/(n+1)k} \leq \left(\frac{C'(n+1,k)}{C(n+1,k)}\right)^{1/(n+1)k} \leq 1.$$

В каждой части неравенства перейдем к пределу при $n \to \infty$:

$$\left(\frac{1}{N}\right)^{1/k} \leq \frac{\alpha(T_k^{(3)})}{\alpha(D_k)} \leq 1.$$

Наконец, устремим $k \to \infty$ и, применяя правило двух милиционеров, получим, что предел $\lim_{k\to\infty} \alpha(T_k^{(3)})/\alpha(D_k)$ существует и равен 1 (напомним, что N не зависит от k). Так как предел $\lim_{k\to\infty}\alpha(D_k)=\alpha(D)$ также существует, то получим:

$$\alpha(D)=\alpha(D)\cdot 1=\lim_{k\to\infty}\alpha(D_k)\cdot\lim_{k\to\infty}\frac{\alpha\big(T_k^{(3)}\big)}{\alpha(D_k)}=\lim_{k\to\infty}\left(\frac{\alpha\big(T_k^{(3)}\big)}{\alpha(D_k)}\cdot\alpha(D_k)\right)=\lim_{k\to\infty}\alpha\big(T_k^{(3)}\big).$$

Значит предел $\lim_{k\to\infty}\alpha\big(T_k^{(3)}\big)$ существует как предел произведения (см. [2]) и равен индексу роста языка D. □

Замечание 1.14. *Как показывают численные эксперименты, настоящее значение N такое, что $P_k(n)\approx N^n P'_k(n)$, намного меньше 28, а именно, примерно 2.119.*

Теорема 1.6 не только доказывает гипотезу 1.1 для случая $m=3$, но и позволяет усилить ее:

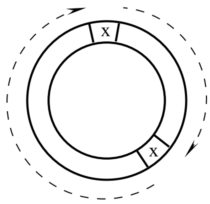
Замечание 1.15 (Усиление гипотезы 1.1)**.** *Для любого $m\geq 3$ предел $\alpha^{(m)}=\lim_{k\to\infty}\alpha\big(T_k^{(m)}\big)$ не только существует, но и совпадает с индексом роста локального двумерного языка над алфавитом Δ, множество разрешенных подслов которого избегает фрагментов из множества Q_m, описанного в предложении 1.1.*

Глава 2

Циклические слова

Циклические слова образуют очень интересный и важный класс комбинаторных объектов. Циклические слова описывают циклы в графах и автоматах, а также подходят для описания любого циклического процесса, от производственных циклов до циклов солнечной активности. Кроме того, циклические слова взаимно однозначно соответствуют классам сопряженности обычных слов и неявно присутствуют во всех связанных с такими классами задачах. Поэтому получение оценок для циклической границы повторяемости также является немаловажной задачей. Напомним, что для 2- и 3-буквенного алфавита значение циклической границы повторяемости уже известно. Поэтому далее мы полагаем $k \geq 4$.

У циклических слов есть одна особенность. В «обычном» слове расстояние между двумя вхождениями одного символа определяется как разность между номерами их позиций. А в циклическом слове расстояние можно измерять в двух направлениях:



2.1. Нижняя оценка для циклической границы повторяемости

Предложение 2.1. *Для любого $k \geq 4$ не существует $\frac{\lceil k/2\rceil+1}{\lceil k/2\rceil}$-свободного циклического слова над алфавитом Σ_k длины $k+1$.*

Доказательство. Рассмотрим циклическое слово (w) над алфавитом Σ_k длины $k+1$. По принципу Дирихле в слове (w) обязательно найдется хотя бы одна буква, б.о.о. x, готорая появляется в (w) по крайней мере дважды. Следовательно, $(w) = (xw_1xw_2)$, причем $\min\{|w_1|, |w_2|\} \leq (k-1)/2 < \lceil k/2 \rceil$. Но тогда $\max\{\exp(xw_1x), \exp(xw_2x)\} \geq \frac{\lceil k/2 \rceil+1}{\lceil k/2 \rceil}$. А значит слово (w) не может быть $\frac{\lceil k/2 \rceil+1}{\lceil k/2 \rceil}$-свободным. $\square$

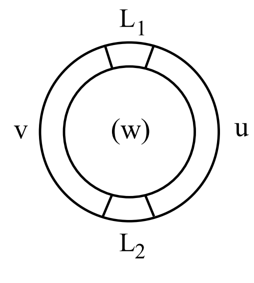
Следствие 2.1. *Из предложения 2.1 непосредственно следует, что* $\mathrm{CRT}(k) \geq \frac{\lceil k/2 \rceil+1}{\lceil k/2 \rceil}$ *для любого* $k \geq 4$.

Замечание 2.1. *Для алфавитов из 2 и 3 букв данная оценка тоже справедлива, но может быть улучшена до точной при рассмотрении более длинных циклических слов (длины 5).*

2.2. Верхние оценки для циклической границы повторяемости

Чтобы получить верхнюю оценку для циклической границы повторяемости $\mathrm{CRT}(k) \leq \beta$, необходимо для любого $n \in \mathbb{N}$ построить β^+-свободное циклическое слово (w) над алфавитом Σ_k длины n. Основная идея построения подходящего слова (w) длины n заключается в следующем:

- выбираем $\bar{k} \in \mathbb{N}$ такое, что $\bar{k} = \min\{i \in \mathbb{N} \mid \beta \geq \mathrm{RT}(i)\}$;
- берем произвольное $\mathrm{RT}(\bar{k})^+$-свободное (а значит и β^+-свободное в силу выбора $\bar{k}$) слово u над $\bar{k}$-буквенным алфавитом A длины $\lceil (n-2c)/2 \rceil$, где c – некоторая константа (такое слово всегда существует согласно граничной теореме);
- берем произвольное $\mathrm{RT}(\bar{k})^+$-свободное слово v над $\bar{k}$-буквенным алфавитом B длины $\lfloor (n-2c)/2 \rfloor$, причем $A \cup B = \Sigma_k$;
- склеиваем слова u и v с помощью двух вставок L_1 и L_2 длины c, состоящих из чередующихся букв алфавитов $A \setminus B$ и $B \setminus A$ (слова читаем по часовой стрелке);



- с помощью определенных перестановок на алфавитах A и B добиваемся, чтобы полученное слово $(w) = (L_1 u L_2 v)$ было искомым (т.е. β^+-свободным).

Далее рассматриваются несколько случаев в зависимости от мощности исходного алфавита, в каждом из которых с помощью изложенной выше идеи (возможно с некоторыми модификациями) получаются верхние оценки для циклической границы повторяемости.

2.2.1. Случай $\mathbf{k} = 4$

Теорема 2.1. *Над* 4*-буквенным алфавитом существует циклическое* $(7/4)^+$*-свободное слово любой длины.*

Доказательство. Для любого $n \in \mathbb{N}$ построим циклическое $(7/4)^+$-свободное слово (w) длины n над алфавитом $\Sigma_4 = \{1, 2, 3, 4\}$. Выделим три случая относительно значения n:

- если $n \leq 4$, то требуемое слово (w) получается тривиально (берем слово, в котором все буквы различны);
- если $5 \leq n \leq 9$, то искомое слово (w) изображено на рис. 2.1;
- если $n \geq 10$, то слово (w) будем искать в виде $w = L_1 u L_2 v$.

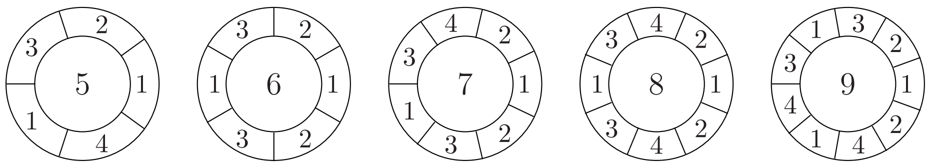


Рис. 2.1. $(7/4)^+$*-свободное слово* (w) *длины* n*, где* $5 \leq n \leq 9$ *(длина слова указана внутри круга).*

Остановимся подробнее на последнем случае. Здесь u – произвольное $(7/4)^+$-свободное слово над алфавитом $A = \{1, 2, 3\}$, длина которого равна $\lceil (n-4)/2 \rceil$, дополнительно потребуем, чтобы $u[1] = u[3]$[1]; v – произвольное $(7/4)^+$-свободное слово над алфавитом $B = \{2, 3, 4\}$, длина которого равна $\lfloor (n-4)/2 \rfloor$; $L_1 = 14$; $L_2 = 41$.

Обозначим через u_1 (соответственно, v_1) – префикс длины 2 слова u (соответственно, v), а через u_2 (соответственно, v_2) – суффикс длины 1 слова u (соответственно, v). Заметим, что в u_1 (соответственно, v_1) обе буквы различны. Кроме того, существует всего три способа распределения одинаковых букв в u_1 и u_2 (соответственно, v_1 и v_2), все они изображены на рис. 2.2. Каждый способ определяет *тип* слова u, который будем обозначать тем же номером.

Пусть биекция $f : A \to B$ такая, что $f(1) = 4,\ f(2) = 2,\ f(3) = 3$. Тогда, если n четно, то $|v| = |u|$ и в качестве слова v можно взять $f(u)$ (это гарантирует, что слова u и v будут одного типа); а если n нечетно, то $|v| = |u| - 1$ и в качестве слова v можно взять слово $f(u)$ без первой буквы (в этом случае тип слова v также будет однозначно определен за счет условия равенства первой и третьей буквы в слове u).

[1] $(7/4)^+$-свободное Z-слово над алфавитом A содержит подслово вида $a_1 a_2 a_1$, которое можно продолжить вправо до любой длины, в частности до длины $\lceil (n-4)/2 \rceil$, полученное слово и возьмем в качестве u.

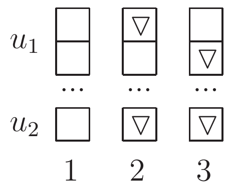


Рис. 2.2. *Варианты расположения общих символов в словах u_1 и u_2 ($k = 4$). Общий символ обозначен $\triangledown$.*

Для каждого типа слова u и для разной четности n рассмотрим варианты перестановок в словах u и v, позволяющие сделать слово (L_1uL_2v) $(7/4)^+$-свободным (см. рис. 2.3). Например, когда слово u имеет тип 1, то независимо от четности n слово v также будет иметь тип 1, поэтому перестановки для любого n будут одинаковы: необходимо переименовать буквы $u[1]$ и 2 (т.е. все вхождения буквы $u[1]$ в слове u заменить на 2 и наоборот), $u[2]$ и 1, $u[|u|]$ и 3 в слове u, а также $v[1]$ и 2, $v[2]$ и 4, $v[|v|]$ и 3 в слове v.

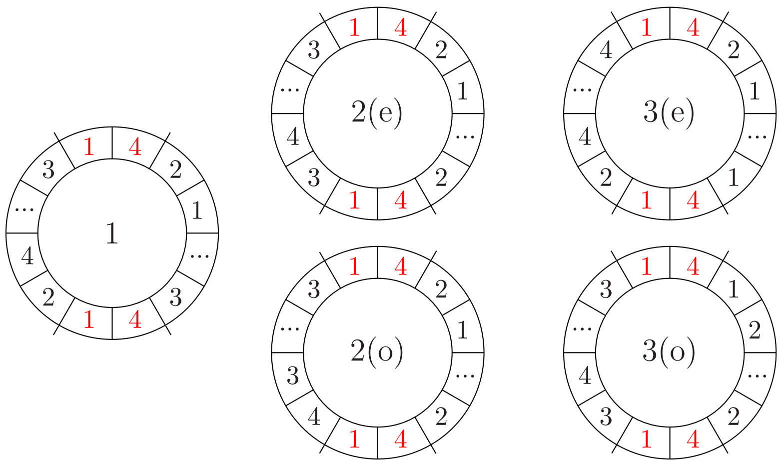


Рис. 2.3. $(7/4)^+$*-свободное слово (w) длины n, где $n \geq 10$ (внутри круга указан тип u и четность n).*

Прежде чем перейти к проверке того, что в каждом случае получилось $(7/4)^+$-свободное слово, сделаем несколько общих наблюдений.

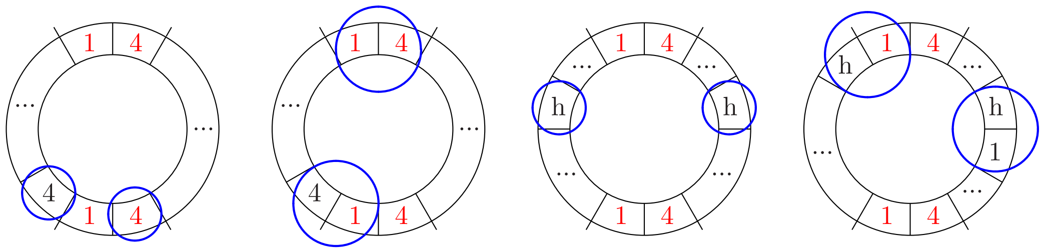
Наблюдение 2.1. *Каждое из слов u_1 и v_1 обязательно содержит букву из $A \setminus B$ и $B \setminus A$, соответственно.*

Наблюдение 2.2. *Если ptp – подслово u или v, то $\exp(ptp) \leq 7/4$ в силу выбора u и v. Причем, если $(w) = (ptpt')$, то $|pt'p| > |ptp|$, а значит $\exp(pt'p) < \exp(ptp) \leq 7/4$.*

Наблюдение 2.3. *Пусть ptp – подслово (w) такое, что одно вхождение p целиком содержится в u, а другое в v. Так как $|A \cap B| = 2$, то максимальная длина слова p в этом случае равна 3 (если $p = 232$ или 323). Если p захватывает одну из вставок L_1 или L_2, то в силу их строения и наблюдения 2.1 p также будет коротким.*

Наблюдение 2.4. *В силу перестановок, сделанных в словах u и v,* $\Sigma(u_1) \cap \Sigma(v_2) = \varnothing$ *и* $\Sigma(u_2) \cap \Sigma(v_1) = \varnothing$.

Проверим, что экспонента любого подслова слова (w) не превосходит $7/4$. Согласно наблюдению 2.2 можно не рассматривать подслова вида ptp, когда оба вхождения p целиком находятся в u (соответственно, v). Остальные возможные варианты расположения «опасных» подслов и их экспоненты представлены ниже (здесь $h \in \{2,3\}^*$ – общее подслово u и v, и $|h| \le 3$):



$$1\underbrace{\cdots}_{\ge 1}1 \qquad \exp(ptp) \le 3/2 < 7/4;$$

$$4\underbrace{\cdots}_{\ge 1}4 \qquad \exp(ptp) \le 3/2 < 7/4;$$

$$14\underbrace{\cdots}_{\ge 2}14 \qquad \exp(ptp) \le 6/4 < 7/4,$$

здесь мы воспользовались тем, что $\min\{|u|,|v|\} = \lfloor (n-4)/2 \rfloor \ge 3$;

$$41\underbrace{\cdots}_{\ge 2}41 \qquad \exp(ptp) \le 6/4 < 7/4;$$

$$\underbrace{h\cdots L_1[L_2]\cdots h}_{\ge 2} \qquad \exp(ptp) \le 8/5 < 7/4;$$

$$h1\ \underbrace{4\cdots 1\cdots}_{\ge 2}\ h1 \qquad \exp(ptp) \le 10/6 < 7/4,$$

здесь мы использовали наблюдения 2.1 и 2.4;

$$1h\ \underbrace{\cdots u_2 4}_{\ge 2}\ 1h \qquad \exp(ptp) \le 6/4 < 7/4,$$

в этом случае $|h| = 1$ благодаря наблюдению 2.1;

$$h4\ \underbrace{1\cdots 4\cdots}_{\ge 2}\ h4 \qquad \exp(ptp) \le 10/6 < 7/4;$$

$$4h\ \underbrace{\cdots v_2 1}_{\ge 2}\ 4h \qquad \exp(ptp) \le 6/4 < 7/4.$$

Отметим, что подслова вида $141 \cdots 141$, $414 \cdots 414$, $h14 \cdots h14$, $14h \cdots 14h$, $h41 \cdots h41$, $41h \cdots 41h$ в построенных словах не появляются, значит все «подозрительные» подслова ptp проверены. Следовательно, для любого $n \in \mathbb{N}$ существует циклическое $(7/4)^+$-свободное слово длины n над 4-буквенным алфавитом. $\square$

Следствие 2.2. $\mathrm{CRT}(4) \le 7/4$.

2.2.2. Случай $\mathbf{k = 5}$

Теорема 2.2. *Над 5-буквенным алфавитом существует циклическое* $(3/2)^+$*-свободное слово любой длины.*

Доказательство. Аналогично случаю $k = 4$ для любого $n \in \mathbb{N}$ построим циклическое $(3/2)^+$-свободное слово (w) длины n над алфавитом $\Sigma_5 = \{1, 2, 3, 4, 5\}$. Выделим три случая относительно значения n:

- если $n \leq 5$, то требуемое слово (w) получается тривиально;
- если $6 \leq n \leq 15$, то искомое слово (w) изображено на рис. 2.4;
- если $n \geq 16$, то слово (w) будем искать в виде $w = L_1 u L_2 v$.

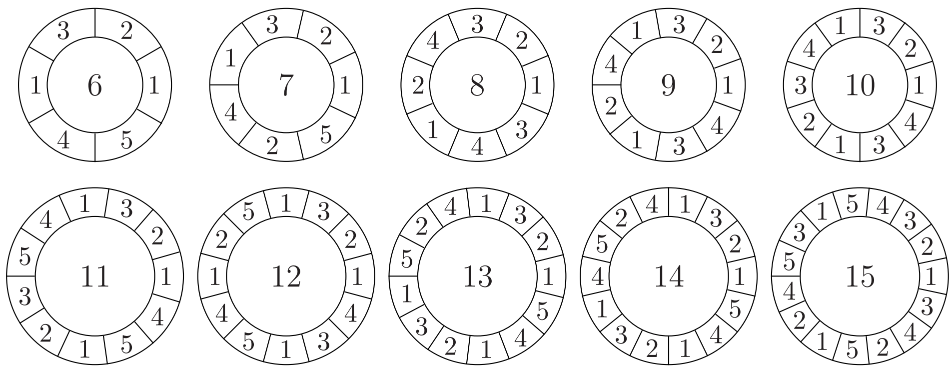


Рис. 2.4. $(3/2)^+$*-свободное слово* (w) *длины* n*, где* $6 \leq n \leq 15$ *(длина слова указана внутри круга).*

Разберем последний случай. Здесь u – произвольное $(7/5)^+$-свободное слово над алфавитом $A = \{1, 2, 3, 4\}$, длина которого равна $\lceil (n-4)/2 \rceil$ и $u[1] = u[4]$; v – $(7/5)^+$-свободное слово над алфавитом $B = \{2, 3, 4, 5\}$, которое равно $f(u)$ (в случае четного n) или $f(u)$ без первой буквы (когда n нечетно) для биекции $f : A \to B$ такой, что $f(1) = 5,\ f(2) = 2,\ f(3) = 3,\ f(4) = 4$; $L_1 = 15$; $L_2 = 51$.

Обозначим через u_1 (соответственно, v_1) и u_2 (соответственно, v_2) – префикс и суффикс длины 3 слова u (соответственно, v). Заметим, что в словах u_1, u_2, v_1, v_2 все буквы различны. В зависимости от того, на каких позициях встречаются одинаковые буквы в словах u_1 и u_2 (соответственно, v_1 и v_2), можно выделить 24 случая (см. рис. 2.5). Причем число случаев нельзя сократить за счет симметрии (например, вместо u, имеющего тип 17, использовать $\overleftarrow{u}$, который имеет тип 1), так как при этом может нарушится условие $\overleftarrow{u}[1] = \overleftarrow{u}[4]$, а значит пропадет однозначность в определении типа слова v в случае нечетного n.

Для каждого типа слова u переставляем буквы в словах u и v так, чтобы их префиксы и суффиксы выглядели как на рис. 2.6.

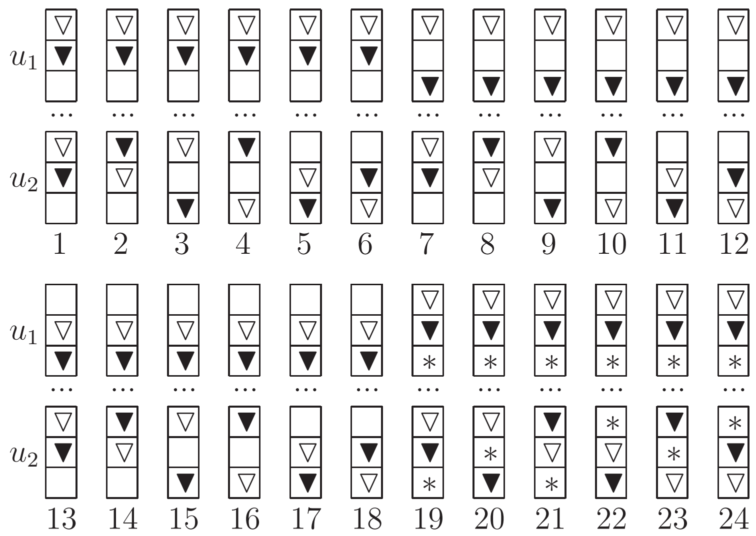


Рис. 2.5. *Варианты расположения общих символов в словах u_1 и u_2 ($k = 5$). Для обозначения общих символов используются $\triangledown$, $\blacktriangledown$ и $*$.*

Наблюдение 2.5. *Слова u_1 и u_2 обязательно содержат букву из $A \setminus B$, а слова v_1 и v_2 – букву из $B \setminus A$.*

Наблюдения 2.2 и 2.3 будут справедливы и в случае $k = 5$, только экспонента будет $3/2$ вместо $7/4$, а длина наибольшего общего подслова u и v будет равна 4 (это слово 2342 и все полученные из него с помощью перестановки букв). Следующие подслова вида ptp будут иметь разрешенную экспоненту в каждом случае (здесь $h \in \{2, 3, 4\}^*$ – общее подслово u и v такое, что $|h| \le 2$; $q \in \{2, 3, 4\}^*$ – общее подслово u и v такое, что $3 \le |q| \le 4$):

$$1\underbrace{\cdots}_{\ge 1}1 \qquad \exp(ptp) \le 3/2;$$

$$5\underbrace{\cdots}_{\ge 1}5 \qquad \exp(ptp) \le 3/2;$$

$$15\underbrace{\cdots}_{\ge 5}15 \qquad \exp(ptp) \le 9/7 < 3/2,$$

здесь мы воспользовались тем, что $|u|, |v| \ge 6$;

$$51\underbrace{\cdots}_{\ge 5}51 \qquad \exp(ptp) \le 9/7 < 3/2;$$

$$h\underbrace{\cdots L_1[L_2]\cdots}_{\ge 2}h \qquad \exp(ptp) \le 3/2;$$

$$q\underbrace{\cdots 5[1]\cdots L_1[L_2]\cdots 1[5]\cdots}_{\ge 5}q \qquad \exp(ptp) \le 13/9 < 3/2,$$

здесь мы используем наблюдение 2.5;

$$h1\ \underbrace{5u_1\cdots}_{\ge 4}\ h1 \qquad \exp(ptp) \le 10/7 < 3/2;$$

$$1h\ \underbrace{\cdots u_2 5}_{\ge 4}\ 1h \qquad \exp(ptp) \le 10/7 < 3/2,$$

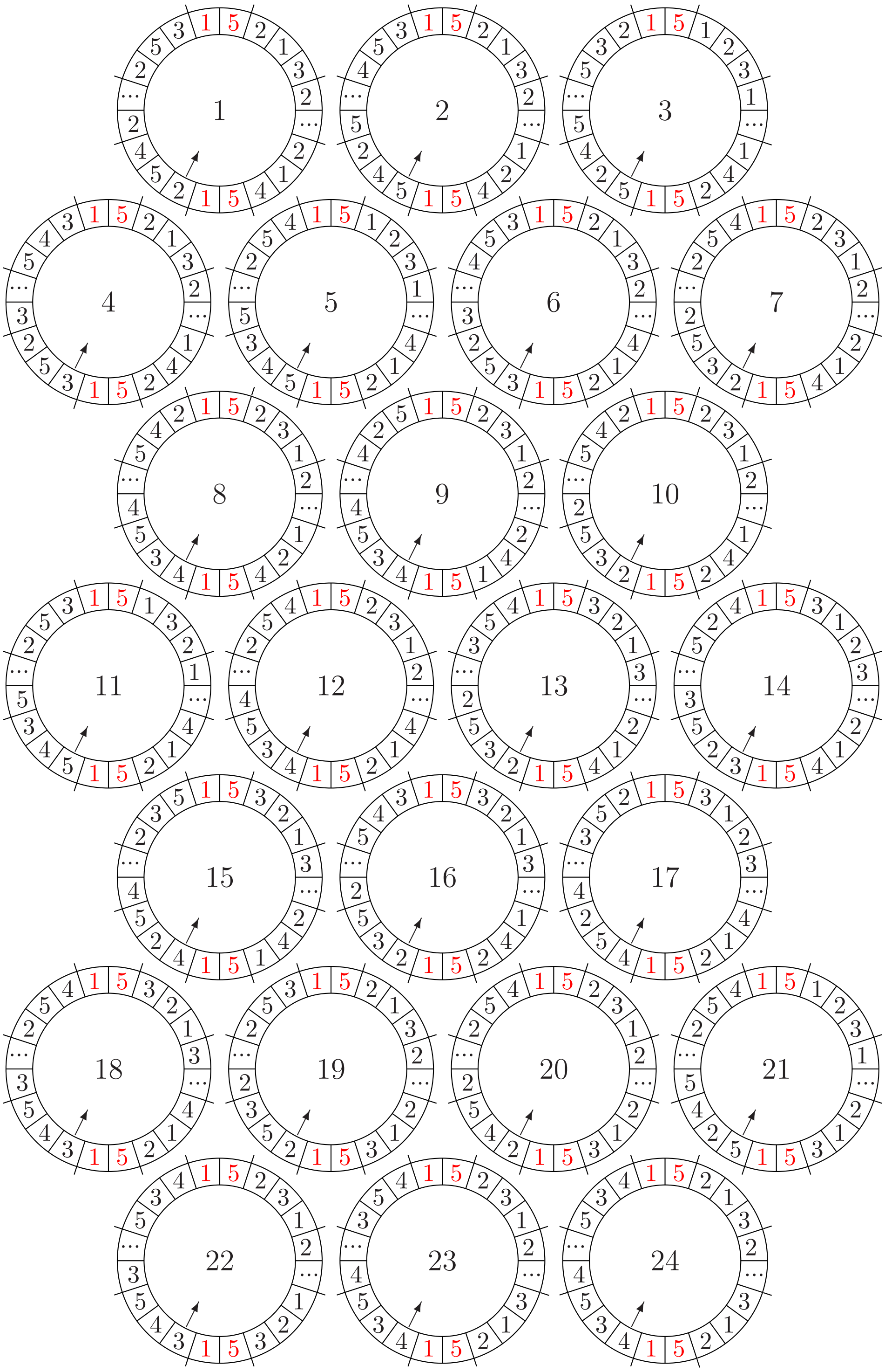


Рис. 2.6. $(3/2)^{+}$-свободное слово (w) длины n, где $n \geq 16$ (внутри круга указан тип u). На диаграммах показан случай четного n, для нечетного n надо лишь удалить первый символ слова v (отмечен стрелкой).

