ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Апрель—июнь 1999. Серия 1. Том 6, № 2, 23–40

УДК 519.114

О КОМБИНАТОРНЫХ СВОЙСТВАХ ПОСЛЕДОВАТЕЛЬНОСТИ АРШОНА*)

А. В. Клепинин, Е. В. Суханов

Рассматриваются комбинаторно-алгебраические свойства формального языка, составленного из конечных слов бесконечной последовательности над трехбуквенным алфавитом, построенной в 30-е годы С. Е. Аршоном [2]. Эта последовательность не содержит два раза подряд повторяющихся слов, т. е. «избегает число 2». Это понятие избегаемости распространяется с целых чисел на рациональные. Показывается, что границей избегаемости рассматриваемого языка является число 7/4. Этот язык задается чередующимся действием двух морфизмов. Предлагаемый метод исследования позволяет изучать данный язык так же, как это делается для формальных языков, задаваемых одним морфизмом. Дается полное описание синтаксической конгруэнции изучаемого языка.

Введение

В той части теории формальных языков, которую принято называть комбинаторикой слов, заметную роль играют бесповторные последовательности в алфавите из конечного числа символов. В начале века А. Туэ построил бесконечную последовательность в алфавите из двух символов, которая не содержит трех одинаковых следующих друг за другом слов. Впоследствии эта последовательность неоднократно переоткрывалась и использовалась в различных комбинаторных построениях (см. например, [7]). С. Е. Аршон [2] построил последовательность в алфавите из трех букв, которая не содержит двух одинаковых следующих друг за другом слов. Именно эта последовательность была использована в качестве одного из базовых элементов для решения ограниченной проблемы Бернсайда в теории групп [1].

Позднее были найдены и другие аналогичные последовательности (см., например, [6, 7, 9]). Исследования бесповторных символьных последовательностей и их обобщений, а также ряда близких задач на рубеже

*) Работа выполнена при финансовой поддержке Грантового центра по исследованиям в области математики при НГУ (грант 46).

70–80-х годов привели различных авторов к более широкой постановке вопроса изучения языков, определяемых множествами запрещенных подслов общего вида. Это нашло свое отражение в исследовании блокирующих (иначе полных) множеств слов [3, 4] и более характерных для алгебраических рассмотрений блокирующих множеств термов и понятия избегаемости слов [5, 8].

Пусть v — слово в алфавите Δ, а u — слово (терм) в алфавите переменных Γ. Слово v содержит значение слова u, если слово $h(u)$, встречается среди подслов в v для некоторого гомоморфизма h: $\Gamma^+ \to \Delta^+$ такого, что $h(x) \neq \varnothing$ для всех $x \in \Gamma$. В этом случае говорят, что слово u *блокирует* слово v. В противном случае, т. е. если среди подслов v нет значений слова u, говорят, что слово v *избегает* слово u.

В этой терминологии любое слово, являющееся частью последовательности Туэ, избегает слово $x^3 = xxx$, а слова в последовательности Аршона избегают слово x^2. Данный подход рассматривает лишь «целые степени» слов. Известно [9], что при подходящем видоизменении понятия избегаемости можно рассматривать и «дробные степени» (определения см. в § 3). Тогда формальный язык, составленный из слов последовательности Туэ, допускает степень 2 и избегает степень $2 + \varepsilon$ при любом $\varepsilon > 0$. Это является отражением того факта, что слова Туэ не только являются «бескубными», но и удовлетворяют более сильному свойству «сильной бескубности» [7].

Таким образом, степень 2 является своеобразной границей избегаемости для языка, составленного из слов последовательности Туэ. Первая из двух основных целей данной работы — найти аналогичную границу для языка, состоящего из слов последовательности Аршона. Эта граница, равная $7/4$, содержится в теореме 1. В несколько иной терминологии в [11] построена последовательность над трехбуквенным алфавитом, для которой граница избегаемости тоже равна $7/4$. Наш результат показывает, что для достижения цели годится ставшая классической последовательность Аршона. Эта последовательность задается чередующимся действием двух морфизмов, в то время как более распространенным является способ задания формального языка с помощью одного морфического генератора. Поэтому авторы предприняли попытку предложить метод, который позволял бы работать с двумя морфизмами примерно так же, как и с одним. Этот метод позволил получить и второй из двух основных результатов данной работы.

Как известно, интересные свойства формального языка можно извлечь, изучая его синтаксическую полугруппу, если рассматриваются непустые слова. Так, например, несложно получаемые свойства синтаксической полугруппы языка, составленного из слов последовательности Аршона, показывают, что этот язык является апериодическим и не является рациональным (соответствующие определения см., например, в [12]). С комбинаторно алгебраической точки зрения синтаксическая полугруппа — это фактор полугруппа по синтаксической конгруэнции языка. Вторым основным результатом данной работы является описание синтаксической конгруэнции языка, составленного из слов последовательности Аршона. Это описание содержится в теореме 2.

Несколько слов о структуре данной работы. Первый параграф содержит сводку основных понятий. Второй параграф посвящен изучению некоторых структурных свойств последовательности Аршона, связанных с ее порождающим правилом. Третий и четвертый параграфы посвящены нахождению границы избегаемости языка, состоящего из слов последовательности Аршона, и изучению его синтаксической конгруэнции соответственно.

§ 1. Основные понятия

Следуя [7] определим основные понятия. *Алфавитом* Σ называется конечное непустое множество. *Буквами* (*символами*) называются элементы из Σ.

Слово над алфавитом Σ — это конечная цепочка, состоящая из букв алфавита Σ. Цепочка, состоящая из нулевого количества букв, называется *пустым словом* и обозначается через λ. Множество всех слов (соответственно всех непустых слов) над Σ обозначается через Σ^* (соответственно через Σ^+). В дальнейшем для обозначения слов, как правило, будем использовать строчные латинские буквы, хотя в некоторых ситуациях для этой цели придется задействовать греческий алфавит.

Конкатенацией слов x и y называется слово xy. Конкатенация является ассоциативной операцией, и пустое слово является единицей по отношению к ней: $x\lambda = \lambda x = x$ для каждого x. Таким образом, Σ^+ — полугруппа, а Σ^+ — моноид. Они носят название *свободных.* Если x — слово, а i — натуральное число, то через x^i будем обозначать слово, полученное конкатенацией i слов, каждое из которых есть x. По определению полагаем $x^0 = \lambda$.

Длина слова x обозначается через $|x|$. По определению полагаем $|\lambda| = 0$. Легко понять, что

$$|xy| = |x| + |y|, \quad |x^i| = i|x|.$$

Пусть x и y — слова и для некоторых $x_1, x_2 \in \Sigma^*$ имеет место равенство $