исключение составляют случаи 18 и 23, в которых построенное слово (w) четной длины может содержать подслова 134 125 134 и 143 125 143, соответственно, которые выбиваются из общей схемы, но имеют разрешенную экспоненту, равную $3/2$;

$$h5\ \underbrace{1v_1\cdots}_{\geq 4}\ h5 \qquad \exp(ptp) \leq 10/7 < 3/2;$$

$$5h\ \underbrace{\cdots v_2 1}_{\geq 4}\ 5h \qquad \exp(ptp) \leq 10/7 < 3/2,$$

здесь исключение тоже составляют случаи 18 и 23, в которых построенное слово (w) может содержать подслова 532 541 532 и 523 541 523, соответственно, которые выбиваются из общей схемы, но тем не менее имеют разрешенную экспоненту, равную $3/2$.

Таким образом, для любого $n \in \mathbb{N}$ построено циклическое $(3/2)^+$-свободное слово длины n над 5-буквенным алфавитом. $\square$

Следствие 2.3. $\mathrm{CRT}(5) \leq 3/2$.

2.2.3. Случай k = 6

Теорема 2.3. *Над 6-буквенным алфавитом существует циклическое $(4/3)^+$-свободное слово любой длины.*

Доказательство. Для любого $n \in \mathbb{N}$ построим циклическое $(4/3)^+$-свободное слово (w) длины n над алфавитом $\Sigma_6 = \{1, 2, 3, 4, 5, 6\}$. Выделим три случая относительно значения n:

- если $n \leq 6$, то требуемое слово (w) получается тривиально;
- если $7 \leq n \leq 19$, то искомое слово (w) изображено на рис. 2.7;
- если $n \geq 20$, то слово (w) будем искать в виде $w = L_1 u L_2 v$.

Разберем последний случай. Здесь u – произвольное $(5/4)^+$-свободное слово над алфавитом $A = \{1, 2, 3, 4, 5\}$, длина которого равна $\lceil (n-4)/2 \rceil$ и $u[1] = u[5]$; v – $(5/4)^+$-свободное слово над алфавитом $B = \{2, 3, 4, 5, 6\}$, которое равно $f(u)$ (в случае четного n) или $f(u)$ без первой буквы (когда n нечетно), где $f : A \to B$ – биекция такая, что $f(1) = 6,\ f(2) = 2,\ f(3) = 3,\ f(4) = 4,\ f(5) = 5$; $L_1 = 16$; $L_2 = 61$.

Обозначим через u_1 (соответственно, v_1) и u_2 (соответственно, v_2) – префикс и суффикс длины 4 слова u (соответственно, v). Заметим, что в слове u_1 (соответственно, u_2, v_1, v_2) все буквы различны. В зависимости от того, на каких позициях встречаются одинаковые буквы в словах u_1 и u_2 (v_1 и v_2), можно выделить 120 случаев (см. рис. 2.8).

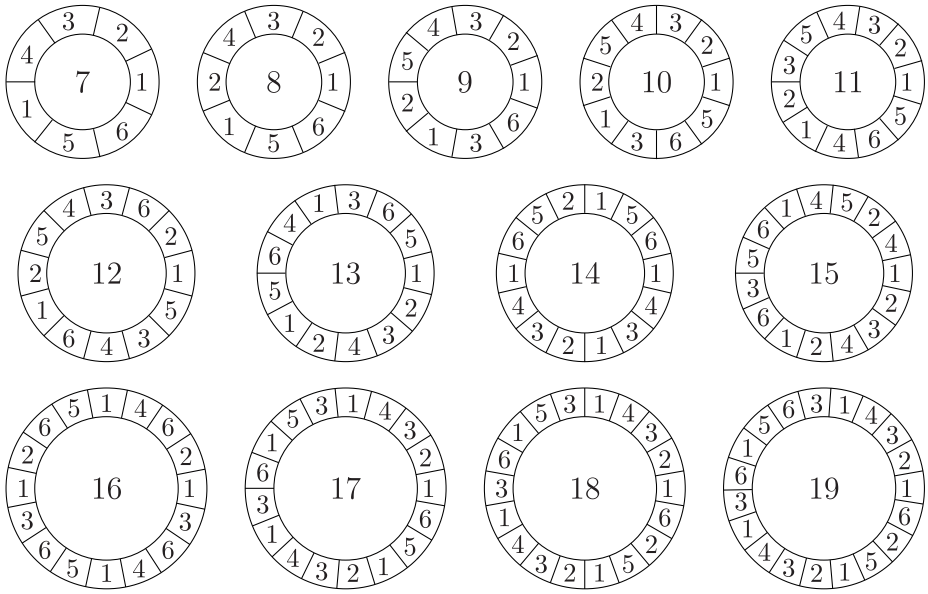


Рис. 2.7. $(4/3)^+$*-свободное слово* (w) *длины* n, где $7 \leq n \leq 19$ *(длина слова указана внутри круга).*

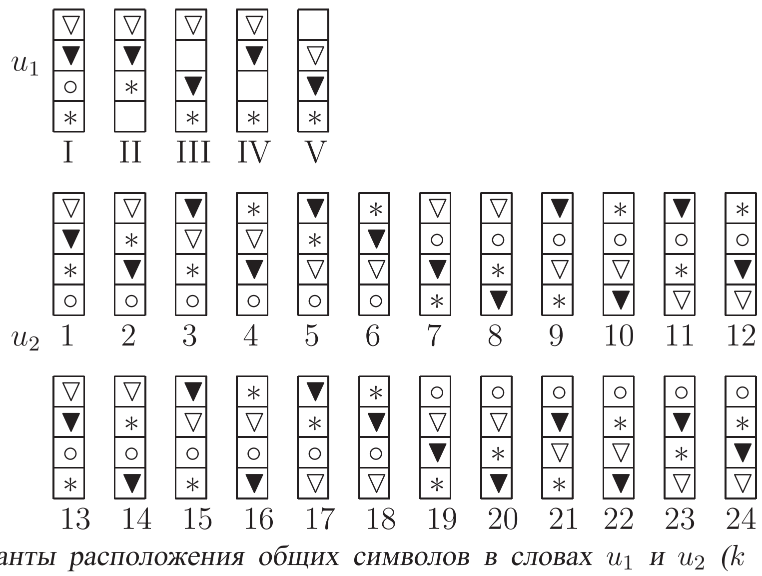


Рис. 2.8. *Варианты расположения общих символов в словах* u_1 *и* u_2 $(k = 6)$*. Римскими цифрами обозначен тип слова* u*, арабскими – тип* v*.*

Для каждого типа u_1 и u_2 переставляем буквы в словах u и v так, чтобы их префиксы и суффиксы выглядели как на рис. 2.9–2.13. Отметим, что в силу выбора слова u мы знаем в нем не только пятую, но и шестую букву, которая однозначно определяется с учетом второго правила Пансьё: $\{u[6]\} = A \setminus \Sigma(u_1)$. Аналогично получаем для слова $v = f(u)$: $\{v[6]\} = B \setminus \Sigma(v_1)$. Поэтому на рисунках слова u_1 и v_1 изображены с «запасом» в две буквы.

Сделаем несколько важных замечаний, которые будут полезны при проверке экспонент подслов вида ptp. Наблюдения 2.5, 2.2 и 2.3 будут справедливы и в случае $k = 6$, только экспонента будет $4/3$, а длина наибольшего общего подслова u и v будет равна 5 (это слово 23452 и все полученные из него с помощью перестановки букв). Согласно наблюдению 2.5 слово u_1 имеет вид $d_1 1 q_1$, причем для $k = 6$ $1 \le |d_1| \le 3$, а $0 \le |q_1| \le 2$. Аналогично получаем: $u_2 = q_2 1 d_2$, $v_1 = \bar{d}_1 6 \bar{q}_1$, $v_2 = \bar{q}_2 6 \bar{d}_2$. Тогда

$$1 \underbrace{\cdots}_{\ge 2} 1 \qquad \exp(ptp) \le 4/3;$$

$$6 \underbrace{\cdots}_{\ge 2} 6 \qquad \exp(ptp) \le 4/3.$$

Пусть $h \in \{2,3,4,5\}^*$. Если $\mathbf{h}$ **– общее подслово $\mathbf{\bar{d}_2}$ и $\mathbf{d_1}$** (аналогично для пары $\bar{d}_1$ и d_2), то $|h| \le 2$ за счет сделанных перестановок, причем

$$|h| = 1 \qquad h \underbrace{\cdots L_1[L_2] \cdots}_{\ge 2} h \qquad \exp(ptp) \le 4/3;$$

$$|h| = 2 \qquad h \underbrace{v[|v|] L_1 u[1]}_{=4} h \ \text{ или } \ h \underbrace{u[|u|] L_2 v[1]}_{=4} h \qquad \exp(ptp) = 4/3,$$

как в случае IV-4 (u_1 имеет тип IV, u_2 – тип 4) с повтором 34 5162 34. Причем повторов вида $h1 \cdots h1$, $1h \cdots 1h$, $h6 \cdots h6$, $6h \cdots 6h$ в этом случае не возникает.

Если $\mathbf{h}$ **не является подсловом $\mathbf{d_1}$, $\mathbf{\bar{d}_2}$** (соответственно, $\bar{d}_1$, d_2), то $|h| \le 5$. Тогда

$$|h| \le 3 \qquad h \underbrace{6 \bar{d}_2 L_1 d_1 1}_{\ge 6} h \qquad \exp(ptp) \le 4/3;$$

$$h \underbrace{1 d_2 L_2 \bar{d}_1 6}_{\ge 6} h \qquad \exp(ptp) \le 4/3.$$

Для $4 \le |h| \le 5$ проверка во всех 120 случаях примерно одинакова (хоть и не выражается общей формулой, как для $|h| \le 3$), поэтому рассмотрим наиболее характерный случай. Проверим, что в случае I-13 (верхняя часть) не возникает подслова h 6416231 h. После 1 нам известно три следующие буквы – 425. Значит $h = 4253$ или $h = 42534$. Но в слове v_2 перед 6 стоит 23, а в $(5/4)^+$-свободном слове v между двумя соседними вхождениями одной буквы, в частности буквы 3, должно находится не менее трех символов, следовательно

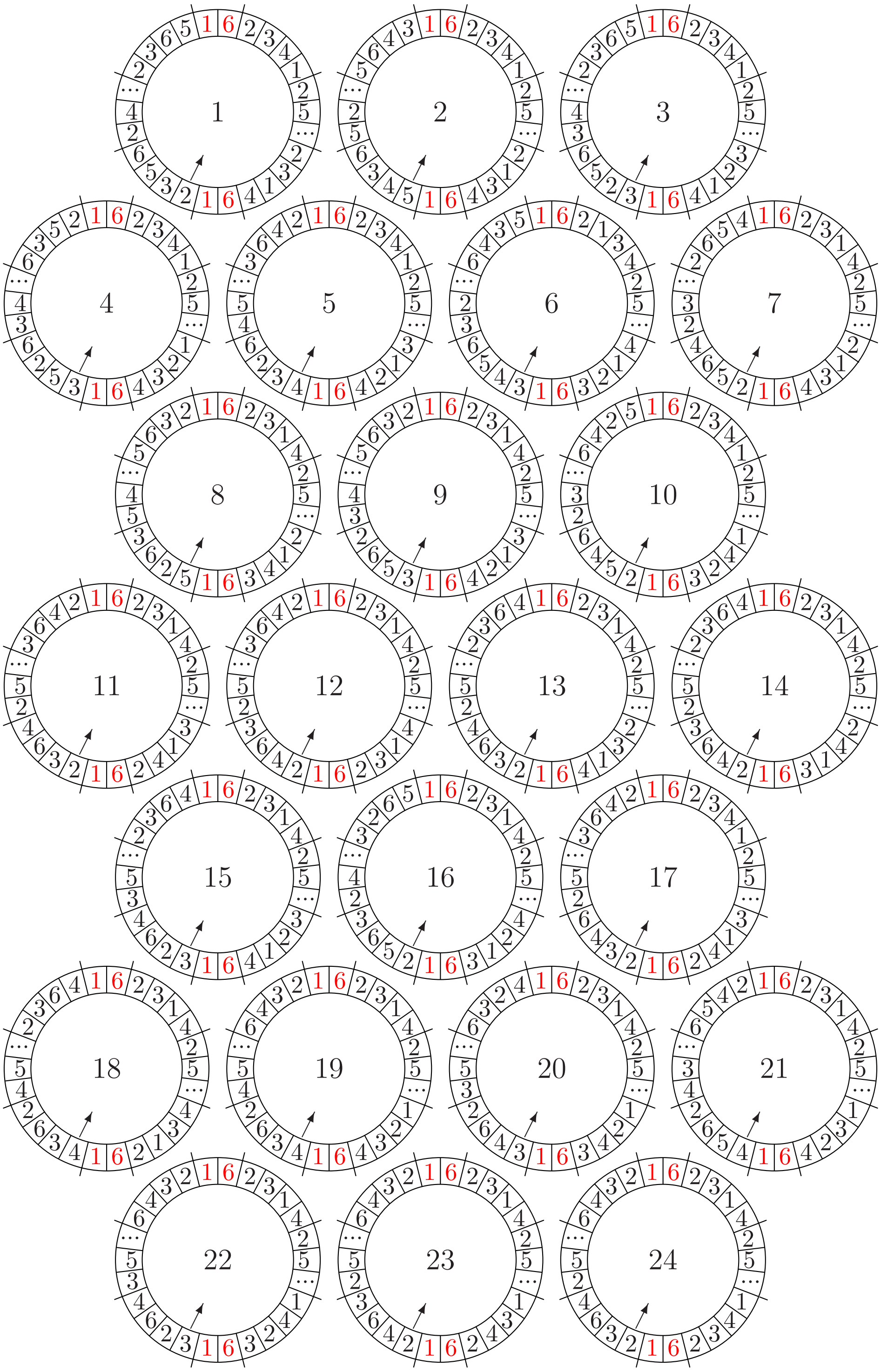


Рис. 2.9. $(4/3)^+$-свободное слово (w) длины n, где $n \geq 20$ (u_1 *имеет тип I, а внутри круга указан тип* u_2). *На диаграммах показан случай четного* n, *для нечетного* n *надо лишь удалить первый символ слова* v *(отмечен стрелкой).*

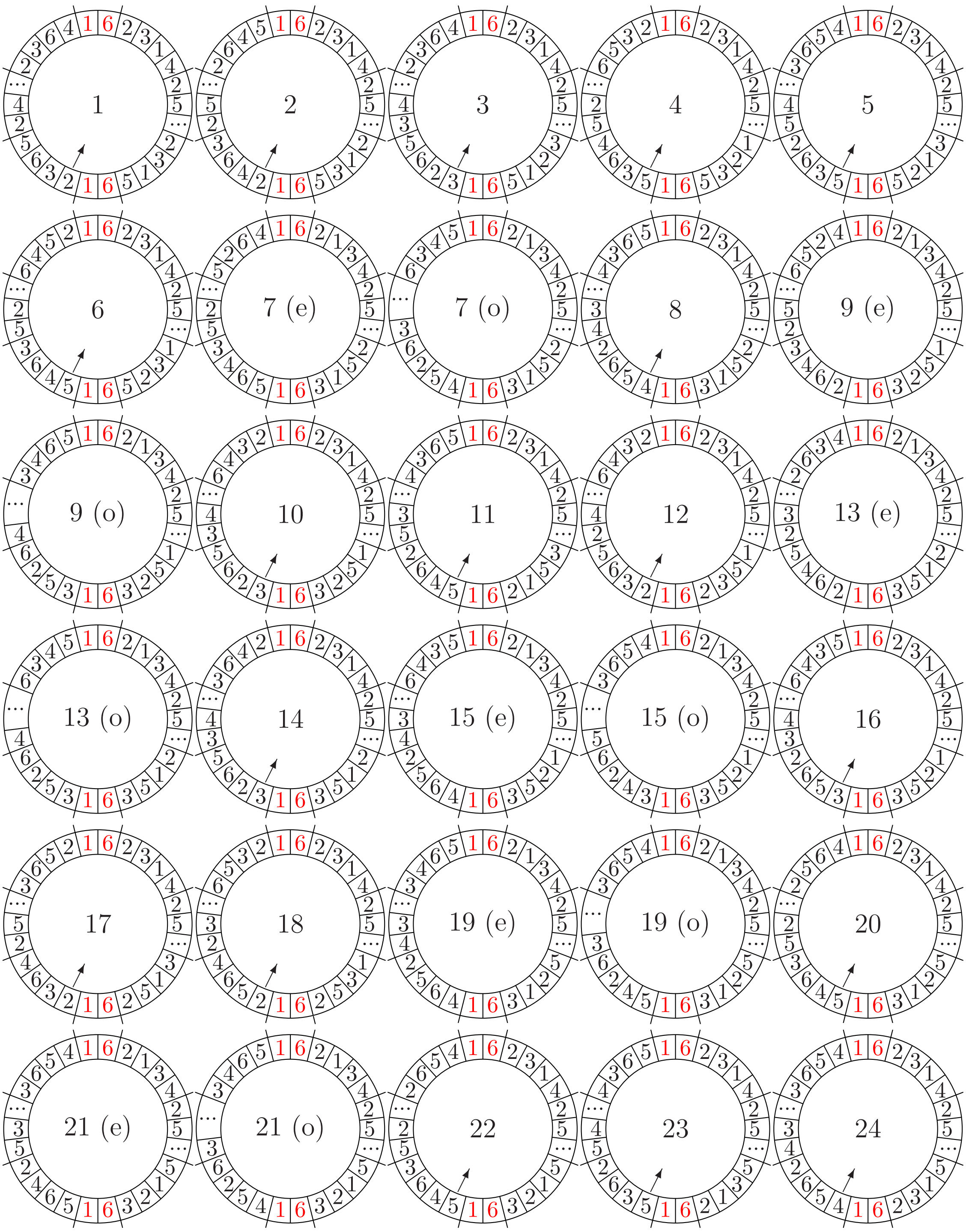


Рис. 2.10. $(4/3)^+$*-свободное слово* (w) *длины* n, *где* $n \geq 20$ $(u_1$ *имеет тип II, а внутри круга указан тип* $u_2)$. *Для некоторых типов* u_2 *случаи четного и нечетного* n *рассмотрены отдельно.*

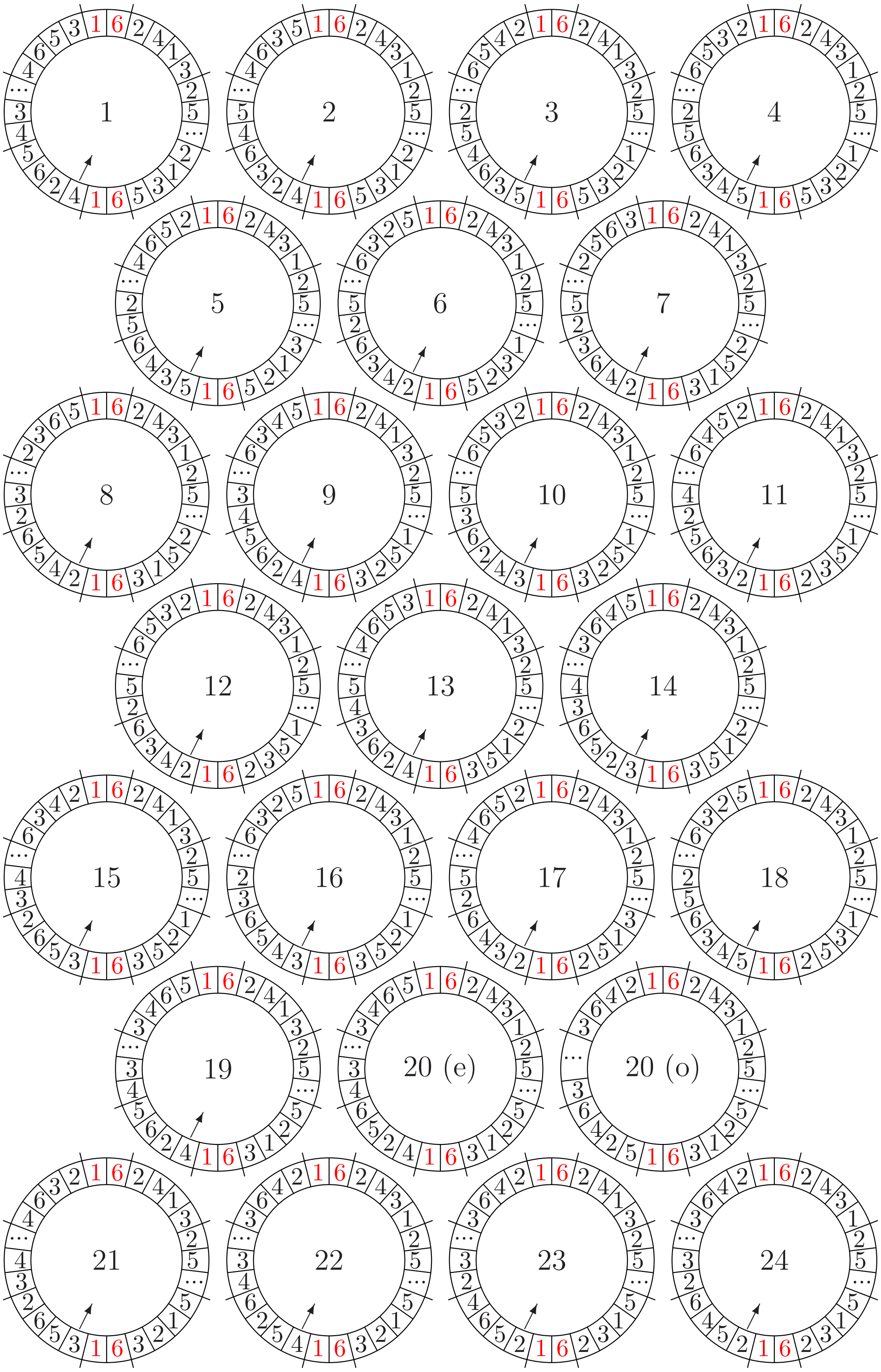


Рис. 2.11. $(4/3)^+$*-свободное слово* (w) *длины* n, где $n \geq 20$ (u_1 *имеет тип III, а внутри круга указан тип* u_2).

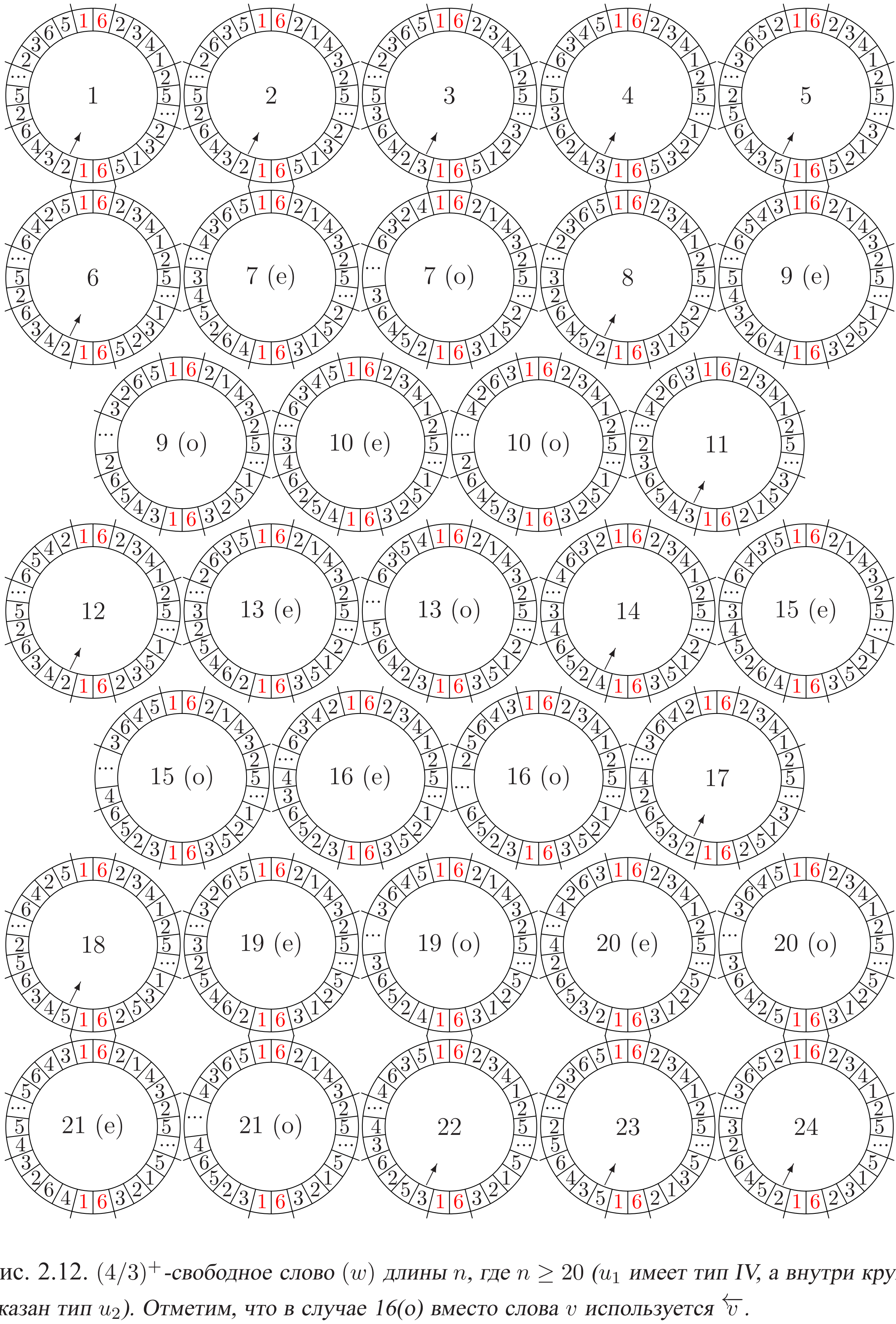


Рис. 2.12. $(4/3)^+$-свободное слово (w) длины n, где $n \geq 20$ (u_1 имеет тип IV, а внутри круга указан тип u_2). Отметим, что в случае 16(о) вместо слова v используется $\overleftarrow{v}$.

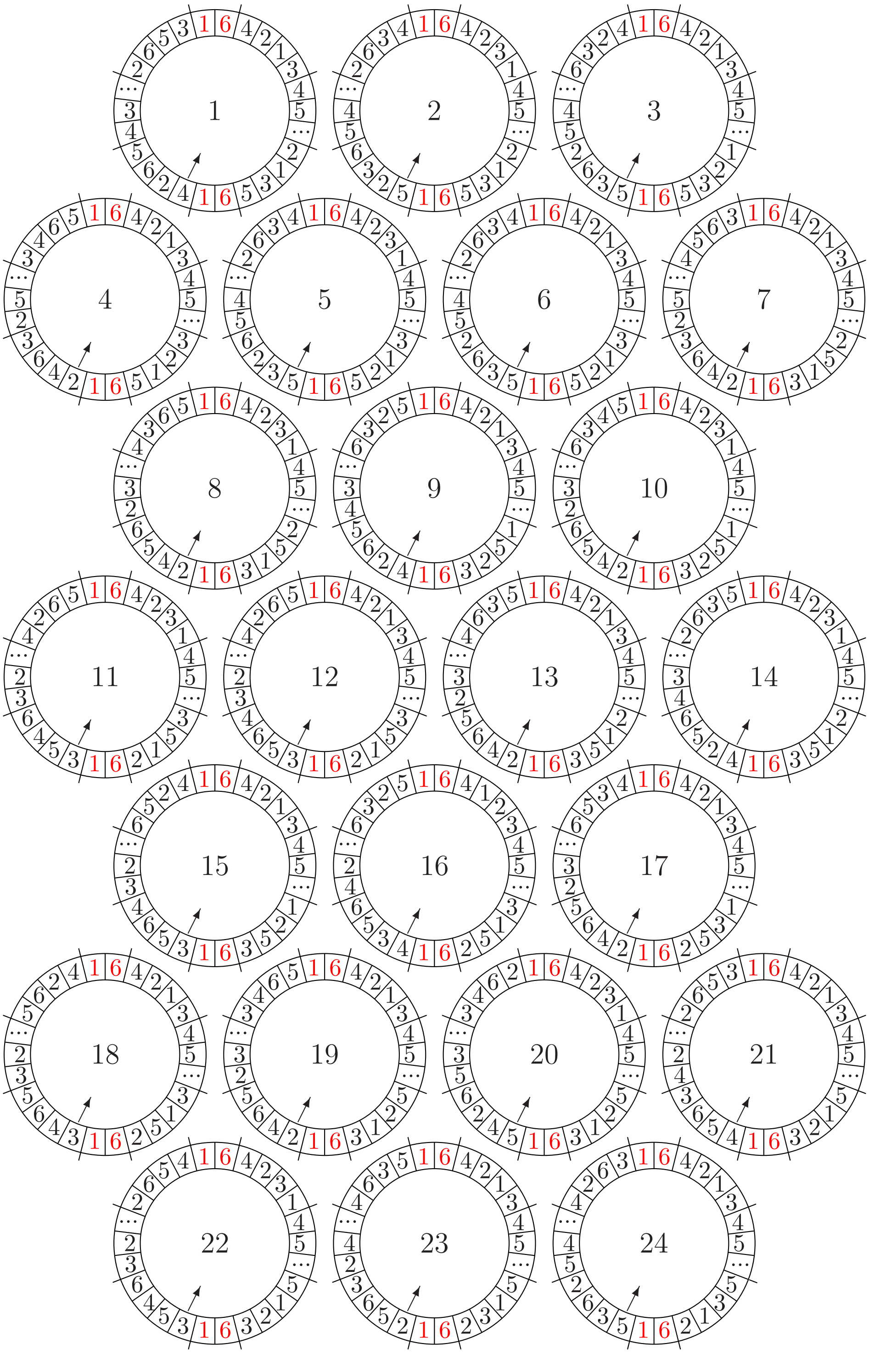


Рис. 2.13. $(4/3)^+$*-свободное слово* (w) *длины* n*, где* $n \geq 20$ *(*u_1 *имеет тип V, а внутри круга указан тип* u_2*).*

$$4253 \underbrace{\cdots}_{\geq 2} 236416231\ 4253 \quad \exp(4253\cdots 4253) \leq 19/15 < 4/3;$$
$$42534 \underbrace{\cdots}_{\geq 1} 236416231\ 42534 \quad \exp(42534\cdots 42534) \leq 20/15 = 4/3.$$

А в случае I-13 (нижняя часть и нечетное n) не возникает запрещенного подслова $h\ 146136\ h$ (варианты для слова h те же):

$$4253 \underbrace{\cdots}_{\geq 2} 23146136\ 4253 \quad \exp(4253\cdots 4253) \leq 18/14 < 4/3;$$
$$42534\ 1 \underbrace{\cdots}_{\geq 1} 23146136\ 42534 \quad \exp(42534\cdots 42534) \leq 20/15 = 4/3,$$

здесь мы используем то, что длина максимального общего подслова u и v равна 5, следовательно, подслово 42534 слева и справа может быть ограничено только 1 в u или 6 в v.

Осталось рассмотреть случай, когда **только одно вхождение h попадает в $\mathbf{d_1}$ или $\bar{\mathbf{d}}_\mathbf{2}$** (соответственно, $\bar{d}_1$ или d_2). В этом случае $|h| \leq 3$.

$$|h| \leq 2 \qquad h\underbrace{\cdots 6\bar{d}_2 L_1 \cdots}_{\geq 4} h \qquad \exp(ptp) \leq 8/6 = 4/3;$$
$$h\underbrace{\cdots L_1 d_1 1 \cdots}_{\geq 4} h \qquad \exp(ptp) \leq 8/6 = 4/3;$$
$$h\underbrace{\cdots 1 d_2 L_2 \cdots}_{\geq 4} h \qquad \exp(ptp) \leq 8/6 = 4/3;$$
$$h\underbrace{\cdots L_2 \bar{d}_1 6 \cdots}_{\geq 4} h \qquad \exp(ptp) \leq 8/6 = 4/3;$$
$$|h| = 3 \qquad h\ \underbrace{\cdots v_2 L_1}_{\geq 6}\ h \qquad \exp(ptp) \leq 12/9 = 4/3,$$

как в случае V-20 $\quad 4\overbrace{23 \underbrace{\cdots}_{\geq 3} 3}^{\geq 3} 46216\ 423 \quad \exp(423\cdots 423) \leq 15/12 < 4/3;$

$$h\ \underbrace{L_1 u_1 \cdots}_{\geq 6}\ h \qquad \exp(ptp) \leq 12/9 = 4/3;$$
$$h\ \underbrace{\cdots u_2 L_2}_{\geq 6}\ h \qquad \exp(ptp) \leq 12/9 = 4/3;$$
$$h\ \underbrace{L_2 v_1 \cdots}_{\geq 6}\ h \qquad \exp(ptp) \leq 12/9 = 4/3;$$

В этом случае также могут появиться повторы вида $h1\cdots h1$, $1h\cdots 1h$, $h6\cdots h6$, $6h\cdots 6h$.

$$
\begin{array}{llll}
|h|=1 & h1\ \underbrace{6d_1 1}_{\geq 3}\underbrace{\cdots}_{\geq 2}\ h1 & & \exp(ptp)\leq 9/7<4/3;\\
 & 1h\ \underbrace{\cdots}_{\geq 2}\underbrace{1d_2 6}_{\geq 3}\ 1h & & \exp(ptp)\leq 9/7<4/3;\\
 & h6\ \underbrace{1\bar{d}_1 6}_{\geq 3}\underbrace{\cdots}_{\geq 2}\ h6 & & \exp(ptp)\leq 9/7<4/3;\\
 & 6h\ \underbrace{\cdots}_{\geq 2}\underbrace{6\bar{d}_2 1}_{\geq 3}\ 6h & & \exp(ptp)\leq 9/7<4/3.
\end{array}
$$

Для $2 \leq |h| \leq 3$ рассмотрим только наиболее технически сложные случаи, которые позволяют продемонстрировать практически все необходимые приемы.

В случае **V-11** (верхняя часть)

$$6\,\overbrace{42\underbrace{\cdots}_{\geq 6}42}^{\geq 3}651\ 642 \qquad \exp(642\cdots 642)\leq 17/14<4/3,$$

здесь мы используем то, что $(5/4)^+$-свободное слово v не может содержать подслов $6\,42\,6$ и $42\underbrace{\cdots}_{<6}42$;

$$6423\ \underbrace{\cdots}_{\geq 5}42651\ 6423 \qquad \exp(6423\cdots 6423)\leq 18/14<4/3.$$

В случае **V-10** (верхняя часть)

$$642\ \underbrace{\cdots}_{\geq 1}63451\ 642 \qquad \exp(642\cdots 642)\leq 12/9=4/3;$$

$6423\ \underbrace{\cdots}_{\geq 2}63451\ 6423$, однако если в промежутке ровно два символа, то это могут быть только 54 (действительно, 42354 – единственное $(5/4)^+$-свободное слово длины 5, начинающееся на 423, которое может быть заключено между двумя 6), но тогда в слове v появится повтор $4\,63\,4$ с экспонентой $4/3$, что невозможно, значит

$$6423\ \underbrace{\cdots}_{\geq 3}63451\ 6423 \qquad \exp(6423\cdots 6423)\leq 16/12=4/3.$$

В случае **V-8** (нижняя часть и четное n)

$$124\ \underbrace{\cdots}_{\geq 2}25136\ 124 \qquad \exp(124\cdots 124)\leq 13/10<4/3;$$

$1245\ \underbrace{\cdots}_{\geq 2}25136\ 1245$, однако если в промежутке ровно два символа, то среди них обязательно есть 1 (над алфавитом $\{2,3,4,5\}$ нельзя построить $(5/4)^+$-свободное слово длины 7), но в слове $1245\,1a25136\ 1245$, где $a\in A$, есть запрещенный повтор $51\,a2\,51$ с экспонентой $3/2$, а в слове $1245\,a125136\ 1245$ запрещенный повтор $1\,25\,1$ с экспонентой $4/3$, значит

$$1245 \underbrace{\cdots}_{\geq 3} 25136\ 1245 \qquad \exp(1245 \cdots 1245) \leq 16/12 = 4/3.$$

Таким образом, для любого $n \in \mathbb{N}$ построено циклическое $(4/3)^{+}$-свободное слово длины n над 6-буквенным алфавитом. $\square$

Следствие 2.4. $\mathrm{CRT}(6) \leq 4/3$.

2.2.4. Случай $\mathbf{k = 7}$

Теорема 2.4. *Над 7-буквенным алфавитом существует циклическое $(5/4)^{+}$-свободное слово любой длины.*

Доказательство. Для любого $n \in \mathbb{N}$ построим циклическое $(5/4)^{+}$-свободное слово (w) длины n над алфавитом $\Sigma_7 = \{1, 2, 3, 4, 5, 6, 7\}$. Выделим три случая относительно значения n:

- если $n \leq 7$, то требуемое слово (w) получается тривиально;
- если $8 \leq n \leq 17$, то искомое слово (w) изображено на рис. 2.14;
- если $n \geq 18$, то слово (w) будем искать в виде $w = L_1 u L_2 v$.

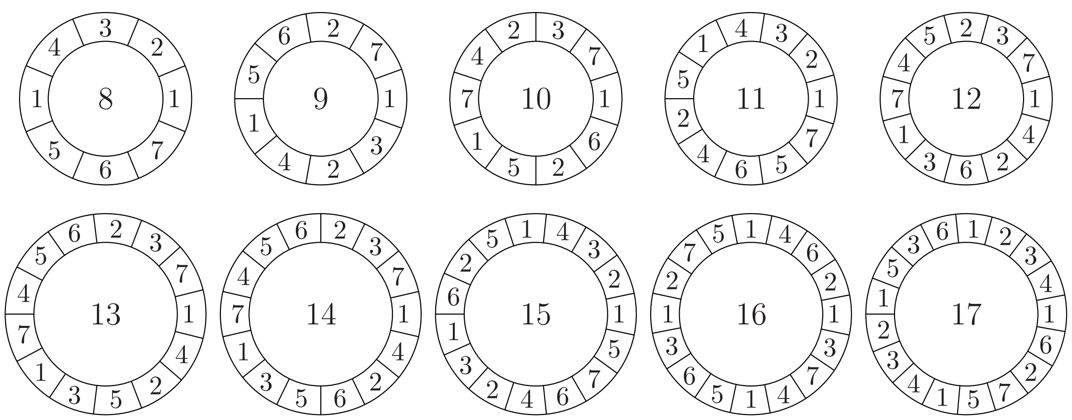


Рис. 2.14. $(5/4)^{+}$*-свободное слово* (w) *длины* n*, где* $8 \leq n \leq 17$ *(длина слова указана внутри круга).*

Разберем последний случай. Здесь u – произвольное $(5/4)^{+}$-свободное слово над алфавитом $A = \{1, 2, 3, 4, 5\}$, длина которого равна $\lceil (n-8)/2 \rceil$; v – $(5/4)^{+}$-свободное слово над алфавитом $B = \{3, 4, 5, 6, 7\}$, которое равно $f(u)$ (в случае четного n) или $f(u)$ без первой буквы (когда n нечетно), где $f : A \to B$ – биекция такая, что $f(1) = 7,\ f(2) = 6,\ f(3) = 3,\ f(4) = 4,\ f(5) = 5$; $L_1 = 1627$; $L_2 \in \{7261, 6172\}$.

За счет увеличения длины вставок L_1 и L_2, теперь можно рассматривать более короткие префиксы и суффиксы слов u и v, что позволит сократить число возможных вариантов расположения общих символов в них, а значит и число рассматриваемых случаев. Обозначим через u_1 (соответственно, v_1) префикс длины 3 слова u (соответственно, v); а через u_2 (соответственно, v_2) – суффикс длины 2 слова u (соответственно, v). Заметим, что в словах u_1, u_2, v_1, v_2 все буквы различны. В зависимости от того, на каких позициях встречаются одинаковые буквы в словах u_1 и u_2 (v_1 и v_2), можно выделить 13 случаев (см. рис. 2.15).

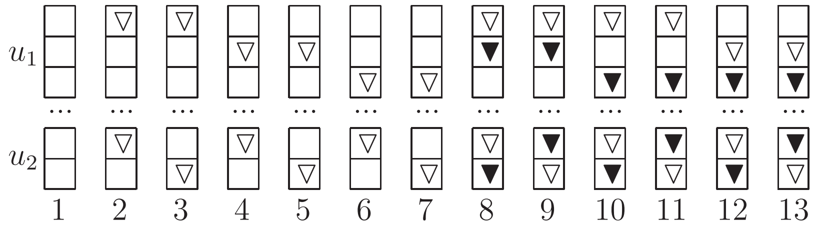


Рис. 2.15. *Варианты расположения общих символов в словах u_1 и u_2 ($k = 7$). Общие символы обозначены с помощью $\triangledown$ и $\blacktriangledown$.*

Если n четно, то u и $v = f(u)$ будут иметь одинаковый тип. Для каждого типа переставляем буквы в словах u и v так, чтобы их префиксы и суффиксы выглядели как на рис. 2.16. Если же n нечетно, то тип слова v определяется по типу слова u с помощью таблицы 2.1, а соответствующие перестановки в u и v представлены на рис. 2.17.

Наблюдения 2.5, 2.2 и 2.3 будут справедливы и в случае $k = 7$, только экспонента будет $5/4$, а длина наибольшего общего подслова u и v будет равна 3 (это слово 345 и все полученные из него с помощью перестановки букв).

Проверим, что в построенных словах все возникающие повторы вида ptp, где оба вхождения p не являются подсловами u (соответственно, v), имеют разрешенную экспоненту.

Пусть $a \in A \setminus B = \{1, 2\}, b \in B \setminus A = \{6, 7\}$, тогда

$$a \underbrace{\cdots}_{\geq 3} a \qquad \exp(ptp) \leq 5/4;$$
$$b \underbrace{\cdots}_{\geq 3} b \qquad \exp(ptp) \leq 5/4.$$

Заметим, что повторы вида $ab \cdots ab$ и $ba \cdots ba$ в построенных словах не появляются.

Пусть $h \in \{3, 4, 5\}^*$ – общее подслово слов u и v, тогда $|h| \leq 3$.

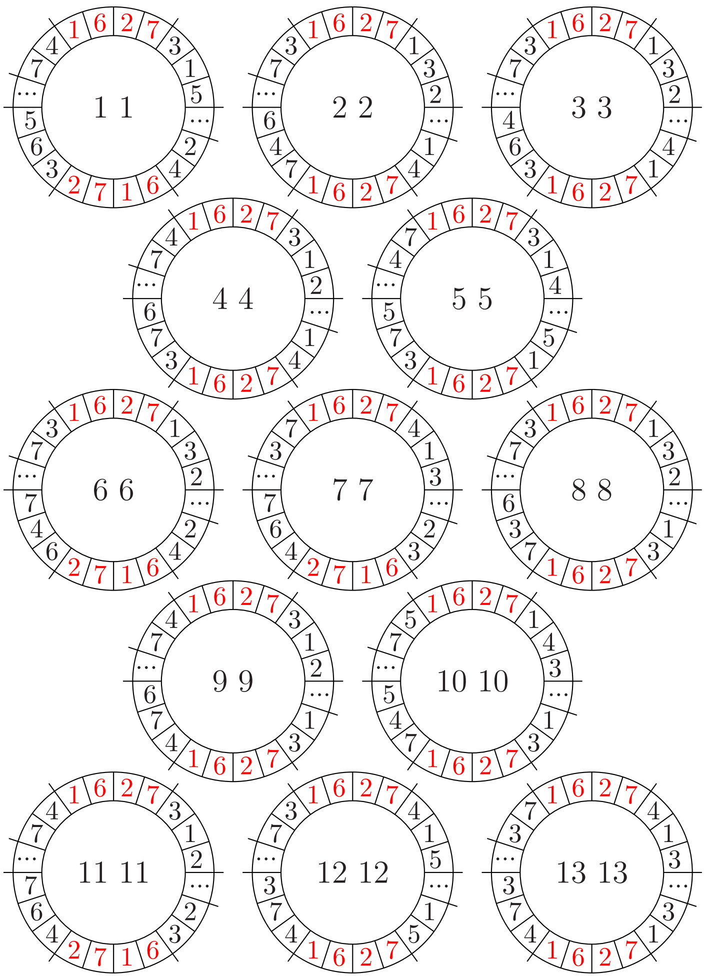


Рис. 2.16. $(5/4)^+$-*свободное слово* (w) *длины* n, *где* $n \geq 18$ *и четно (внутри круга указан тип* v *и* u*).*

Таблица 2.1. *Зависимость типа* v *от типа* u *в случае нечетного* n $(k = 7)$.

Тип u	1	2	3	4	5	6	7	8	9	10	11	12	13
Возможный тип v	6 7	1 7	1 6	2 10	3 11	4 12	5 13	3	2	5	4	8	9

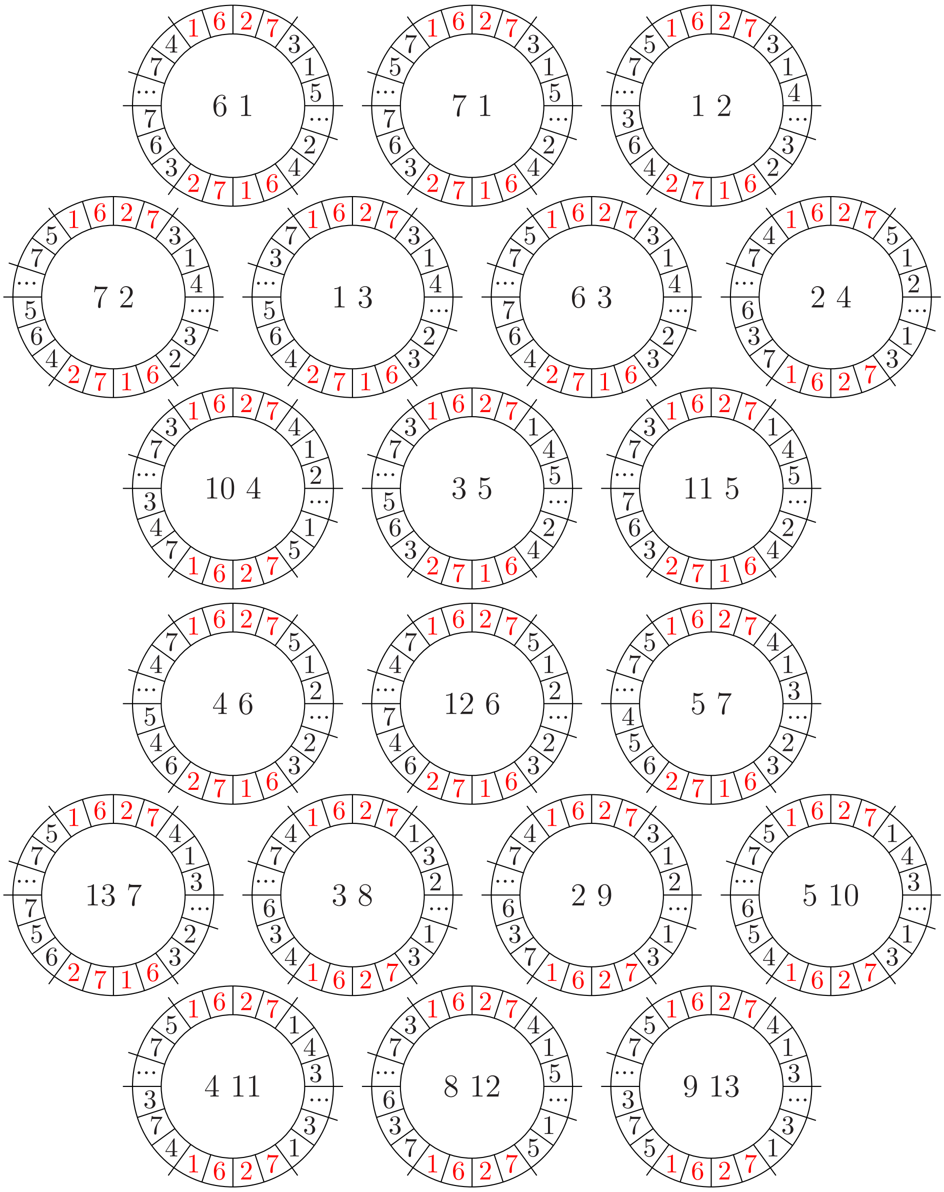


Рис. 2.17. $(5/4)^{+}$-*свободное слово* (w) *длины* n, *где* $n \geq 18$ *и нечетно (внутри круга указан тип* v *и* u*).*

$$
\begin{array}{lcl}
|h| = 1 & h\underbrace{\cdots L_1 \cdots}_{\geq 4}h & \exp(ptp) \leq 6/5 < 5/4; \\
 & h\underbrace{\cdots L_2 \cdots}_{\geq 4}h & \exp(ptp) \leq 6/5 < 5/4; \\
|h| = 2 & h\underbrace{\cdots b \cdots L_1 \cdots a \cdots}_{\geq 7}h & \exp(ptp) \leq 11/9 < 5/4; \\
 & h\underbrace{\cdots a \cdots L_2 \cdots b \cdots}_{\geq 7}h & \exp(ptp) \leq 11/9 < 5/4.
\end{array}
$$

Для $|h| = 2$ исключение составляют случаи $(3, 8)$ и $(5, 10)$, здесь первая координата – это тип слова v, вторая – тип u. Рассмотрим их отдельно:

$$
\begin{array}{lcl}
(3,8) & 43\underbrace{\cdots}_{\geq 2}137261\ 43 & \exp(43\cdots 43) \leq 12/10 < 5/4; \\
(5,10) & 45\underbrace{\cdots}_{\geq 0}137261\ 45 & \exp(45\cdots 45) \leq 10/8 = 5/4.
\end{array}
$$

Для $|h| = 3$ удается избежать запрещенных повторов вида $h \cdots h$ благодаря некоторым приемам, которые можно проиллюстрировать на конкретном примере.

В случае **(1,1)** проверим отсутствие подслов вида $h \cdots 74162731\ h$ с запрещенной экспонентой. Зная, что после 1 в слове u идет 5, получаем, что $h = 543$ (иначе в слове u появится запрещенный повтор $3\underbrace{\cdots}_{<3}3$). Тогда

$$
543\underbrace{\cdots}_{\geq 1}74162731\ 543 \quad \exp(543\cdots 543) \leq 15/12 = 5/4.
$$

Если рассмотреть подслова вида $h\,74162731 \cdots h$, то возможны два варианта: $h = 435$ или $h = 453$ (иначе появится запрещенный повтор $4\underbrace{\cdots}_{<3}4$). Тогда

$$
\begin{array}{ll}
435\ 741627315\underbrace{\cdots}_{\geq 1}435 & \exp(435\cdots 435) \leq 16/13 < 5/4; \\
453\ 741627315\underbrace{\cdots}_{\geq 2}453 & \exp(453\cdots 453) \leq 17/14 < 5/4.
\end{array}
$$

Однако в случаях **(1, 3)**, **(4, 6)**, **(12, 6)** сделанных перестановок может оказаться недостаточно, чтобы гарантировать отсутствие запрещенных подслов вида $h \cdots h$. А именно, могут появиться следующие подслова:

$$
\begin{array}{ccl}
(1,3) & 453\ 7162731\ 453 & \exp(453\cdots 453) = 13/10 > 5/4; \\
(4,6) \text{ и } (12,6) & 354\ 71627512\ 354 & \exp(354\cdots 354) = 14/11 > 5/4.
\end{array}
$$

Если возникла одна из описанных ситуаций, то необходимо дополнительно переименовать буквы 4 и 5 только в слове u.

Теперь рассмотрим повторы вида $ha \cdots ha$, $ah \cdots ah$, $hb \cdots hb$, $bh \cdots bh$. Отметим, что в этом случае $|h| \leq 2$.

$$
\begin{array}{lll}
|h| = 1 & ha\underbrace{\cdots}_{\geq 3}a\underbrace{\cdots}_{\geq 2}ha & \exp(ptp) \leq 10/8 = 5/4; \\
 & ah\underbrace{\cdots}_{\geq 2}a\underbrace{\cdots}_{\geq 3}ah & \exp(ptp) \leq 10/8 = 5/4; \\
 & hb\underbrace{\cdots}_{\geq 3}b\underbrace{\cdots}_{\geq 2}hb & \exp(ptp) \leq 10/8 = 5/4; \\
 & bh\underbrace{\cdots}_{\geq 2}b\underbrace{\cdots}_{\geq 3}bh & \exp(ptp) \leq 10/8 = 5/4.
\end{array}
$$

Для $|h| = 2$ подобные подслова могут появиться только в случаях **(3, 8)** и **(5, 10)**. Причем с помощью сделанных перестановок нам не всегда удается избежать появления запрещенных повторов. В построенных словах могут появиться следующие запрещенные подслова:

$$
\begin{array}{lll}
(3,8) & 143\underbrace{\cdots}_{2-3}13726\,143 & \exp(143\cdots143) \geq 14/11 > 5/4; \\
(5,10) & 145\underbrace{\cdots}_{1-3}13726\,145 & \exp(145\cdots145) \geq 14/11 > 5/4.
\end{array}
$$

Если в случае (3, 8) появилось описанное подслово, то необходимо дополнительно переименовать буквы 4 и 5 только в слове u; а в случае (5, 10) необходимо переименовать буквы 2 и 5 в слове u.

Повторы вида $bha\cdots bha$ и $ahb\cdots ahb$ в полученных словах отсутствуют.

Таким образом, для любого $n \in \mathbb{N}$ построено циклическое $(5/4)^+$-свободное слово длины n над 7-буквенным алфавитом. □

Следствие 2.5. $\mathrm{CRT}(7) \leq 5/4$.

2.2.5. Случай k = 9

Теорема 2.5. *Над 9-буквенным алфавитом существует циклическое $(6/5)^+$-свободное слово любой длины.*

Доказательство. Для любого $n \in \mathbb{N}$ построим циклическое $(6/5)^+$-свободное слово (w) длины n над алфавитом $\Sigma_9 = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Выделим три случая относительно значения n:

- если $n \leq 9$, то требуемое слово (w) получается тривиально;
- если $10 \leq n \leq 23$, то искомое слово (w) изображено на рис. 2.18;
- если $n \geq 24$, то слово (w) будем искать в виде $w = L_1uL_2v$.

Разберем последний случай. Здесь u – произвольное $(6/5)^+$-свободное слово над алфавитом $A = \{1, 2, 3, 4, 5, 6\}$, длина которого равна $\lceil (n-12)/2 \rceil$; v – $(6/5)^+$-свободное слово над алфавитом $B = \{4, 5, 6, 7, 8, 9\}$, которое равно $f(u)$ (в случае

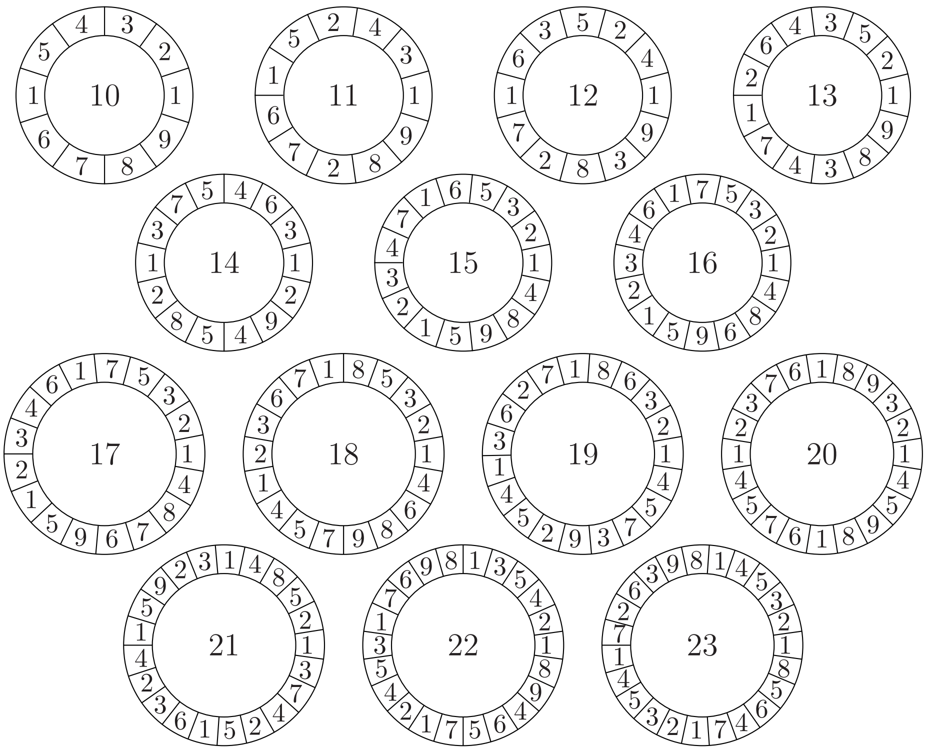


Рис. 2.18. $(6/5)^+$*-свободное слово* (w) *длины* n*, где* $10 \le n \le 23$ *(длина слова указана внутри круга).*

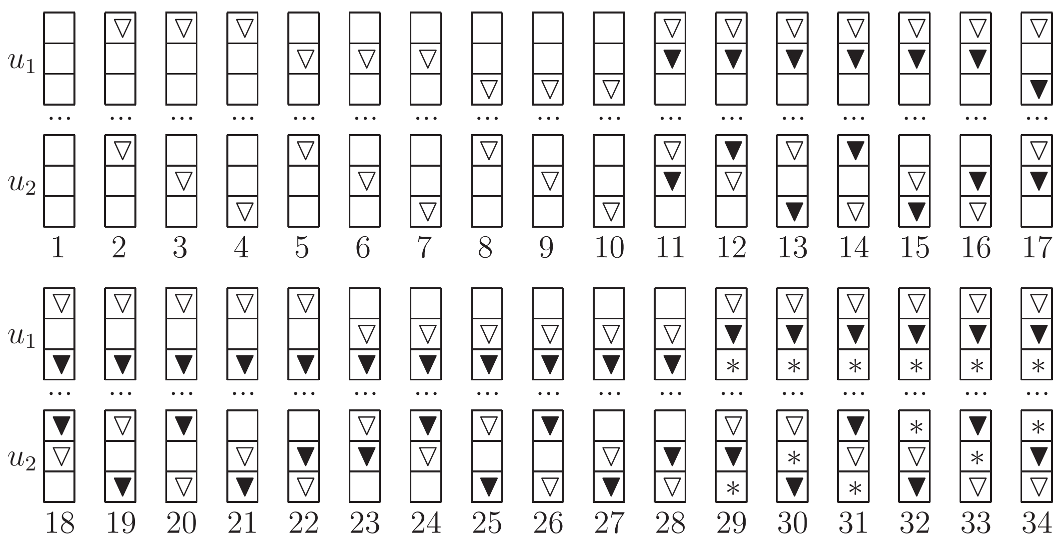


Рис. 2.19. *Варианты расположения общих символов в словах* u_1 *и* u_2 $(k = 9)$*.*

четного n) или $f(u)$ без первой буквы (когда n нечетно), где $f : A \to B$ – биекция такая, что $f(1) = 9,\ f(2) = 8,\ f(3) = 7,\ f(4) = 4,\ f(5) = 5,\ f(6) = 6$; $L_1 = 183729$; $L_2 \in \{738192, 739281, 927381, 928173\}$.

Обозначим через u_1 (соответственно, v_1) и u_2 (соответственно, v_2) – префикс и суффикс длины 3 слова u (соответственно, v). Заметим, что в словах u_1, u_2, v_1, v_2 все буквы различны. В зависимости от того, на каких позициях встречаются одинаковые буквы в словах u_1 и u_2 (v_1 и v_2), можно выделить 34 случая (см. рис. 2.19). Так как никаких дополнительных условий на слово u мы не накладывали, то можно отбросить симметричные относительно взятия реверса варианты. Тогда из следующих пар можно оставить только первые варианты: 2 и 10; 3 и 7; 5 и 9; 11 и 27; 12 и 28; 13 и 21; 14 и 22; 17 и 25; 18 и 26; 30 и 31.

Если n четно, то u и $v = f(u)$ будут иметь одинаковый тип. Для каждого типа переставляем буквы в словах u и v так, чтобы их префиксы и суффиксы выглядели как на рис. 2.20. Если же n нечетно, то тип слова v определяется по типу u с помощью таблицы 2.2. Соответствующие перестановки в словах u и v представлены на рис. 2.21.

Таблица 2.2. *Зависимость типа v от типа u в случае нечетного n ($k = 9$).*

Тип u	1	2	3	4	5	6	8	11	12	13	14	15	16	17	18	19	20	23	24	29	30	32	33	34
Тип v	2 5 8	1 2 5	1 2 8	1 5 8	2 17 19	3 13 18	5 17 23	3 13	2 19	4 14	2 17	4 20	3 18	6 11	5 17	3 12	5 23	11 29	12 30	15	16	14	11	12

Наблюдения 2.5, 2.2 и 2.3 будут справедливы и в случае $k = 9$, только экспонента будет $6/5$, а длина наибольшего общего подслова u и v будет равна 3 (это слово 456 и все полученные из него с помощью перестановки букв). Проверим, что полученные в каждом случае циклические слова не содержат подслов вида ptp, где оба вхождения p не являются подсловами u (соответственно, v), с запрещенной экспонентой.

Пусть $a \in \{1, 2, 3\}$, $b \in \{7, 8, 9\}$, тогда всегда разрешены следующие повторы:

$$a\underbrace{\cdots}_{\geq 4}a \qquad \exp(ptp) \leq 6/5;$$
$$b\underbrace{\cdots}_{\geq 4}b \qquad \exp(ptp) \leq 6/5.$$

Заметим, что повторы вида $ab\cdots ab$ и $ba\cdots ba$ в построенных словах не появляются.

Пусть $h \in \{3, 4, 5\}^*$ – общее подслово слов u и v, тогда $|h| \leq 3$.

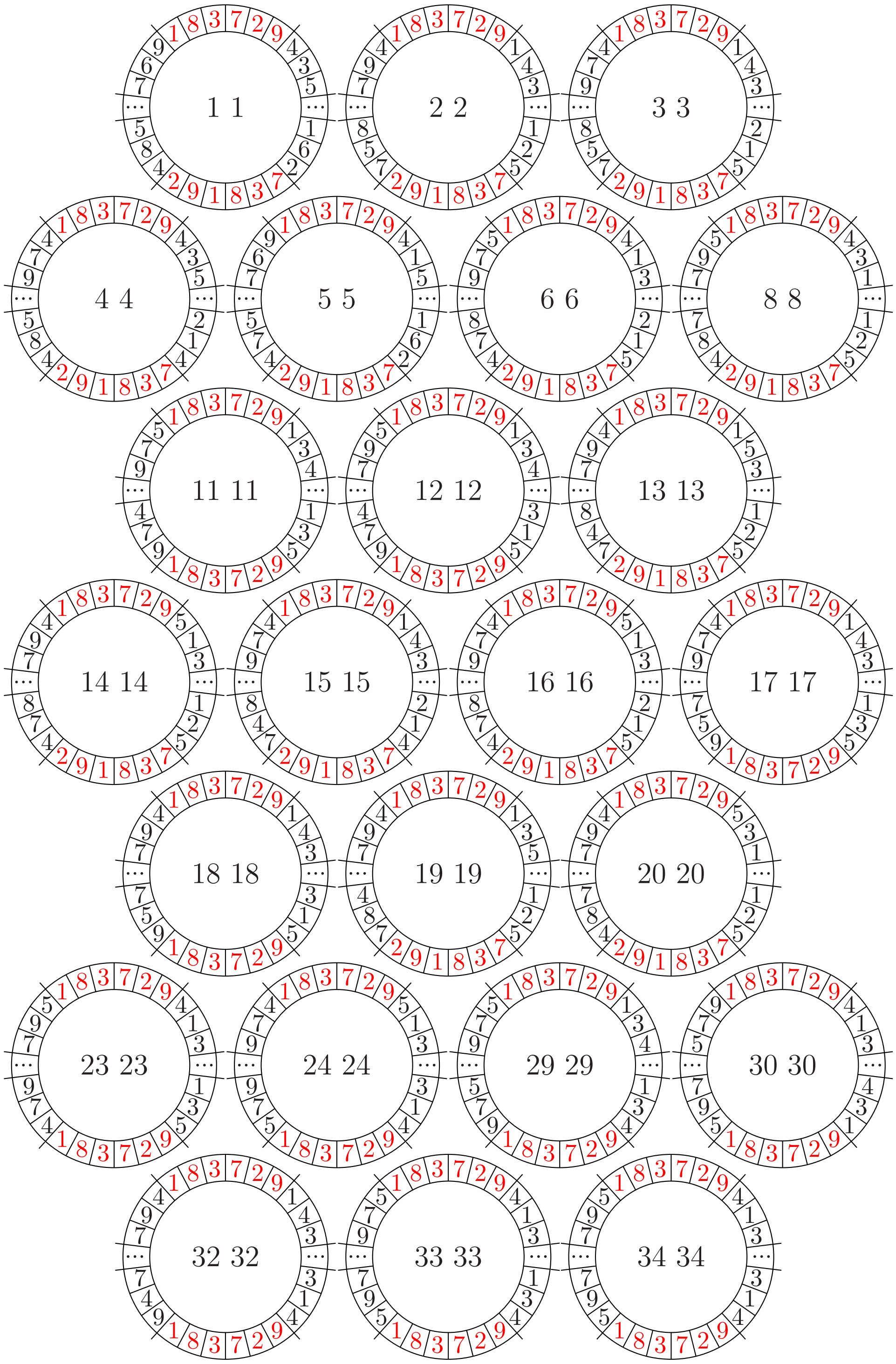


Рис. 2.20. $(6/5)^+$-*свободное слово* (w) *длины* n, *где* $n \geq 24$ *и четно (внутри круга указан тип* v *и тип* u*).*

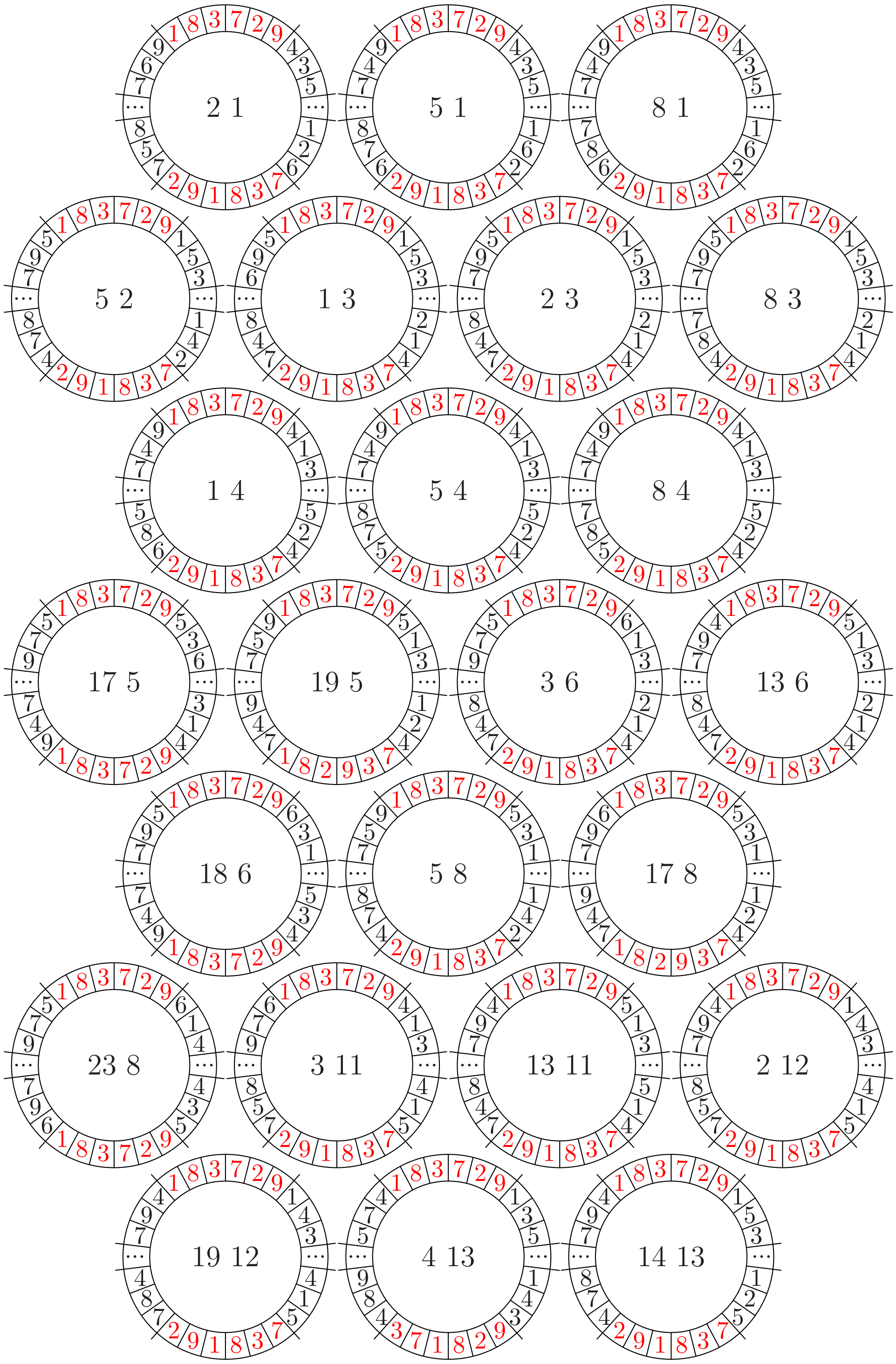


Рис. 2.21. $(6/5)^+$*-свободное слово* (w) *длины* n*, где* $n \geq 24$ *и нечетно (внутри круга указан тип* v *и тип* u*).*

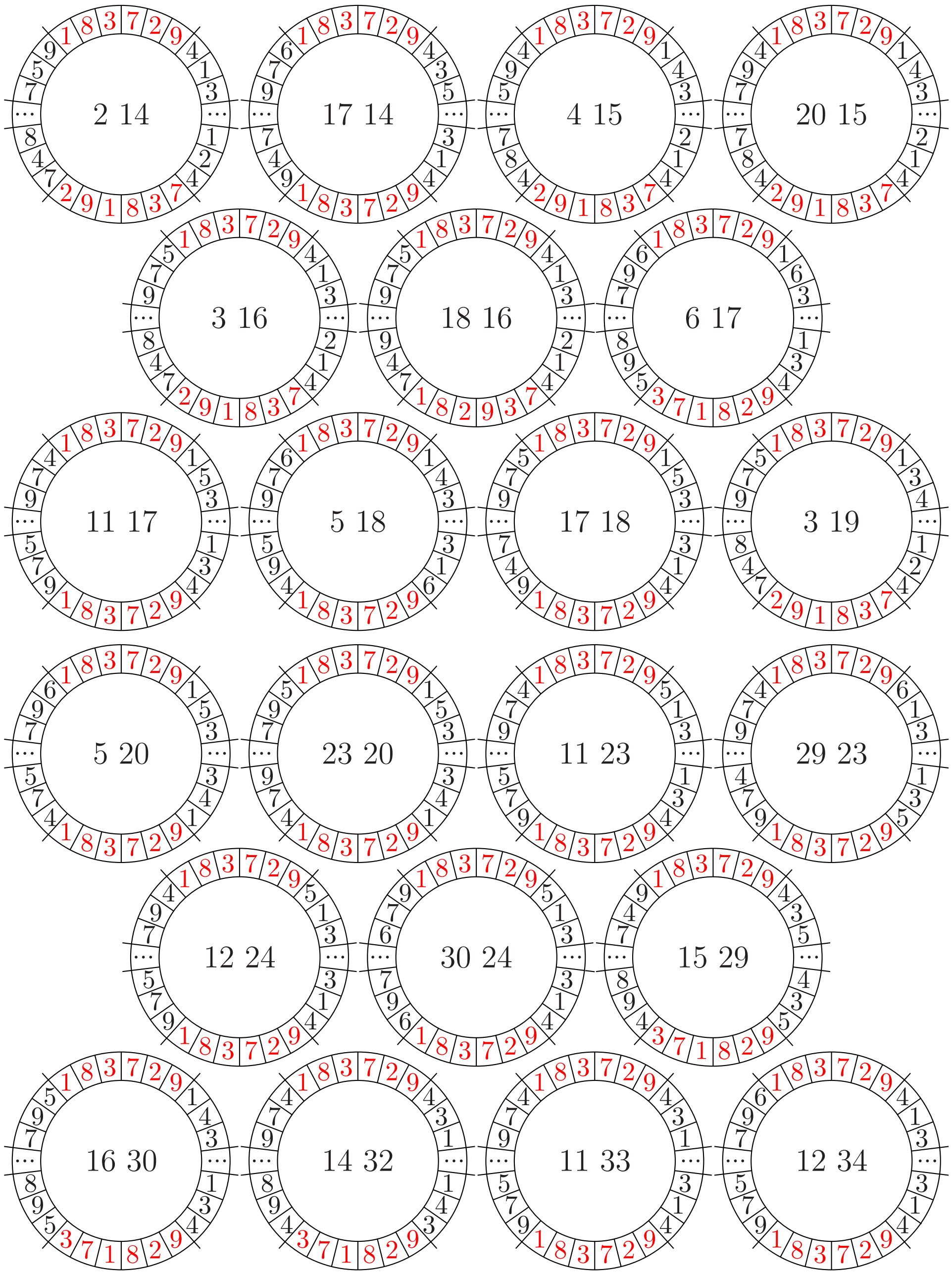


Рис. 2.22. $(6/5)^+$-*свободное слово* (w) *длины* n, *где* $n \geq 24$ *и нечетно (продолжение).*

$$
\begin{array}{lll}
|h| \le 2 & h\underbrace{\cdots b \cdots L_1 \cdots a \cdots}_{\ge 9}h & \exp(ptp) \le 13/11 < 6/5; \\
& h\underbrace{\cdots a \cdots L_2 \cdots b \cdots}_{\ge 9}h & \exp(ptp) \le 13/11 < 6/5; \\
|h| = 3 & h\underbrace{\cdots v_2 L_1 u_1 \cdots}_{\ge 12}h & \exp(ptp) \le 18/15 = 6/5; \\
& h\underbrace{\cdots u_2 L_2 v_1 \cdots}_{\ge 12}h & \exp(ptp) \le 18/15 = 6/5;
\end{array}
$$

например, в случае **(1, 1)** при $|h| = 3$ проверим отсутствие подслов $h \cdots 76918372943h$. Если $h = 546$, то в слове u появится повтор $4\,35\,4$ с экспонентой $4/3$; а если $h = 564$, то появится подслово $4\,356\,4$ с экспонентой $5/4$, что невозможно, так как слово u является $(6/5)^+$-свободным.

При $|h| = 3$ появляются и исключения, например, как в случае $(12, 12)$ (слово h начинается в u_1 (аналогично u_2, v_1, v_2), хотя подслово $h \cdots h$ по-прежнему имеет разрешенную экспоненту). Отметим, что в этом случае возможно два варианта: $h = 456$ или $h = 465$.

$$
\begin{array}{ll}
456\underbrace{\cdots}_{\ge 1}79518372913\,456 & \exp(ptp) \le 18/15 = 6/5; \\
465\underbrace{\cdots}_{\ge 2}79518372913\,465 & \exp(ptp) \le 19/16 < 6/5.
\end{array}
$$

Благодаря сделанным перестановкам повторы вида $ha \cdots ha$, $ah \cdots ah$, $hb \cdots hb$ или $bh \cdots bh$ возможны только при $|h| = 1$.

$$
\begin{array}{ll}
ha\underbrace{L_1' u_1 \cdots}_{\ge 8}ha & \exp(ptp) \le 12/10 = 6/5; \\
ah\underbrace{\cdots u_2 L_2'}_{\ge 8}ah & \exp(ptp) \le 12/10 = 6/5; \\
hb\underbrace{L_2' v_1 \cdots}_{\ge 8}hb & \exp(ptp) \le 12/10 = 6/5; \\
bh\underbrace{\cdots v_2 L_1'}_{\ge 8}bh & \exp(ptp) \le 12/10 = 6/5.
\end{array}
$$

Таким образом, для любого $n \in \mathbb{N}$ построено циклическое $(6/5)^+$-свободное слово длины n над 9-буквенным алфавитом. $\square$

Следствие 2.6. $\mathrm{CRT}(9) \le 6/5$.

2.2.6. Случай нечетных k ≥ 11

Теорема 2.6. *Для любого нечетного $k \ge 11$ существует циклическое $\left(\frac{\lceil k/2 \rceil + 1}{\lceil k/2 \rceil}\right)^+$-свободное слово над k-буквенным алфавитом любой длины.*

Доказательство. Положим $r = \lceil k/2 \rceil$. Для любого $n \in \mathbb{N}$ построим циклическое $((r+1)/r)^+$-свободное слово (w) длины n над алфавитом $\Sigma_k = \{1, \ldots, k\}$. Выделим три случая относительно значения n:

- случай $n \leq k$ тривиален;
- случай $k+1 \leq n \leq 2k+5$ будет рассмотрен позже;
- в случае $n \geq 2k+6$ слово (w) будем искать в виде $w = L_1 u L_2 v$.

Остановимся подробнее на последнем случае. Здесь

- u – произвольное $((r+1)/r)^+$-свободное слово над алфавитом $A = \{1, \ldots, r+1\}$, длина которого равна $\lceil (n-4)/2 \rceil$, дополнительно потребуем, чтобы $u[1] = u[r+1]$;
- v – $((r+1)/r)^+$-свободное слово над алфавитом $B = \{r-1, \ldots, k\}$, длина которого равна $\lfloor (n-4)/2 \rfloor$;
- $L_1 = a_{i_0} b_{i_0}$, где $a_{i_0} \in A \smallsetminus B$, $b_{i_0} \in B \smallsetminus A$;
- $L_2 = b_{j_0} a_{j_0}$, где $a_{j_0} \in A \smallsetminus B$, $b_{j_0} \in B \smallsetminus A$.

Обозначим общие буквы алфавитов A и B через x, y и z. Пусть $f : A \to B$ – биекция такая, что $f(x) = x,\ f(y) = y,\ f(z) = z$. Тогда при четном n возьмем в качестве v слово $f(u)$, а при нечетном n слово $f(u)$ без первой буквы. Через $u_1 = a_{i_1} \cdots a_{i_r}$ и $u_2 = a_{j_r} \cdots a_{j_1}$, ($v_1 = b_{j_1} \cdots b_{j_r}$ и $v_2 = b_{i_r} \cdots b_{i_1}$) обозначим префикс и суффикс слова u (соответственно, v) длины r. Причем u_1 и u_2 (v_1 и v_2) не перекрываются, и все буквы в каждом из этих слов различны (см. условие (1) Пансьё).

Можно выделить два случая в зависимости от количества различных букв в словах u_1 и u_2:

I. $\Sigma(u_1) = \Sigma(u_2)$ (а также $\Sigma(v_1) = \Sigma(v_2)$);

II. $\Sigma(u_1)$ и $\Sigma(u_2)$ (а также $\Sigma(v_1)$ и $\Sigma(v_2)$) различаются всего одной буквой.

Начиная с $k = 11$, в нашем распоряжении находится достаточное количество «необщих» букв для слов u и v, поэтому чтобы избежать запрещенных подслов необязательно знать все буквы в выделенных префиксах и суффиксах, достаточно лишь правильно выбрать позиции для общих букв x, y и z. Выбор позиций в слове u автоматически будет переноситься и на слово $f(u)$, а стало быть и на слово v (за исключением нескольких случаев, требующих специфичного выбора позиций в слове v). А расстановка букв x, y, z на выбранных позициях в словах u и v будет осуществляться по-разному.

Случай I.

Пусть $\hat{a} = A \setminus \Sigma(u_1)$. Тогда переименуем буквы $\hat{a}$ и z в слове u (соответственно, $f(\hat{a})$ и z в слове v). Каждой букве $a \in \Sigma(u_1)$ поставим в соответствие пару позиций $g(a) = (s_1, s_2)$ такую, что $a_{i_{s_1}} = a_{j_{s_2}} = a$. Среди r пар позиций есть по крайней мере $(r-3) \geq 3$ пары, которые не имеют следующий вид: $(1, \_)$, $(\_, 1)$, $(\_, r)$. Из них выберем пару $(\tilde{s}_1, \tilde{s}_2)$ с наибольшей первой координатой и переименуем буквы $g^{-1}(\tilde{s}_1, \tilde{s}_2)$ и x в слове u. Заметим, что $\tilde{s}_1 \in \{r-2, r-1, r\}$. Возьмем другую пару $(\hat{s}_1, \hat{s}_2)$, где $\hat{s}_1 < \tilde{s}_1 - 1$ (в частности можно выбрать пару с наименьшей первой координатой). Переименуем буквы $g^{-1}(\hat{s}_1, \hat{s}_2)$ и y в слове u. Теперь в словах u и v буквы x, y, z стоят на одинаковых выбранных позициях. Осталось переименовать буквы y и z только в слове v.

Вставки L_1 и L_2 определим следующим образом: $a_{j_0} = a_{j_r}$, $a_{i_0} \in \{a_{i_{r-1}}, a_{i_r}\} \setminus \{x\}$, $b_{i_0} = b_{i_r}$, $b_{j_0} \in \{b_{j_{r-1}}, b_{j_r}\} \setminus \{x\}$, если n четно (см. рис. 2.23, I(e)), $b_{j_0} \in \{b_{j_1}, b_{j_r}\} \setminus \{x\}$, если n нечетно (см. рис. 2.23, I(o)).

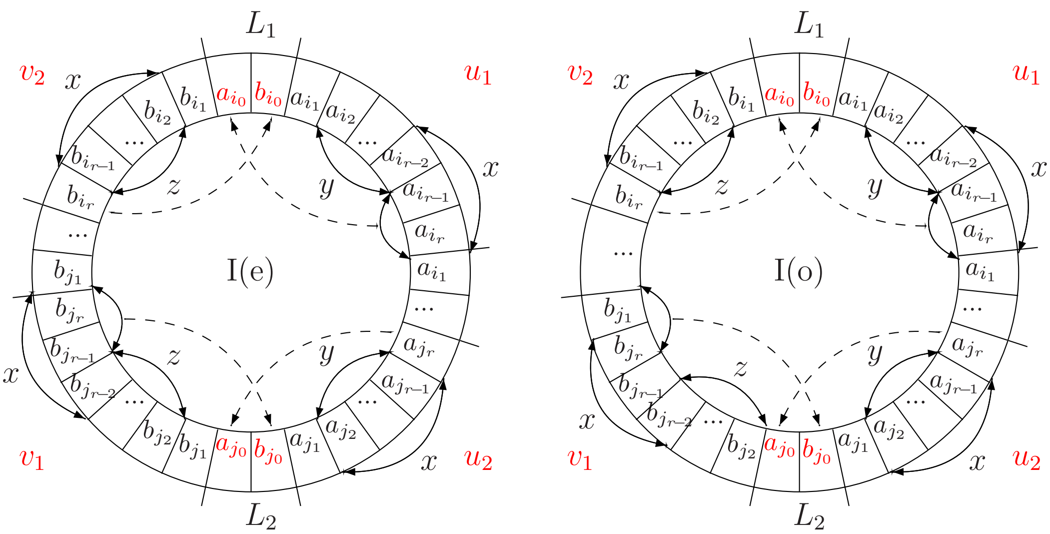


Рис. 2.23. *Циклическое слово* (w) *длины* n, *где* $n \geq 2k+6$. *Случай I.*

Случай II.

Пусть $a_{i_{s_1}} = \Sigma(u_1) \setminus \Sigma(u_2)$ и $a_{j_{s_2}} = \Sigma(u_2) \setminus \Sigma(u_1)$. Переименуем буквы $a_{i_{s_1}}$ и y; $a_{j_{s_2}}$ и z. В зависимости от расположения позиций s_1 и s_2 в слове u выделим 25 подслучаев. Для каждого из них выберем букву, подходящую на роль x. Так как все возможные запрещенные подслова достаточно короткие (в силу наблюдений), то будет удобнее разбить слово (w) на верхнюю (подслово $v_2L_1u_1$) и нижнюю (подслово $u_2L_2v_1$) части, которые играют ключевую роль. Причем для удобства вместо $u_2L_2v_1$ будем рассматривать его реверс $\overleftarrow{u_2L_2v_1}$. Подходящие позиции в u_1 и u_2 (v_1 и v_2) будут выделены фигурной скобкой. Так как $r \geq 6$, то среди выделенных позиций обязательно найдется хотя бы одна пара $(\hat{s}_1, \hat{s}_2)$, на которой стоит одна и та же буква, т.е. $(\hat{s}_1, \hat{s}_2) = g(a)$, где $a \in \Sigma(u_1)$. Переименуем буквы a и x в слове u. Тогда в слове v будут переименованы буквы $f(a)$ и x, если не указано ничего другого. Кроме того, в слове v возможны дополнительные перестановки между буквами x, y и z.

Слова $L_1 = a_{i_0}b_{i_0}$ и $L_2 = b_{j_0}a_{j_0}$ выбираем таким образом, чтобы выполнялись следующие условия:

(1) $a_{i_0} \in \{a_{i_{r-1}}, a_{i_r}\}$, $a_{j_0} \in \{a_{j_{r-1}}, a_{j_r}\}$, $b_{i_0} \in \{b_{i_{r-1}}, b_{i_r}\}$, $b_{j_0} \in \{b_{j_{r-1}}, b_{j_r}\}$, если n четно, и $b_{j_0} \in \{b_{j_1}, b_{j_r}\}$, если n нечетно;

(2) $\{a_{i_0}, a_{j_0}, b_{i_0}, b_{j_0}\} \cap \{x, y, z\} = \varnothing$;

(3) если возможно, $a_{i_0} \neq a_{j_0}$, $b_{i_0} \neq b_{j_0}$.

Заметим, что в силу сделанных перестановок в словах u и v вставки L_1 и L_2 в каждом случае можно выбрать так, чтобы условия (1) и (2) обязательно выполнялись.

Случай 1: $(s_1, s_2) = (1, 1)$.

$v_2L_1u_1$: $b_{i_r}\underbrace{b_{i_{r-1}}\ \cdots\ b_{i_3}}_{x} b_{i_2} z\ a_{i_0}b_{i_0}\ y a_{i_2}\cdots \underbrace{a_{i_{r-3}}a_{i_{r-2}}a_{i_{r-1}}}_{x} a_{i_r}$

$\overleftarrow{u_2L_2v_1}$ (e) : $b_{j_r}\underbrace{b_{j_{r-1}}b_{j_{r-2}}b_{j_{r-3}}}_{x}\cdots b_{j_2} y\ a_{j_0}b_{j_0}\ z a_{j_2}\underbrace{a_{j_3}\ \cdots\ a_{j_{r-1}}}_{x} a_{j_r}$

$\overleftarrow{u_2L_2v_1}$ (o) : $y b_{j_r}\underbrace{b_{j_{r-1}}b_{j_{r-2}}b_{j_{r-3}}}_{x}\cdots b_{j_2}\ a_{j_0}b_{j_0}\ z a_{j_2}\underbrace{a_{j_3}\ \cdots\ a_{j_{r-1}}}_{x} a_{j_r}$

Случай 2: $(s_1, s_2) = (2, 2)$.

$v_2L_1u_1$ (e) : $b_{i_r}b_{i_{r-1}}\underbrace{b_{i_{r-2}}\ \cdots\ b_{i_3}}_{x} z b_{i_1}\ a_{i_0}b_{i_0}\ a_{i_1} y\cdots \underbrace{a_{i_{r-3}}a_{i_{r-2}}a_{i_{r-1}}a_{i_r}}_{x} [a_{i_1}]$

$\overleftarrow{u_2L_2v_1}$ (e) : $[b_{j_1}]\underbrace{b_{j_r}b_{j_{r-1}}b_{j_{r-2}}b_{j_{r-3}}}_{x}\cdots y b_{j_1}\ a_{j_0}b_{j_0}\ a_{j_1} z \underbrace{a_{j_3}\ \cdots\ a_{j_{r-2}}}_{x} a_{j_{r-1}}a_{j_r}$

Если n нечетно, то, чтобы избежать повтора $xya_{j_0}\cdots xya_{j_0}$, выберем новую позицию для x в слове v. Также потребуем, чтобы $a_{j_0} \in \{a_{j_{r-1}}, a_{j_r}\} \setminus \{a_{i_1}\}$.

$v_2L_1u_1$ (o) : $b_{i_r}\underbrace{b_{i_{r-1}}\ \cdots\ b_{i_3}}_{x} z b_{i_1}\ a_{i_0}b_{i_0}\ a_{i_1} y\cdots \underbrace{a_{i_{r-3}}a_{i_{r-2}}a_{i_{r-1}}a_{i_r}}_{x} [a_{i_1}]$

$\overleftarrow{u_2L_2v_1}$ (o) : $b_{j_1}\underbrace{b_{j_r}b_{j_{r-1}}b_{j_{r-2}}}_{x}\cdots b_{j_3} y\ a_{j_0}b_{j_0}\ a_{j_1} z \underbrace{a_{j_3}\ \cdots a_{j_{r-2}}}_{x} a_{j_{r-1}}a_{j_r}$

Случай 3: $(s_1, s_2) \in \{3, \ldots, r-2\} \times \{3, \ldots, r-2\}$. В слове v переименуем буквы z и x.

$v_2L_1u_1$: $b_{i_r}b_{i_{r-1}}\underbrace{\overbrace{b_{i_{r-2}}\ \cdots\ b_{i_3}}^{x} b_{i_2}b_{i_1}}_{z}\ a_{i_0}b_{i_0}\ a_{i_1}a_{i_2}\overbrace{a_{i_3}\cdots a_{i_{r-3}}a_{i_{r-2}}}^{y}\underbrace{\vphantom{a}}_{} a_{i_{r-1}}a_{i_r}\ [a_{i_1}]$, где $\underbrace{a_{i_{r-3}}a_{i_{r-2}}a_{i_{r-1}}a_{i_r}}_{x}$

$\overleftarrow{u_2L_2v_1}$ (e) : $[b_{j_1}]\ \underbrace{b_{j_r}b_{j_{r-1}}b_{j_{r-2}}b_{j_{r-3}}}_{z}$, $\overbrace{b_{j_{r-2}}b_{j_{r-3}}\cdots b_{j_3}}^{y}$ $b_{j_2}b_{j_1}\ a_{j_0}b_{j_0}\ \underbrace{a_{j_1}a_{j_2}\overbrace{a_{j_3}\ \cdots\ a_{j_{r-2}}}^{z}}_{x} a_{j_{r-1}}a_{j_r}$

$\overleftarrow{u_2L_2v_1}$ (o) : $b_{j_1}\underbrace{b_{j_r}b_{j_{r-1}}b_{j_{r-2}}b_{j_{r-3}}}_{z}$, $\overbrace{b_{j_{r-2}}b_{j_{r-3}}\cdots b_{j_3}}^{y}$ $b_{j_2}\ a_{j_0}b_{j_0}\ \underbrace{a_{j_1}a_{j_2}\overbrace{a_{j_3}\ \cdots\ a_{j_{r-2}}}^{z}}_{x} a_{j_{r-1}}a_{j_r}$

Случай 4: $(s_1, s_2) = (r-1, r-1)$. В слове v переименуем буквы z и x.

$$v_2L_1u_1 \quad : \quad b_{i_r} x \underbrace{b_{i_{r-2}} \cdots b_{i_2}}_{z} b_{i_1} \;\; a_{i_0} b_{i_0} \;\; a_{i_1} \underbrace{a_{i_2} \cdots a_{i_{r-2}}}_{x} y a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad b_{j_r} y \underbrace{b_{j_{r-2}} \cdots b_{j_2}}_{z} b_{j_1} \;\; a_{j_0} b_{j_0} \;\; a_{j_1} \underbrace{a_{j_2} \cdots a_{j_{r-2}}}_{x} z a_{j_r}$$

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad b_{j_1} b_{j_r} y \underbrace{b_{j_{r-2}} \cdots b_{j_2}}_{z} \;\; a_{j_0} b_{j_0} \;\; a_{j_1} \underbrace{a_{j_2} \cdots a_{j_{r-2}}}_{x} z a_{j_r}$$

Случай 5: $(s_1, s_2) = (r, r)$. В слове v переименуем буквы y и x.

$$v_2L_1u_1 \quad : \quad z b_{i_{r-1}} \underbrace{b_{i_{r-2}} \cdots b_{i_2}}_{y} b_{i_1} \;\; a_{i_0} b_{i_0} \;\; a_{i_1} \underbrace{a_{i_2} \cdots a_{i_{r-2}}}_{x} a_{i_{r-1}} y \; [a_{i_1}]$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad [b_{j_1}] \; x b_{j_{r-1}} \underbrace{b_{j_{r-2}} \cdots b_{j_2}}_{y} b_{j_1} \;\; a_{j_0} b_{j_0} \;\; a_{j_1} \underbrace{a_{j_2} \cdots a_{j_{r-2}}}_{x} a_{j_{r-1}} z$$

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad b_{j_1} x b_{j_{r-1}} \underbrace{b_{j_{r-2}} \cdots b_{j_2}}_{y} \;\; a_{j_0} b_{j_0} \;\; a_{j_1} \underbrace{a_{j_2} \cdots a_{j_{r-2}}}_{x} a_{j_{r-1}} z$$

Случай 6: $(s_1, s_2) = (1, 2)$.

$$v_2L_1u_1 \quad : \quad b_{i_r} \underbrace{b_{i_{r-1}} \quad \cdots \quad b_{i_3}}_{x} z b_{i_1} \;\; a_{i_0} b_{i_0} \;\; y a_{i_2} \cdots \underbrace{a_{i_{r-3}} a_{i_{r-2}} a_{i_{r-1}}}_{x} a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad b_{j_r} \underbrace{b_{j_{r-1}} b_{j_{r-2}} b_{j_{r-3}}}_{x} \cdots b_{j_2} y \;\; a_{j_0} b_{j_0} \;\; a_{j_1} z \underbrace{a_{j_3} \quad \cdots \quad a_{j_{r-1}}}_{x} a_{j_r}$$

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad y b_{j_r} \underbrace{b_{j_{r-1}} b_{j_{r-2}} b_{j_{r-3}}}_{x} \cdots b_{j_2} \;\; a_{j_0} b_{j_0} \;\; a_{j_1} z \underbrace{a_{j_3} \quad \cdots \quad a_{j_{r-1}}}_{x} a_{j_r}$$

Случай 7: $(s_1, s_2) = (2, 1)$ симметричен случаю 6, т.е., если в случае 7 вместо слов u и v взять их реверсы, то получится в точности случай 6 (такое преобразование корректно, так как символы $u[r{+}1]$ и $v[r{+}1]$ в случае 6 не участвуют, что делает возможным разворачивание слов u и v без изменения описанной конструкции). При этом, если n нечетно, то удаляем первую букву слова $\overleftarrow{v}$ (это b_{i_1}) и считаем, что $\overleftarrow{v}[r{+}1] = \overleftarrow{v}[1]$.

Случай 8: $(s_1, s_2) \in \{1\} \times \{3, \dots, r{-}2\}$. В слове v переименуем буквы z и x.

$$v_2L_1u_1 \quad : \quad b_{i_r} b_{i_{r-1}} \underbrace{\overbrace{b_{i_{r-2}} \cdots b_{i_3}}^{x} b_{i_2} b_{i_1}}_{z} \;\; a_{i_0} b_{i_0} \;\; y a_{i_2} \cdots \underbrace{a_{i_{r-3}} a_{i_{r-2}} a_{i_{r-1}}}_{x} a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad b_{j_r} \underbrace{b_{j_{r-1}} b_{j_{r-2}} b_{j_{r-3}}}_{z} \cdots b_{j_2} y \;\; a_{j_0} b_{j_0} \;\; \underbrace{a_{j_1} a_{j_2} \overbrace{a_{j_3} \cdots a_{j_{r-2}}}^{z}}_{x} a_{j_{r-1}} a_{j_r}$$

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad y b_{j_r} \underbrace{b_{j_{r-1}} b_{j_{r-2}} b_{j_{r-3}}}_{z} \cdots b_{j_2} \;\; a_{j_0} b_{j_0} \;\; \underbrace{a_{j_1} a_{j_2} \overbrace{a_{j_3} \cdots a_{j_{r-2}}}^{z}}_{x} a_{j_{r-1}} a_{j_r}$$

Случай 9: $(s_1, s_2) \in \{3, \dots, r{-}2\} \times \{1\}$. Если n четно, то случай 9 симметричен случаю 8. Если n нечетно, то выберем новую позицию для x и переименуем буквы y и x в слове v.

$$v_2L_1u_1\ (o)\ :\quad b_{i_r}\,\underbrace{b_{i_{r-1}}b_{i_{r-2}}b_{i_{r-3}}}_{y}\cdots b_{i_2}z\ \ a_{i_0}b_{i_0}\ \ a_{i_1}a_{i_2}\underbrace{\overbrace{a_{i_3}\cdots a_{i_{r-2}}}^{y}a_{i_{r-1}}a_{i_r}}_{x}\ [a_{i_1}]$$

$$\overleftarrow{u_2L_2v_1}\ (o)\ :\quad b_{j_1}\,\underbrace{b_{j_r}b_{j_{r-1}}\overbrace{b_{j_{r-2}}\cdots b_{j_3}}^{x}}_{y}b_{j_2}\ \ a_{j_0}b_{j_0}\ \ za_{j_2}\cdots\underbrace{a_{j_{r-3}}a_{j_{r-2}}a_{j_{r-1}}}_{x}a_{j_r}$$

Случай 10: $(s_1, s_2) = (1, r{-}1)$. В слове v переименуем буквы z и x.

$$v_2L_1u_1\quad :\quad b_{i_r}x\,\underbrace{b_{i_{r-2}}\cdots b_{i_2}}_{z}b_{i_1}\ \ a_{i_0}b_{i_0}\ \ ya_{i_2}\underbrace{a_{i_3}\cdots a_{i_{r-1}}}_{x}a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e)\ :\quad b_{j_r}\,\underbrace{b_{j_{r-1}}\cdots b_{j_3}}_{z}b_{j_2}y\ \ a_{j_0}b_{j_0}\ \ a_{j_1}\underbrace{a_{j_2}\cdots a_{j_{r-2}}}_{x}za_{j_r}$$

$$\overleftarrow{u_2L_2v_1}\ (o)\ :\quad yb_{j_r}\,\underbrace{b_{j_{r-1}}\cdots b_{j_3}}_{z}b_{j_2}\ \ a_{j_0}b_{j_0}\ \ a_{j_1}\underbrace{a_{j_2}\cdots a_{j_{r-2}}}_{x}za_{j_r}$$

Случай 11: $(s_1, s_2) = (r{-}1, 1)$ симметричен случаю 10.

Случай 12: $(s_1, s_2) = (1, r)$.

$$v_2L_1u_1\ (e)\ :\quad zb_{i_{r-1}}\,\underbrace{b_{i_{r-2}}\ \cdots\ b_{i_2}}_{x}b_{i_1}\ \ a_{i_0}b_{i_0}\ \ ya_{i_2}\cdots\underbrace{a_{i_{r-3}}a_{i_{r-2}}a_{i_{r-1}}}_{x}a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e)\ :\quad b_{j_r}\,\underbrace{b_{j_{r-1}}b_{j_{r-2}}b_{j_{r-3}}}_{x}\cdots b_{j_2}y\ \ a_{j_0}b_{j_0}\ \ a_{j_1}\underbrace{a_{j_2}\ \cdots\ a_{j_{r-2}}}_{x}a_{j_{r-1}}z$$

Здесь $\bar{x} \in A\diagdown\{x\}$ и $\bar{z} \in B\diagdown\{z\}$. В слове v переименуем буквы z и x.

$$v_2L_1u_1\ (o)\ :\quad xb_{i_{r-1}}\,\underbrace{b_{i_{r-2}}\ \cdots\ b_{i_1}}_{z}\ \ a_{i_0}b_{i_0}\ \ ya_{i_2}\ \cdots\ \underbrace{a_{i_{r-2}}a_{i_{r-1}}}_{x}a_{i_r}\ [yz\bar{x}]$$

$$\overleftarrow{u_2L_2v_1}\ (o)\ :\quad [\bar{z}x]\ yb_{j_r}\,\underbrace{b_{j_{r-1}}b_{j_{r-2}}}_{z}\ \cdots\ b_{j_2}\ \ a_{j_0}b_{j_0}\ \ \underbrace{a_{j_1}\ \cdots\ a_{j_{r-2}}}_{x}a_{j_{r-1}}z$$

Случай 13: $(s_1, s_2) = (r, 1)$ симметричен случаю 12 (если n нечетно, то переименуем буквы z и x в слове v.

Случай 14: $(s_1, s_2) \in \{2\} \times \{3, \ldots, r{-}2\}$. В слове v переименуем буквы z и x.

$$v_2L_1u_1\quad :\quad \underbrace{b_{i_r}b_{i_{r-1}}\overbrace{b_{i_{r-2}}\cdots b_{i_3}}^{x}b_{i_2}b_{i_1}}_{z}\ \ a_{i_0}b_{i_0}\ \ a_{i_1}ya_{i_3}\ \ \cdots\ \ xa_{i_{r-1}}a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e)\ :\quad b_{j_r}b_{j_{r-1}}z\ \ \cdots\ \ b_{j_3}yb_{j_1}\ \ a_{j_0}b_{j_0}\ \ \underbrace{a_{j_1}a_{j_2}\overbrace{a_{j_3}\cdots a_{j_{r-2}}}^{z}a_{j_{r-1}}a_{j_r}}_{x}$$

$$\overleftarrow{u_2L_2v_1}\ (o)\ :\quad b_{j_1}b_{j_r}b_{j_{r-1}}z\ \ \cdots\ \ b_{j_3}y\ \ a_{j_0}b_{j_0}\ \ \underbrace{a_{j_1}a_{j_2}\overbrace{a_{j_3}\cdots a_{j_{r-2}}}^{z}a_{j_{r-1}}a_{j_r}}_{x}$$

Случай 15: $(s_1, s_2) \in \{3, \ldots, r{-}2\} \times \{2\}$. При четном n случай 15 симметричен случаю 14. если n нечетно, то переименуем буквы x и y в слове v. Чтобы избежать повтор $xya_{j_0}\cdots xya_{j_0}$, выберем новую позицию для y в слове v.

$$v_2L_1u_1\ (o)\ :\quad b_{i_r}b_{i_{r-1}}\underbrace{b_{i_{r-2}}b_{i_{r-3}}}_{y}\ \cdots\ zb_{i_1}\ \ a_{i_0}b_{i_0}\ \ a_{i_1}\underbrace{a_{i_2}\overbrace{a_{i_3}\cdots a_{i_{r-2}}}^{y}a_{i_{r-1}}a_{i_r}}_{x}$$

$$\overleftarrow{u_2L_2v_1}\ (o)\ :\quad \underbrace{b_{j_1}b_{j_r}b_{j_{r-1}}\overbrace{b_{j_{r-2}}\cdots b_{j_3}}^{x}}_{y}b_{j_2}\ \ a_{j_0}b_{j_0}\ \ a_{j_1}z\cdots\underbrace{a_{j_{r-3}}a_{j_{r-2}}}_{x}a_{j_{r-1}}a_{j_r}$$

Случай 16: $(s_1, s_2) = (2, r-1)$. В слове v переименуем буквы z и x.

$$v_2L_1u_1 \quad : \quad b_{i_r} x \underbrace{b_{i_{r-2}} \cdots b_{i_2} b_{i_1}}_{z} a_{i_0} b_{i_0} a_{i_1} y \cdots \underbrace{a_{i_{r-3}} a_{i_{r-2}}}_{x} a_{i_{r-1}} a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad b_{j_r} b_{j_{r-1}} \underbrace{b_{j_{r-2}} b_{j_{r-3}}}_{z} \cdots y b_{j_1} a_{j_0} b_{j_0} \underbrace{a_{j_1} a_{j_2} \cdots a_{j_{r-2}}}_{x} z a_{j_r}$$

Чтобы избежать повтор $zya_{j_0} \cdots zya_{j_0}$, выберем новую позицию для z в слове v.

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad b_{j_1} b_{j_r} \underbrace{b_{j_{r-1}} b_{j_{r-2}}}_{z} \cdots b_{j_3} y a_{j_0} b_{j_0} \underbrace{a_{j_1} a_{j_2} \cdots a_{j_{r-2}}}_{x} z a_{j_r}$$

Случай 17: $(s_1, s_2) = (r-1, 2)$ симметричен случаю 16.

Случай 18: $(s_1, s_2) = (2, r)$. В слове v переименуем буквы z и y.

$$v_2L_1u_1 \quad : \quad y b_{i_{r-1}} \underbrace{b_{i_{r-2}} \cdots b_{i_2}}_{x} b_{i_1} a_{i_0} b_{i_0} a_{i_1} y a_{i_3} \cdots \underbrace{a_{i_{r-2}} a_{i_{r-1}} a_{i_r}}_{x} \ [a_{i_1}]$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : [b_{j_1}] \underbrace{b_{j_r} b_{j_{r-1}} b_{j_{r-2}}}_{x} \cdots b_{j_3} z b_{j_1} a_{j_0} b_{j_0} a_{j_1} \underbrace{a_{j_2} \cdots a_{j_{r-2}}}_{x} a_{j_{r-1}} z$$

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad b_{j_1} \underbrace{b_{j_r} b_{j_{r-1}} b_{j_{r-2}}}_{x} \cdots b_{j_3} z a_{j_0} b_{j_0} a_{j_1} \underbrace{a_{j_2} \cdots a_{j_{r-2}}}_{x} a_{j_{r-1}} z$$

Случай 19: $(s_1, s_2) = (r, 2)$. В слове v переименуем буквы y и x.

$$v_2L_1u_1 \quad : \quad b_{i_r} \underbrace{b_{i_{r-1}} \cdots b_{i_3}}_{y} z b_{i_1} a_{i_0} b_{i_0} a_{i_1} \underbrace{a_{i_2} \cdots a_{i_{r-2}}}_{x} a_{i_{r-1}} y \ [a_{i_1}]$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad [b_{j_1}] x b_{j_{r-1}} \underbrace{b_{j_{r-2}} \cdots b_{j_2}}_{y} b_{j_1} a_{j_0} b_{j_0} a_{j_1} z \underbrace{a_{j_3} \cdots a_{j_{r-1}}}_{x} a_{j_r}$$

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad b_{j_1} x b_{j_{r-1}} \underbrace{b_{j_{r-2}} \cdots b_{j_2}}_{y} a_{j_0} b_{j_0} a_{j_1} z \underbrace{a_{j_3} \cdots a_{j_{r-1}}}_{x} a_{j_r}$$

Случай 20: $(s_1, s_2) \in \{3, \ldots, r-2\} \times \{r-1\}$. В слове v переименуем буквы z и x.

$$v_2L_1u_1 \quad : \quad b_{i_r} x \underbrace{b_{i_{r-2}} \cdots b_{i_1}}_{z} a_{i_0} b_{i_0} a_{i_1} \underbrace{a_{i_2} \overbrace{a_{i_3} \cdots a_{i_{r-2}}}^{y}}_{x} a_{i_{r-1}} a_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad b_{j_r} b_{j_{r-1}} \underbrace{\overbrace{b_{j_{r-2}} \cdots b_{j_3}}^{y} b_{j_2}}_{z} b_{j_1} a_{j_0} b_{j_0} \underbrace{a_{j_1} \cdots a_{j_{r-2}}}_{x} z a_{j_r}$$

Чтобы избежать повтор $yza_{j_0} \cdots yza_{j_0}$, выберем новую позицию для z в слове v.

$$\overleftarrow{u_2L_2v_1}\ (o) \quad : \quad b_{j_1} b_{j_r} \underbrace{b_{j_{r-1}} \overbrace{b_{j_{r-2}} \cdots b_{j_3}}^{y}}_{z} b_{j_2} a_{j_0} b_{j_0} \underbrace{a_{j_1} \cdots a_{j_{r-2}}}_{x} z a_{j_r}$$

Случай 21: $(s_1, s_2) \in \{r-1\} \times \{3, \ldots, r-2\}$ симметричен случаю 20.

Случай 22: $(s_1, s_2) \in \{3, \ldots, r-2\} \times \{r\}$. В слове v переименуем буквы y и x.

$$v_2L_1u_1\ (e) \quad : \quad z b_{i_{r-1}} y \cdots b_{i_1} a_{i_0} b_{i_0} \underbrace{a_{i_1} a_{i_2} \overbrace{a_{i_3} \cdots a_{i_{r-2}}}^{y} a_{i_{r-1}} a_{i_r}}_{x}$$

$$\overleftarrow{u_2L_2v_1}\ (e) \quad : \quad \underbrace{b_{j_r} b_{j_{r-1}} \overbrace{b_{j_{r-2}} \cdots b_{j_3}}^{x} b_{j_2} b_{j_1}}_{y} a_{j_0} b_{j_0} a_{j_1} \cdots x a_{j_{r-1}} z$$

Чтобы избежать повтор $xya_{j_0} \cdots xya_{j_0}$, выберем новую позицию для y в слове v.

$$v_2L_1u_1\ (o)\ :\ zb_{i_{r-1}}\underbrace{b_{i_{r-2}}b_{i_{r-3}}}_{y}\ \cdots\ b_{i_1}\ a_{i_0}b_{i_0}\ a_{i_1}\underbrace{a_{i_2}\overbrace{a_{i_3}\cdots a_{i_{r-2}}}^{y}a_{i_{r-1}}a_{i_r}}_{x}$$

$$\overleftarrow{u_2L_2v_1}\ (o)\ :\ \underbrace{b_{j_1}b_{j_r}b_{j_{r-1}}\overbrace{b_{j_{r-2}}\cdots b_{j_3}}^{x}}_{y}b_{j_2}\ a_{j_0}b_{j_0}\ a_{j_1}\ \cdots\ \underbrace{a_{j_{r-3}}a_{j_{r-2}}}_{x}a_{j_{r-1}}z$$

Случай 23: $(s_1,s_2)\in\{r\}\times\{3,\ldots,r-2\}$ симметричен случаю 22.

Случай 24: $(s_1,s_2)=(r-1,r)$. В слове v переименуем буквы y и x.

$$v_2L_1u_1\ :\ zb_{i_{r-1}}\underbrace{b_{i_{r-2}}\cdots b_{i_2}}_{y}b_{i_1}\ a_{i_0}b_{i_0}\ a_{i_1}\underbrace{a_{i_2}\ \cdots\ a_{i_{r-2}}}_{x}ya_{i_r}$$

$$\overleftarrow{u_2L_2v_1}\ (e)\ :\ b_{j_r}x\underbrace{b_{j_{r-2}}\ \cdots\ b_{j_2}}_{y}b_{j_1}\ a_{j_0}b_{j_0}\ a_{j_1}\underbrace{a_{j_2}\cdots a_{j_{r-2}}}_{x}a_{j_{r-1}}z$$

$$\overleftarrow{u_2L_2v_1}\ (o)\ :\ b_{j_1}b_{j_r}x\underbrace{b_{j_{r-2}}\cdots b_{j_2}}_{y}\ a_{j_0}b_{j_0}\ a_{j_1}\underbrace{a_{j_2}\cdots a_{j_{r-2}}}_{x}a_{j_{r-1}}z$$

Случай 25: $(s_1,s_2)=(r,r-1)$ симметричен случаю 24 (если n нечетно, то переименуем $x\to y\to z\to x$ в слове v).

Теперь проверим, что в каждом случае построенное слово (w) является $((r+1)/r)^+$ -свободным. В силу наблюдения 2.2, которое справедливо и для нечетных $k\geq 11$, только с экспонентой $(r+1)/r$, исключим из рассмотрения подслова вида ptp, где оба p являются подсловами u или v. Проверим остальные подслова. В силу конструкции каждое из слов u_1, u_2, v_1 и v_2 содержит ровно две буквы из x, y и z. Во всех случаях кроме 18, без ограничения общности можно предположить, что x появляется как в v_2, так и в u_1, y только в u_1, а z только в v_2 (аналогично разбирается пара (u_2,v_1)). Ниже перечислены все возможные повторы с их экспонентами (повторы, имеющие другие экспоненты в случае 18, даны в квадратных скобках).

$$x\underbrace{\cdots}_{\geq r-1}x \qquad\Rightarrow\ \exp(ptp)\leq(r+1)/r;$$

$$y\underbrace{\cdots v_2L_1\cdots}_{\geq r+2}y\quad[y\underbrace{\cdots}_{\geq r+1}y]\qquad\Rightarrow\ \exp(ptp)\leq(r+3)/(r+2);$$

$$z\underbrace{\cdots L_1u_1\cdots}_{\geq r+2}z\quad[z\underbrace{\cdots v_2L_1u_1\cdots}_{\geq 2r+2}z]\qquad\Rightarrow\ \exp(ptp)\leq(r+4)/(r+3);$$

$$zx\underbrace{\cdots}_{\geq r-1}x\underbrace{\cdots}_{\geq r-2}zx\quad[zx\underbrace{\cdots}_{\geq r-1}x\underbrace{\cdots}_{\geq r-1}x\underbrace{\cdots}_{\geq r-2}zx]\qquad\Rightarrow\ \exp(ptp)\leq(2r+2)/(2r);$$

$$yx\underbrace{\cdots}_{\geq r-1}x\underbrace{\cdots}_{\geq r-2}yx\quad[yx\underbrace{\cdots}_{\geq r-1}x\underbrace{\cdots}_{\geq r-1}x\underbrace{\cdots}_{\geq r-2}yx]\qquad\Rightarrow\ \exp(ptp)\leq(2r+2)/(2r);$$

$$yz\underbrace{\cdots v_2'L_1u_1'\cdots}_{\geq 2r}yz\quad[zy\underbrace{\cdots}_{\geq r+1}y\underbrace{\cdots}_{\geq r-2}zy]\qquad\Rightarrow\ \exp(ptp)\leq(2r+4)/(2r+2),$$

здесь u_1' – это либо u_1, либо u_1 без буквы a_{i_r}; то же самое для v_2';

$$xyz\underbrace{\cdots}_{\geq r-3}x\underbrace{\cdots}_{\geq r-1}x\underbrace{\cdots}_{\geq r-1}xyz\qquad\Rightarrow\ \exp(ptp)\leq(3r+3)/(3r)$$

(в случае 12(o), используем z вместо x).

Если $b_{i_1} = z$ (аналогично симметричные случаи), то возможно появление следующего повтора:

$$z a_{i_0} \underbrace{\cdots}_{\geq r-1} a_{i_{r-1}} [a_{i_r}] \underbrace{\cdots}_{\geq r-2} z a_{i_0} \quad \Rightarrow \quad \exp(ptp) \leq (2r{+}2)/(2r).$$

Следующие повторы всегда разрешены в силу выбора a_{i_0} (аналогично a_{j_0}, b_{i_0}, b_{j_0}):

$$\begin{aligned} & a_{i_0} \underbrace{\cdots}_{\geq r-1} a_{i_{r-1}} [a_{i_r}] && \Rightarrow \quad \exp(ptp) \leq (r{+}1)/r; \\ & a_{i_0} b_{i_0} \underbrace{\cdots}_{\geq r-2} a_{i_{r-1}} [a_{i_r}] \underbrace{\cdots}_{\geq r-1} a_{j_1} b_{j_0} && \Rightarrow \quad \exp(ptp) \leq (2r{+}2)/2r. \end{aligned}$$

Напомним, что $r{=}\lceil k/2 \rceil \geq 6$. Если в случае I появится повтор $b_{j_0} a_{j_0} b_{j_1} \cdots b_{i_1} a_{i_0} b_{i_0}$ [или симметричный $a_{i_0} b_{i_0} a_{i_1} \cdots a_{j_1} b_{j_0} a_{j_0}$], тогда $a_{j_r} = a_{j_0} = a_{i_0} \in \{a_{i_{r-1}}, a_{i_r}\}$ [соответственно, $b_{i_r} \in \{b_{j_{r-1}}, b_{j_r}\}$, если n четно, $b_{i_r} \in \{b_{j_1}, b_{j_r}\}$, если n нечетно]. Тогда есть по крайней мере $r{-}1$ буква между $a_{i_{r-1}}$ и a_{j_r}. Откуда получаем $u_1 \underbrace{a_{i_1} z \cdots \overbrace{a_{j_{r+2}} a_{j_{r+1}}}^{z}}_{\geq r-2} u_2$. Есть по крайней мере $r{-}1$ буква между двумя вхождениями z, следовательно, $|v| \geq |u|{-}1 \geq 3r{+}1$ [соответственно, $|u| \geq |v| \geq 3r{+}1$] и $b_{j_0} a_{j_0} b_{j_1} \underbrace{\cdots}_{\geq |v|-2} b_{i_1} a_{i_0} b_{i_0}$ [соответственно, $a_{i_0} b_{i_0} a_{i_1} \underbrace{\cdots}_{\geq |u|-2} a_{j_1} b_{j_0} a_{j_0}$]. Таким образом, $\exp(ptp) \leq (3r{+}5)/(3r{+}2) < (r{+}1)/r$. Аналогично получаем $u_1 \underbrace{y \cdots \overbrace{a_{j_{r+2}} a_{j_{r+1}}}^{y}}_{\geq r-1} u_2$ в случае 6(o), $a_{i_1} \cdots y \underbrace{\cdots \overbrace{a_{j_{r+2}} a_{j_{r+1}}}^{y}}_{\geq r-3} u_2$ в случаях 19 и 23, а также $u_1 \underbrace{a_{i_1} z \cdots}_{\geq r-3} \overbrace{a_{j_r} a_{j_{r-1}}}^{z} \cdots a_{j_1}$ в случаях 4, 5, 10(o), 12(o), 18, 22, 24 и 25. Во всех этих случаях $|v| \geq |u|{-}1 \geq 3r{-}1$ и $b_{j_0} a_{j_0} b_{j_1} \underbrace{\cdots}_{\geq |v|-2} b_{i_1} a_{i_0} b_{i_0}$ [соответственно, $|u| \geq |v| \geq 3r{-}1$ и $a_{i_0} b_{i_0} a_{i_1} \underbrace{\cdots}_{\geq |u|-2} a_{j_1} b_{j_0} a_{j_0}$]. Таким образом, $\exp(ptp) \leq (3r{+}3)/(3r) = (r{+}1)/r$. А значит существование $((r{+}1)/r)^+$-свободных циклических слов длины n доказано для любого $n \geq 2k + 6$.

Для завершения доказательства осталось рассмотреть случай $k{+}1 \leq n \leq 2k{+}5$. Если $2k{+}2 \leq n \leq 2k{+}5$, то циклическое слово (w) длины n будем искать в виде $L_1 u L_2 v$, где $u \in A^*$ и $v \in B^*$ – $((r{+}1)/r)^+$-свободные слова, удовлетворяющие условиям $|u| = \lceil (n{-}4)/2 \rceil, u[1] = u[r{+}1], |v| = \lfloor (n{-}4)/2 \rfloor$ и $v[1] = v[r{+}1]$. В отличие от случая $n \geq 2k{+}6$ здесь в словах u и v их префиксы и суффиксы длины r будут перекрываться, т.е. $u_1 \cap u_2 \neq \varnothing$ и $v_1 \cap v_2 \neq \varnothing$. В этом случае $u[r{+}2] = \Sigma(u_2) \setminus \Sigma(u_1)$ и $\{\overleftarrow{u}[r{+}1], \overleftarrow{u}[r{+}2]\} \cap (\Sigma(u_1) \setminus \Sigma(u_2)) \neq \varnothing$ (аналогично для слова v). Переименуем буквы в слове (w) как показано на рис. 2.24 (верхняя половина). Проверим, что построенное слово (w) является $((r{+}1)/r)^+$-свободным.

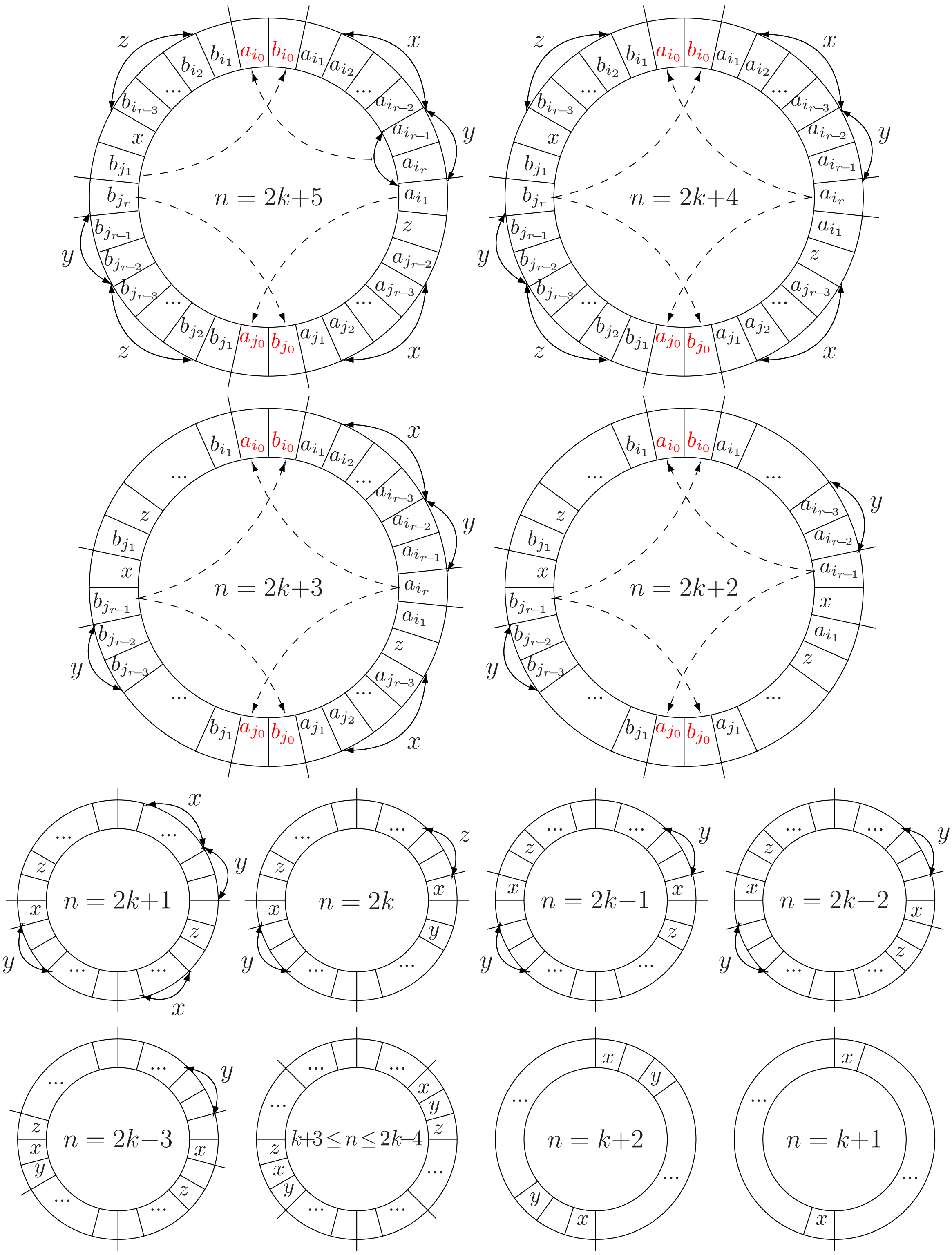


Рис. 2.24. *Циклическое слово* (w) *длины* n, *где* $k+1 \leq n \leq 2k+5$.

n = 2k+5

$$
\begin{array}{ll}
x\underbrace{\cdots}_{\geq r}x & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
y\underbrace{\cdots u_2L_2[v_2L_1]\cdots}_{\geq 2r-1}y & \Rightarrow\ \exp(ptp) \leq (2r{+}1)/2r;\\
z\underbrace{\cdots}_{\geq r+1}z & \Rightarrow\ \exp(ptp) \leq (r{+}3)/(r{+}2);\\
a_{i_0}\underbrace{\cdots}_{\geq r-1}a_{i_{r-1}}[a_{i_r}] & \Rightarrow\ \exp(ptp) \leq (r{+}1)/r;\\
a_{i_1}[=a_{i_{r+1}}=a_{j_r}]\underbrace{\cdots}_{\geq r}a_{j_0} & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
b_{j_1}[=b_{j_{r+1}}=b_{i_{r-1}}]\underbrace{\cdots}_{\geq r-1}b_{i_0} & \Rightarrow\ \exp(ptp) \leq (r{+}1)/r;\\
b_{j_0}\underbrace{\cdots}_{\geq r}b_{j_r} & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1).
\end{array}
$$

Повторов вида $xy\cdots xy$, $cx\cdots cx$ (где $c \in A\setminus B$ или $c \in B\setminus A$) или $xyz\cdots xyz$ в данном случае не возникает (также как и во всех остальных). Повторы вида $a_{i_0}b_{i_0}\cdots a_{j_1}b_{j_0}[a_{j_0}b_{j_1}]$ или $b_{i_1}a_{i_0}b_{i_0}\cdots b_{j_0}a_{j_0}b_{j_1}$ невозможны, так как $a_{i_0} \neq a_{j_0}$ и $b_{i_0} \neq b_{j_0}$.

n = 2k+4

$$
\begin{array}{ll}
x\underbrace{\cdots}_{\geq r}x & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
y\underbrace{\cdots u_2L_2[v_2L_1]\cdots}_{\geq 2r-1}y & \Rightarrow\ \exp(ptp) \leq (2r{+}1)/2r;\\
z\underbrace{\cdots}_{\geq r}z & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
a_{i_0}\underbrace{\cdots}_{\geq r}a_{i_r} & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
a_{i_r}[=a_{j_r}]\underbrace{\cdots}_{\geq r}a_{j_0} & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
b_{j_r}[=b_{i_r}]\underbrace{\cdots}_{\geq r}b_{i_0} & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
b_{j_0}\underbrace{\cdots}_{\geq r}b_{j_r} & \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1).
\end{array}
$$

Повторы вида $a_{i_0}b_{i_0}\cdots a_{j_0}b_{j_1}$ или $a_{i_0}b_{i_0}\cdots a_{j_1}b_{j_0}$ невозможны, так как $a_{i_0} = a_{j_0}$, $b_{i_0} = b_{j_0}$, но $b_{i_0}[= b_{j_r}] \neq b_{j_1}$ и $a_{i_0}[= a_{j_r}] \neq a_{j_1}$ (аналогично в остальных двух случаях).

$\mathbf{n = 2k{+}3}$

$$\begin{aligned}
&x\underbrace{\cdots}_{\geq r+1}x && \Rightarrow\ \exp(ptp) \leq (r{+}3)/(r{+}2);\\
&y\underbrace{\cdots u_2L_2[v_2L_1]\cdots}_{\geq 2r-2}y && \Rightarrow\ \exp(ptp) \leq 2r/(2r{-}1);\\
&z\underbrace{\cdots L_1u_1[L_2v_1]}_{\geq 2r-1}z && \Rightarrow\ \exp(ptp) \leq (2r{+}1)/2r;\\
&a_{i_0}\underbrace{\cdots}_{\geq r}a_{i_r} && \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
&a_{i_r}[= a_{j_r}]\underbrace{\cdots}_{\geq r}a_{j_0} && \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
&b_{j_{r-1}}[= b_{i_r}]\underbrace{\cdots}_{\geq r}b_{i_0} && \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
&b_{j_0}\underbrace{\cdots}_{\geq r-1}b_{j_{r-1}} && \Rightarrow\ \exp(ptp) \leq (r{+}1)/r.
\end{aligned}$$

$\mathbf{n = 2k{+}2}$

$$\begin{aligned}
&x\underbrace{\cdots}_{\geq 2r-1}x && \Rightarrow\ \exp(ptp) \leq (2r{+}1)/2r;\\
&y\underbrace{\cdots u_2L_2[v_2L_1]\cdots}_{\geq 2r-2}y && \Rightarrow\ \exp(ptp) \leq 2r/(2r{-}1);\\
&z\underbrace{\cdots}_{\geq 2r-1}z && \Rightarrow\ \exp(ptp) \leq (2r{+}1)/2r;\\
&a_{i_0}\underbrace{\cdots}_{\geq r-1}a_{i_{r-1}} && \Rightarrow\ \exp(ptp) \leq (r{+}1)/r;\\
&a_{i_{r-1}}[= a_{j_r}]\underbrace{\cdots}_{\geq r}a_{j_0} && \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
&b_{j_{r-1}}[= b_{i_r}]\underbrace{\cdots}_{\geq r}b_{i_0} && \Rightarrow\ \exp(ptp) \leq (r{+}2)/(r{+}1);\\
&b_{j_0}\underbrace{\cdots}_{\geq r-1}b_{j_{r-1}} && \Rightarrow\ \exp(ptp) \leq (r{+}1)/r.
\end{aligned}$$

Для $k{+}3 \leq n \leq 2k{+}1$ отбросим вставки L_1 и L_2 и будем искать слово (w) в виде uv, где слова u и v такие же, как в предыдущем случае. Чтобы построенное слово (w) было $((r{+}1)/r)^+$-свободным, переименуем буквы в словах u и v так, чтобы буквы x, y, z занимали указанные на рис. 2.24 (нижняя половина) позиции. Отметим, что на рисунке показаны все вхождения букв x, y, z в слове (w). Необходимо проверить только повторы вида ptp, где $p \in \{x, y, z\}$ (повторы с $|p| > 1$ не появляются в слове (w) благодаря сделанным перестановкам букв), причем первое и второе вхождения слова p принадлежат разным половинкам слова (w) (т.е. одно лежит в слове u, другое – в слове v).

n = 2k+1

$$x\underbrace{\cdots}_{\geq r}x \quad\Rightarrow\quad \exp(ptp) \leq (r{+}2)/(r{+}1);$$
$$y\underbrace{\cdots}_{\geq 2r-3}y \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}1)/(2r{-}2);$$
$$z\underbrace{\cdots}_{\geq 2r-2}z \quad\Rightarrow\quad \exp(ptp) \leq 2r/(2r{-}1).$$

n = 2k

$$x\underbrace{\cdots}_{=2r-2}x \quad\Rightarrow\quad \exp(ptp) = 2r/(2r{-}1);$$
$$y\underbrace{\cdots}_{\geq 2r-6}y \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}4)/(2r{-}5);$$
$$z\underbrace{\cdots}_{\geq 2r-6}z \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}4)/(2r{-}5).$$

n = 2k−1

$$x\underbrace{\cdots}_{\geq 2r-3}x \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}1)/(2r{-}2);$$
$$y\underbrace{\cdots}_{\geq 2r-4}y \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}2)/(2r{-}3);$$
$$z\underbrace{\cdots}_{\geq 2r-3}z \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}1)/(2r{-}2).$$

n = 2k−2

$$x\underbrace{\cdots}_{=2r-3}x \quad\Rightarrow\quad \exp(ptp) = (2r{-}1)/(2r{-}2);$$
$$y\underbrace{\cdots}_{\geq 2r-4}y \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}2)/(2r{-}3);$$
$$z\underbrace{\cdots}_{=2r-3}z \quad\Rightarrow\quad \exp(ptp) = (2r{-}1)/(2r{-}2).$$

n = 2k−3

$$x\underbrace{\cdots}_{\geq 2r-4}x \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}2)/(2r{-}3);$$
$$y\underbrace{\cdots}_{\geq 2r-5}y \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}3)/(2r{-}4);$$
$$z\underbrace{\cdots}_{\geq 2r-5}z \quad\Rightarrow\quad \exp(ptp) \leq (2r{-}3)/(2r{-}4).$$

k+3 ≤ n ≤ 2k−4

$$x\underbrace{\cdots}_{\geq r-1}x \quad\Rightarrow\quad \exp(ptp) \leq (r{+}1)/r;$$
$$y\underbrace{\cdots}_{\geq r-1}y \quad\Rightarrow\quad \exp(ptp) \leq (r{+}1)/r;$$
$$z\underbrace{\cdots}_{\geq r}z \quad\Rightarrow\quad \exp(ptp) \leq (r{+}2)/(r{+}1).$$

Для $n = k{+}2$ построим циклическое слово (w) следующим образом: положим

$w[1] = w[\lceil n/2 \rceil] = x$, $w[3] = w[\lceil n/2 \rceil + 2] = y$, а все остальные $k-2$ буквы в слове (w) будут различны. Тогда, учитывая, что $\lceil n/2 \rceil = \lceil (k+2)/2 \rceil = \lceil k/2 \rceil + 1 = r+1$, получаем:

$\mathbf{n = k+2}$

$$x \underbrace{\cdots}_{\geq \lceil n/2 \rceil - 2} x \;\Rightarrow\; \exp(ptp) \leq \lceil n/2 \rceil / (\lceil n/2 \rceil - 1) = (r+1)/r;$$
$$y \underbrace{\cdots}_{\geq \lceil n/2 \rceil - 2} y \;\Rightarrow\; \exp(ptp) \leq \lceil n/2 \rceil / (\lceil n/2 \rceil - 1) = (r+1)/r.$$

А для $n = k+1$ построим циклическое слово (w), в котором $w[1] = w[\lceil n/2 \rceil + 1] = x$, а все остальные буквы различны. Так как число $n = k+1$ четное, то $\lceil n/2 \rceil = \lfloor n/2 \rfloor = \lceil k/2 \rceil = r$. Получаем:

$\mathbf{n = k+1}$

$$x \underbrace{\cdots}_{\geq \lfloor n/2 \rfloor - 1} x \;\Rightarrow\; \exp(ptp) \leq (\lfloor n/2 \rfloor + 1) / \lfloor n/2 \rfloor = (r+1)/r.$$

Таким образом, для любого $n \in \mathbb{N}$ существует циклическое $\left(\frac{\lceil k/2 \rceil + 1}{\lceil k/2 \rceil}\right)^{+}$-свободное слово длины n над k-буквенным алфавитом, где $k \geq 11$ – нечетное натуральное число. □

Следствие 2.7. *Для любого нечетного* $k \geq 11$ $\mathrm{CRT}(k) \leq \frac{\lceil k/2 \rceil + 1}{\lceil k/2 \rceil}$.

2.3. Аналог гипотезы Дежан для циклических слов

Теорема 2.7. *Для любого* $k \geq 6$ $\mathrm{CRT}(k) = \frac{\lceil k/2 \rceil + 1}{\lceil k/2 \rceil}$.

Доказательство. Объединяя нижнюю (см. предложение 2.1) и верхние (см. теоремы 2.3, 2.4, 2.5 и 2.6) оценки для циклической границы повторяемости, получим, что для $k = 6$ и любого нечетного $k \geq 7$ $\mathrm{CRT}(k) = \frac{\lceil k/2 \rceil + 1}{\lceil k/2 \rceil}$. Что касается четных $k \geq 8$, то нижнюю оценку по-прежнему дает предложение 2.1, а, используя следующее неравенство, легко получить недостающие верхние оценки, совпадающие с нижними:

$$\mathrm{CRT}(k) \leq \mathrm{CRT}(k-1) \leq \frac{\lceil (k-1)/2 \rceil + 1}{\lceil (k-1)/2 \rceil} = \frac{\lceil k/2 \rceil + 1}{\lceil k/2 \rceil}.$$

□

Замечание 2.2. *Для любого* $n \leq 50$ *существует циклическое* $\left(\frac{\lceil k/2 \rceil + 1}{\lceil k/2 \rceil}\right)^{+}$*-свободное слово длины* n *над* k*-буквенным алфавитом, где* $4 \leq k \leq 5$. Действительно, данные слова, за исключением тривиальных случаев, когда $n \leq k$, представлены на рис. А.1–А.8 в приложении А.

Таким образом, можно предложить следующий аналог гипотезы Дежан для циклических слов:

Гипотеза 2.1. *Циклические β-свободные слова любой длины над алфавитом Σ_k существуют тогда и только тогда, когда $\beta > \mathrm{CRT}(k)$, где $\mathrm{CRT}(k)$ принимает следующие значения:*

k	2	3	4	5	6	$\cdots$	i	$\cdots$
$\mathrm{CRT}(k)$	$5/2$	2	$3/2$	$4/3$	$4/3$	$\cdots$	$(\lceil i/2\rceil+1)/\lceil i/2\rceil$	$\cdots$

Отметим, что в данной гипотезе не доказаны только два случая: $k = 4$ и $k = 5$, хотя в каждом из них известен интервал, в котором может находится значение циклической границы повторяемости: $3/2 \le \mathrm{CRT}(4) \le 7/4$ и $4/3 \le \mathrm{CRT}(5) \le 3/2$ (см. предложение 2.1 и теоремы 2.1 и 2.2). Гипотеза о том, что граница повторяемости совпадает с нижним концом этого интервала, опирается на замечание 2.2 и тот факт, что с увеличением длины циклического слова вероятность улучшить нижнюю оценку для CRT убывает. Кроме того, заметим, что все случаи кроме $k = 2$ и $k = 3$ удовлетворяют общей формуле.

Литература

[1] Аршон С. Е. Доказательство существования n-значных бесконечных асимметричных последовательностей // Матем. сборник. 1937. Том 2(44). С. 769–779.

[2] Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления. Москва : Физматлит, 2001. Том 1.

[3] Aberkane A., Currie J. D. There exist binary circular $5/2^+$ power free words of every length // Electronic J. Combinatorics. 2004. Vol. 11(1). Article R10.

[4] Aberkane A., Currie J. D. Attainable lengths for circular binary words avoiding k-powers // Bull. Belg. Math. Soc. Simon Stevin. 2005. Vol. 12(4). P. 525–534.

[5] Badkobeh G., Crochemore M. Finite-repetition threshold for infinite ternary words // Proc. 8th Internat. Conf. Words 2011 (WORDS 2011). Vol. 63 of EPTCS. 2011. P. 37–43.

[6] Badkobeh G., Crochemore M., Rao M. Finite-repetition threshold for large alphabets // Proc. 14th Mons Days of Theoretical Computer Science.

[7] Berstel J., Karhumäki J. Combinatorics on words: A tutorial // Bull. European Assoc. Theor. Comput. Sci. 2003. Vol. 79. P. 178–228.

[8] Blondel V. D., Cassaigne J., Jungers R. M. On the number of α-power-free binary words for $2 < \alpha \leq 7/3$ // Theoret. Comput. Sci. 2009. Vol. 410. P. 2823–2833.

[9] Brandenburg F.-J. Uniformly growing k-th power-free homomorphisms // Theoret. Comput. Sci. 1983. Vol. 23. P. 69–82.

[10] Capobianco S. Multidimensional cellular automata and generalization of Fekete's lemma // Discrete Math. & Theoret. Comput. Sci. 2008. Vol. 10(3).

[11] Carpi A. On Dejean's conjecture over large alphabets // Theoret. Comput. Sci. 2007. Vol. 385. P. 137–151.

[12] Cassaigne J. Counting overlap-free binary words // STACS 93, Proc. 10th Symp. Theoretical Aspects of Comp. Sci. / Ed. by P. Enjalbert, A. Finkel, K. W. Wagner. Springer-Verlag, 1993. Vol. 665 of LNCS. P. 216–225.

[13] Choffrut C., Karhumäki J. Combinatorics of words // Handbook of formal languages / Ed. by G. Rozenberg, A. Salomaa. Springer-Verlag, 1997. Vol. 1. P. 329–438.

[14] Crochemore M., Mignosi F., Restivo A. Automata and forbidden words // Inform. Process. Lett. 1998. Vol. 67. P. 111–117.

[15] Currie J. D., Rampersad N. Dejean's conjecture holds for $n \geq 27$ // RAIRO Inform. Théor. App. 2009. Vol. 43. P. 775–778.

[16] Currie J. D., Rampersad N. A proof of Dejean's conjecture // Math. Comp. 2011. Vol. 80. P. 1063–1070.

[17] Dejean F. Sur un théorème de Thue // J. Combin. Theory. Ser. A. 1972. Vol. 13. P. 90–99.

[18] Euwe M. Mengentheoretische Betrachtungen über das Schachspiel // Proc. Konin. Akad. Wetenschappen, Amsterdam. 1929. Vol. 32. P. 633–642.

[19] Fekete M. Über die Verteilung der Wurzeln bei gewisser algebraichen Gleichungen mit ganzzahlingen Koeffizienten // Math. Zeitschr. 1923. Vol. 17. P. 228–249.

[20] Fiorenzi F., Ochem P., Vaslet E. Bounds for the generalized repetition threshold // Theoret. Comput. Sci. 2011. Vol. 412. P. 2955–2963.

[21] Giammarresi D., Restivo A. Two-dimensional languages // Handbook of formal languages / Ed. by G. Rozenberg, A. Salomaa. New York : Springer-Verlag, 1997. Vol. 3. P. 215–267.

[22] Ilie L., Ochem P., Shallit J. A generalization of repetition threshold // Theoret. Comput. Sci. 2005. Vol. 345. P. 359–369.

[23] Karhumäki J., Shallit J. Polynomial versus exponential growth in repetition-free binary words // J. Combin. Theory. Ser. A. 2004. Vol. 104. P. 335–347.

[24] Keränen V. Abelian squares are avoidable on 4 letters // Proc. 19th Int'l Conf. on Automata, Languages, and Programming (ICALP) / Ed. by W. Kuich. Springer-Verlag, 1992. Vol. 623 of LNCS. P. 41–52.

[25] Kolpakov R., Rao M. On the number of Dejean words over alphabets of 5, 6, 7, 8, 9 and 10 letters // Theoret. Comput. Sci. 2011. Vol. 412. P. 6507–6516.

[26] Lindgren K., Moore C., Nordahl M. Complexity of two-dimensional patterns // J. Stat. Physics. 1998. Vol. 91. P. 909–951.

[27] Local squares, periodicity and finite automata / M. Huova, J. Karhumäki, A. Saarela, K. Saari // Rainbow of Computer Science / Ed. by C. Calude, G. Rozenberg, A. Salomaa. 2011. P. 90–101.

[28] Lothaire M. Combinatorics on Words. Addison-Wesley, 1983. Vol. 17 of Encyclopedia of Mathematics and Its Applications.

[29] Mercas R., Saarela A. 3-abelian cubes are avoidable on binary alphabets // Proc. 17th Developments in Language Theory (DLT 2013). Vol. 7907 of LNCS. Springer, 2013. P. 374–383.

[30] Mohammad-Noori M., Currie J. D. Dejean's conjecture and Sturmian words // European J. Combinatorics. 2007. Vol. 28. P. 876–890.

[31] Morse M. Recurrent geodesics on a surface of negative curvature // Trans. Amer. Math. Soc. 1921. Vol. 22. P. 84–100.

[32] Morse M., Hedlund G. A. Symbolic dynamics // Amer. J. Math. 1938. Vol. 60. P. 815–866.

[33] Moulin-Ollagnier J. Proof of Dejean's conjecture for alphabets with $5, 6, 7, 8, 9, 10$ and 11 letters // Theoret. Comput. Sci. 1992. Vol. 95. P. 187–205.

[34] Mousavi H., Shallit J. Repetition avoidance in circular factors // Proc. 17th Developments in Language Theory (DLT 2013). Vol. 7907 of LNCS. Springer, 2013. P. 384–395.

[35] Ochem P. A generator of morphisms for infinite words // RAIRO Inform. Théor. App. 2006. Vol. 40. P. 427–441.

[36] Pansiot J.-J. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots // Discrete Appl. Math. 1984. Vol. 7. P. 297–311.

[37] Prouhet E. Mémoire sur quelques relations entre les puissances des nombres // C. R. Acad. Sci. Paris. 1851. Vol. 33. P. 225.

[38] Rao M. Last cases of Dejean's conjecture // Theoret. Comput. Sci. 2011. Vol. 412. P. 3010–3018.

[39] Rauzy G. Suites à termes dans un alphabet fini // Seminaire de Théorie des Nombres de Bordeaux. 1982–1983. Vol. 12. P. 1–16.

[40] Restivo A., Salemi S. Overlap free words on two symbols // Automata on Infinite Words. Ecole de Printemps d'Informatique Theorique, Le Mont Dore, 1984 / Ed. by M. Nivat, D. Perrin. Springer-Verlag, 1985. Vol. 192 of LNCS. P. 198–206.

[41] Rumyantsev A. Upper bound for the generalized repetition threshold. 2010. arXiv:1009.4454.

[42] Samsonov A. V., Shur A. M. On Abelian repetition threshold // RAIRO Inform. Théor. App. 2012. Vol. 46. P. 147–163.

[43] Shallit J. Simultaneous avoidance of large squares and fractional powers in infinite binary words // Internat. J. Found. Comp. Sci. 2004. Vol. 15. P. 317–327.

[44] Shur A. M. Rational approximations of polynomial factorial languages // Internat. J. Found. Comp. Sci. 2007. Vol. 18. P. 655–665.

[45] Shur A. M. Combinatorial complexity of regular languages // Proc. 3rd International Computer Science Symposium in Russia. CSR 2008. Vol. 5010 of LNCS. Springer, 2008. P. 289–301.

[46] Shur A. M. Comparing complexity functions of a language and its extendable part // RAIRO Inform. Théor. App. 2008. Vol. 42. P. 647–655.

[47] Shur A. M. Growth rates of complexity of power-free languages // Theoret. Comput. Sci. 2010. Vol. 411. P. 3209–3223.

[48] Shur A. M. On the existence of minimal β-powers // Internat. J. Found. Comp. Sci. 2011. Vol. 22. P. 1683–1696.

[49] Shur A. M. Deciding context equivalence of binary overlap-free words in linear time // Semigroup Forum. 2012. Vol. 84. P. 447–471.

[50] Thue A. Über unendliche Zeichenreihen // Norske vid. Selsk. Skr. Mat. Nat. Kl. 1906. Vol. 7. P. 1–22.

[51] Thue A. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen // Norske vid. Selsk. Skr. Mat. Nat. Kl. 1912. Vol. 1. P. 1–67.

[52] Tunev I. N., Shur A. M. On two stronger versions of dejean's conjecture // Proc. 37th Symposium, Mathematical Foundations of Computer Science 2012. Vol. 7464 of LNCS. Springer, 2012. P. 800–812.

[53] van Lint J. H., Wilson R. M. A course in combinatorics. Cambridge University Press, 1992. P. I–XII, 1–530. ISBN: 978-0-521-41057-1.

Публикации автора по теме диссертации

[54] Горбунова И. А. Структура равномерных 7-повторов и их влияние на индекс роста граничных языков / Уральский федеральный университет имени первого Президента России Б. Н. Ельцина. Екатеринбург, 2013. 20 с. Библ.: 16 назв. русский. Деп. в ВИНИТИ РАН 05.03.2013 № 63-В2013.

[55] Gorbunova I. A. Repetition threshold for circular words // Proc. 2nd Russian Finnish Symposium on Discrete Mathematics. Vol. 17 of TUCS Lecture Notes. Turku, 2012. P. 76–78.

[56] Gorbunova I. A. Repetition threshold for circular words // Electronic J. Combinatorics. 2012. Vol. 19(4). Article P11.

[57] Gorbunova I. A., Shur A. M. On Pansiot words avoiding 3-repetitions // Proc. 8th Internat. Conf. Words 2011 (WORDS 2011). Vol. 63 of EPTCS. 2011. P. 138–146.

[58] Gorbunova I. A., Shur A. M. On Pansiot words avoiding 3-repetitions // Internat. J. Found. Comp. Sci. 2012. Vol. 23(8). P. 1583–1594.

[59] Shur A. M., Gorbunova I. A. On the growth rates of complexity of threshold languages // Proc. 12th Mons Days of Theoretical Computer Science. Mons : Univ. de Mons-Hainaut, 2008. P. 1–10.

[60] Shur A. M., Gorbunova I. A. On the growth rates of complexity of threshold languages // RAIRO Inform. Théor. App. 2010. Vol. 44. P. 175–192.

Приложение А

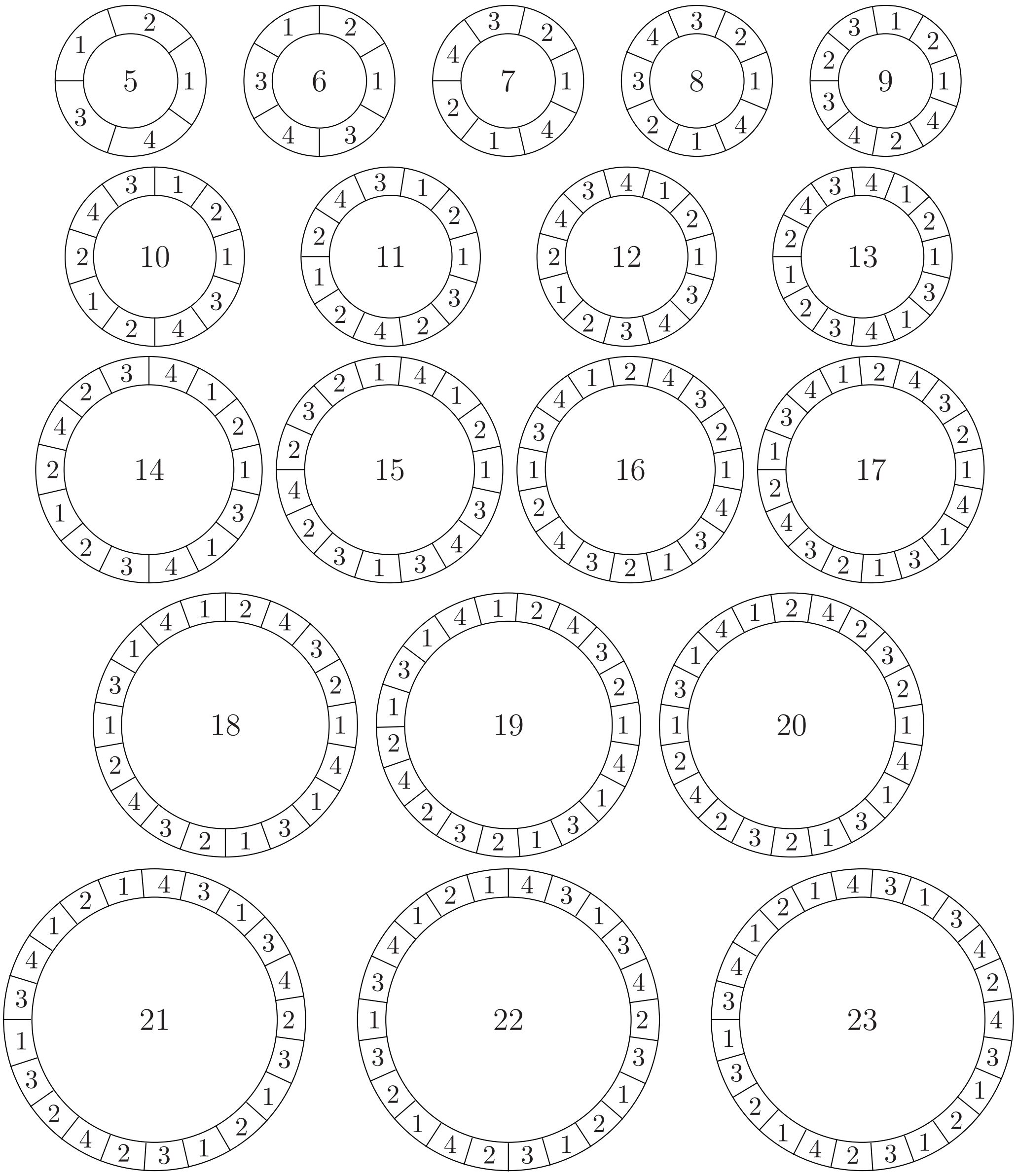


Рис. А.1. $(3/2)^+$-*свободное слово* (w) *над* 4-*буквенным алфавитом длины* n, *где* $5 \leq n \leq 23$.

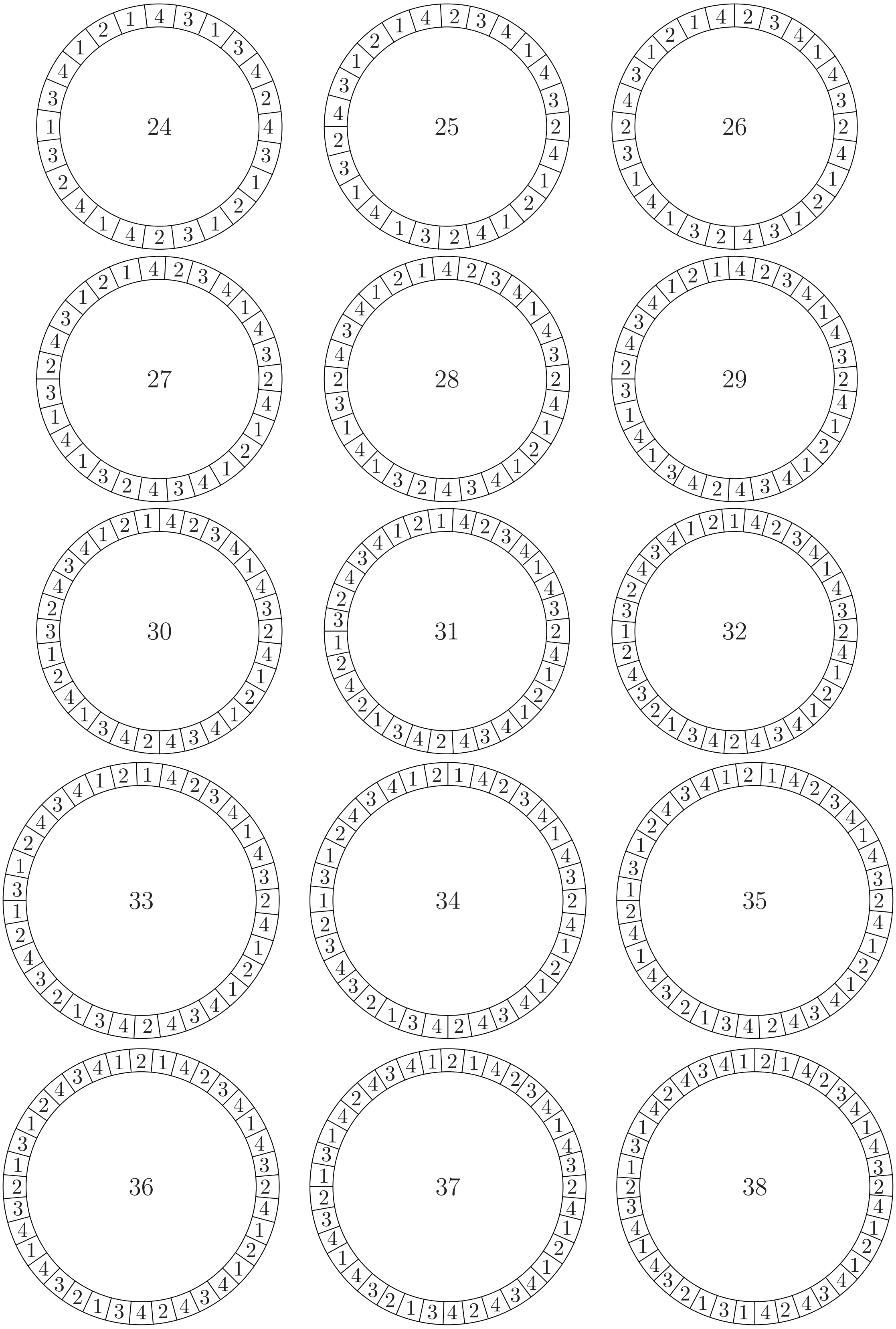


Рис. А.2. $(3/2)^+$-*свободное слово* (w) *над* 4-*буквенным алфавитом длины* n, *где* $24 \leq n \leq 38$.

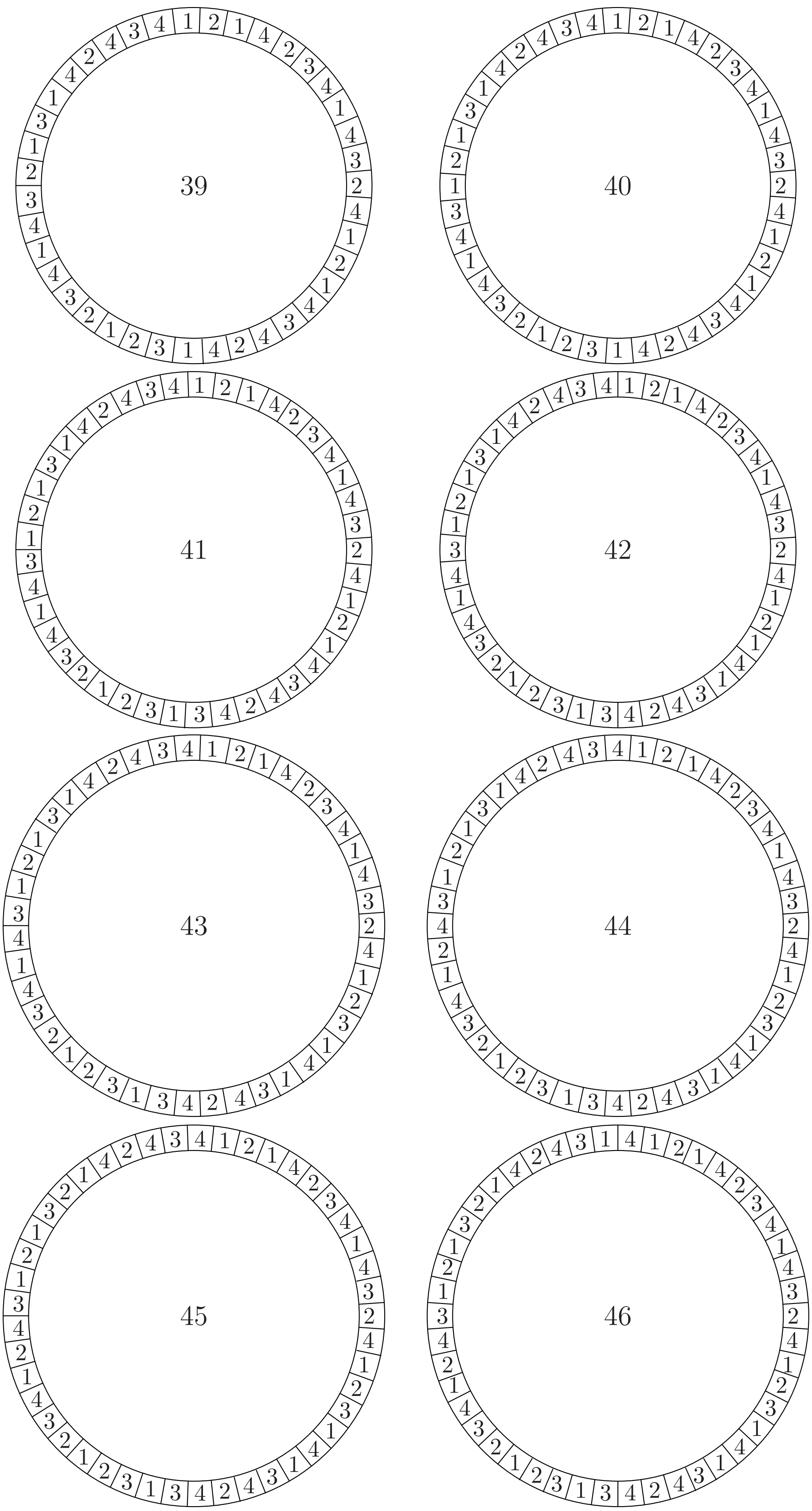


Рис. А.3. $(3/2)^+$*-свободное слово* (w) *над* 4*-буквенным алфавитом длины* n*, где* $39 \leq n \leq 46$.

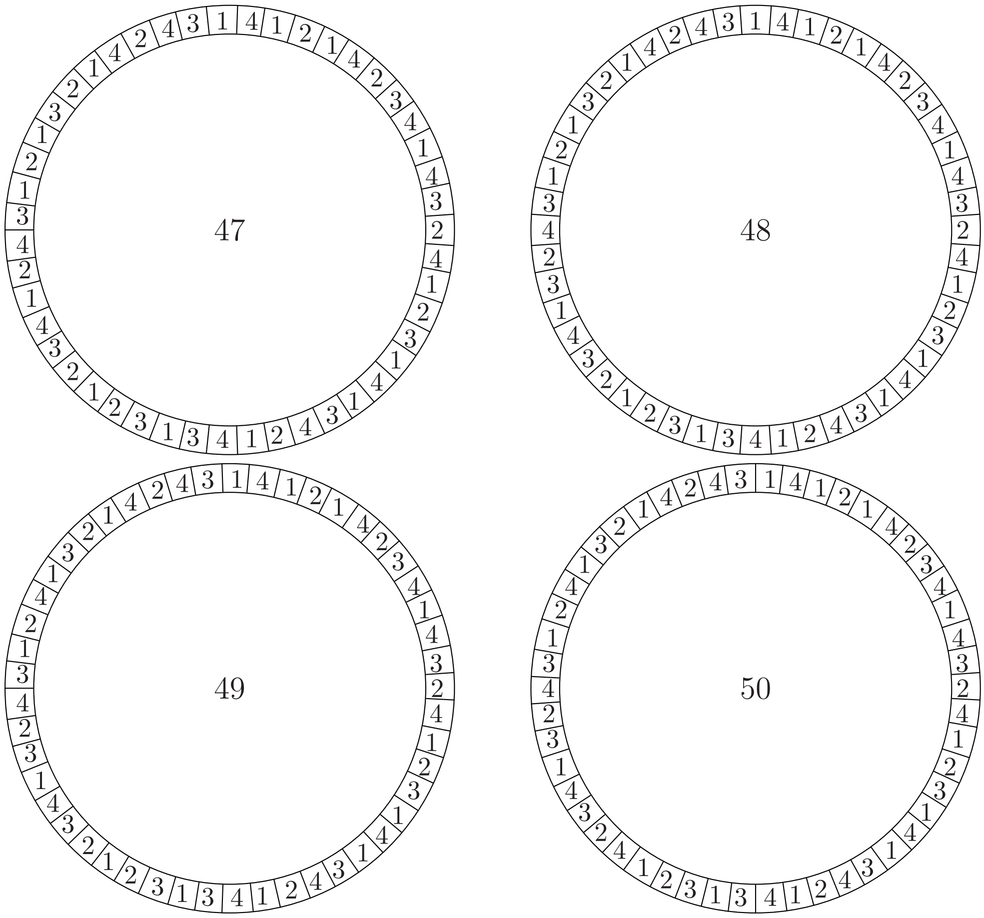


Рис. А.4. $(3/2)^{+}$*-свободное слово* (w) *над* 4*-буквенным алфавитом длины* n, *где* $47 \le n \le 50$.

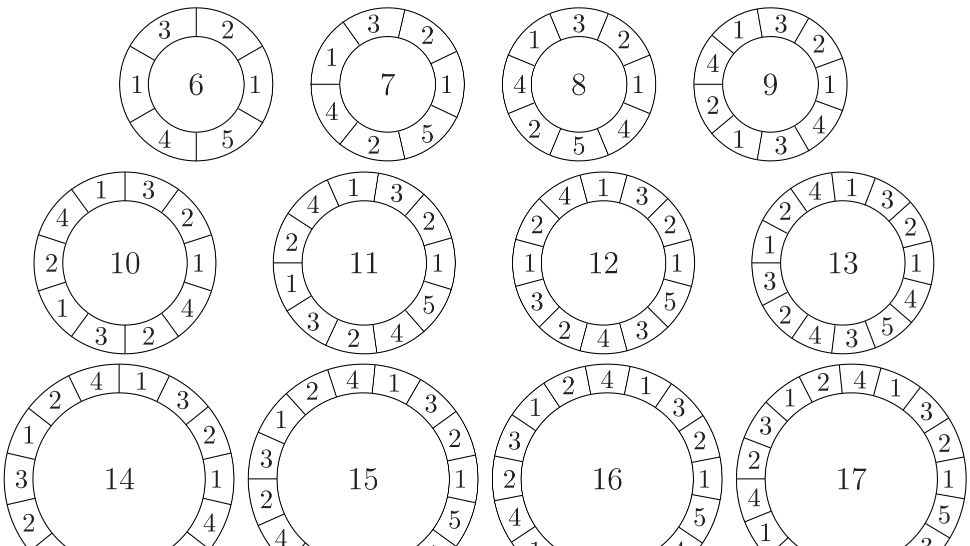


Рис. А.5. $(4/3)^{+}$*-свободное слово* (w) *над* 5*-буквенным алфавитом длины* n, *где* $6 \le n \le 17$.

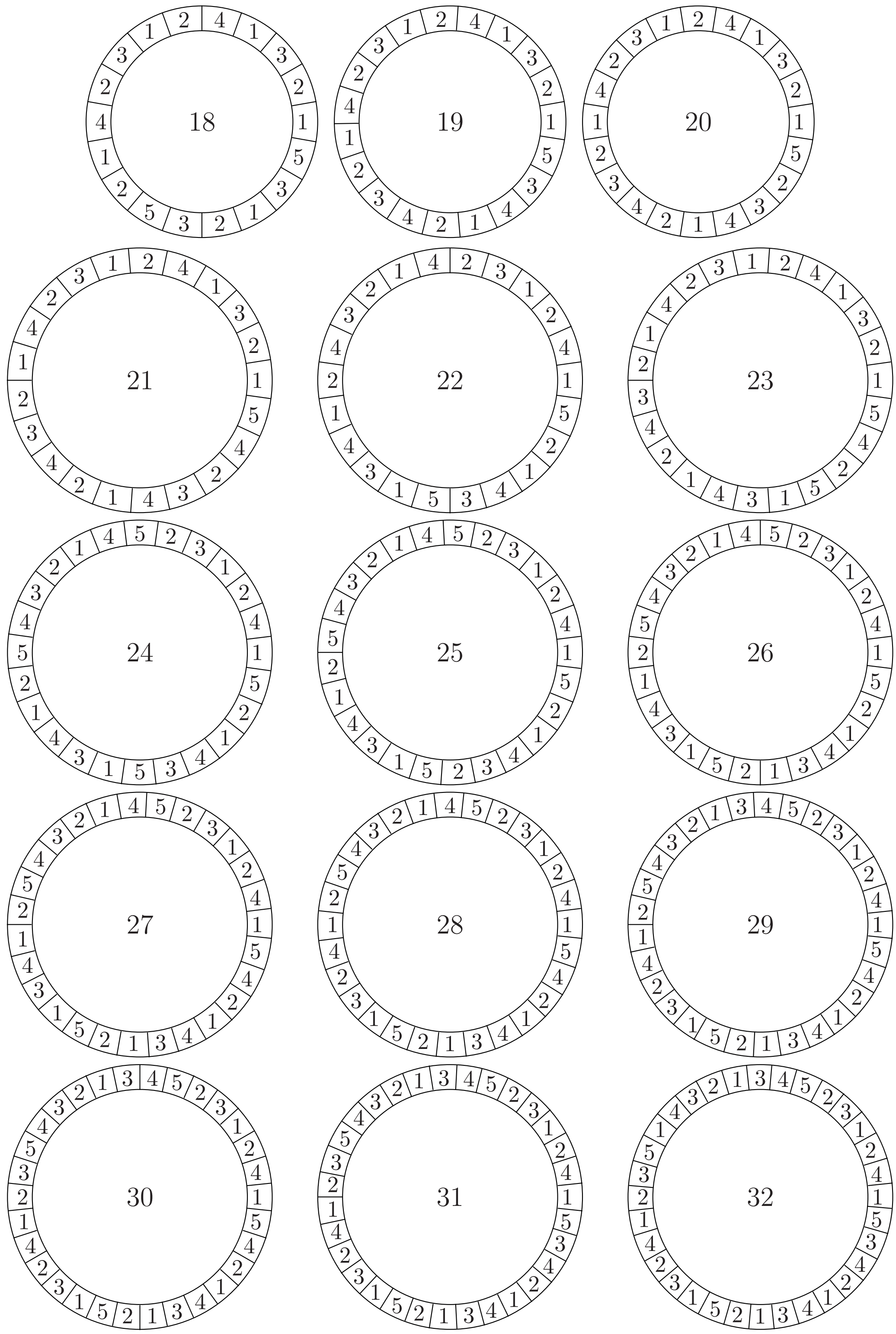


Рис. А.6. $(4/3)^+$-*свободное слово* (w) *над* 5-*буквенным алфавитом длины* n, *где* $18 \le n \le 32$.

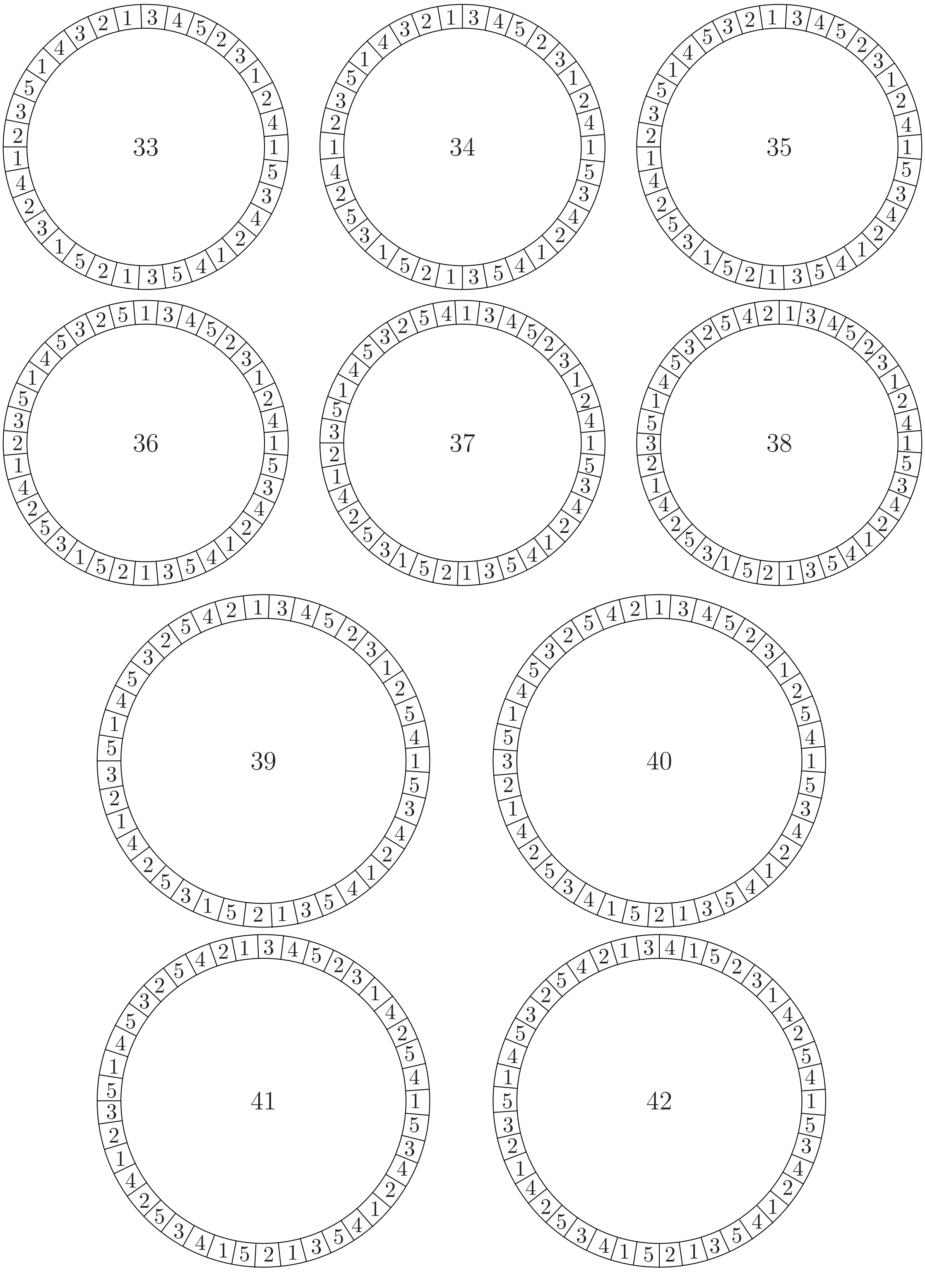


Рис. А.7. $(4/3)^{+}$-*свободное слово* (w) *над* 5-*буквенным алфавитом длины* n, *где* $33 \leq n \leq 42$.

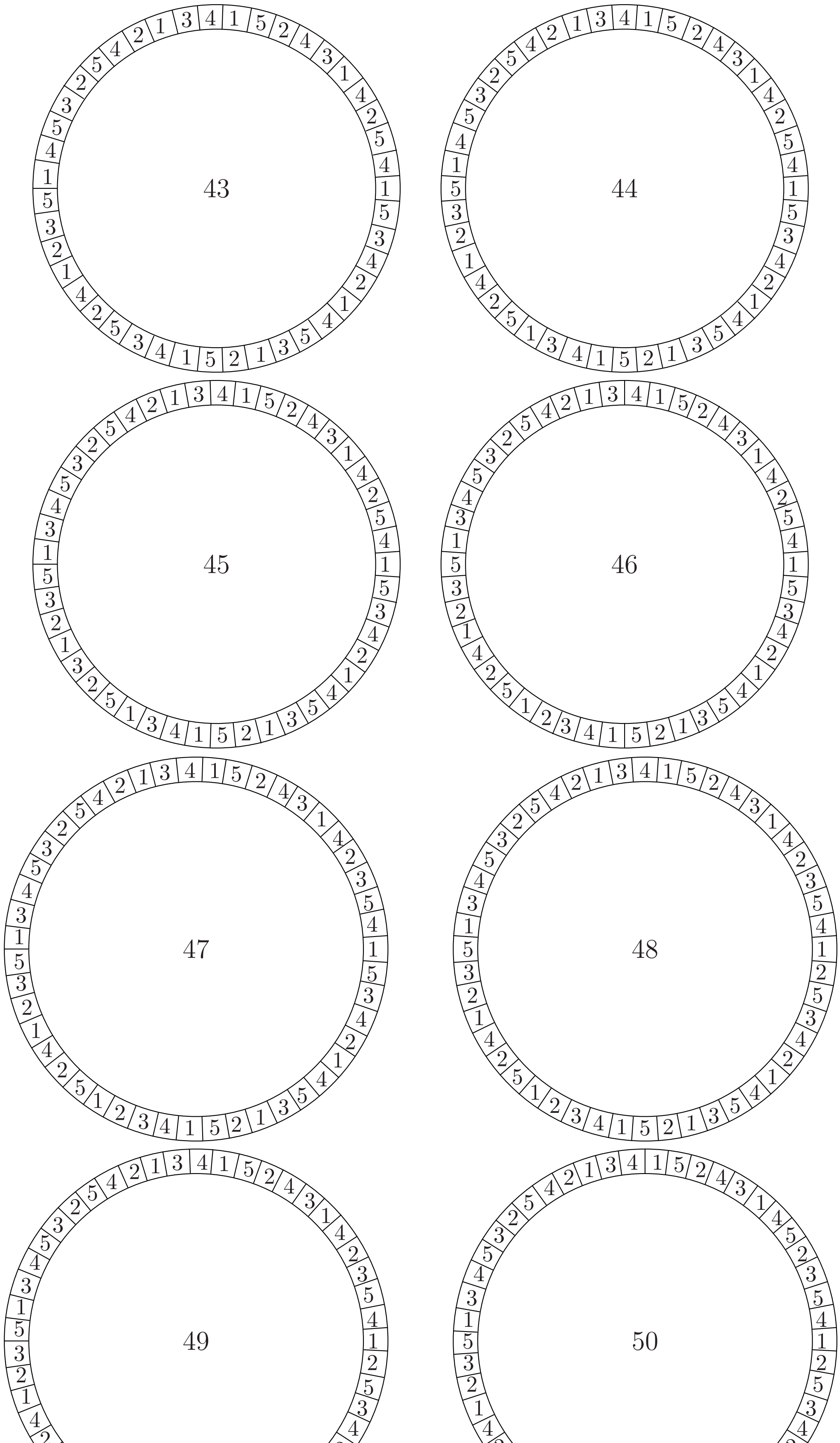


Рис. А.8. $(4/3)^+$-*свободное слово* (w) *над* 5-*буквенным алфавитом длины* n, *где* $43 \leq n \leq 50$.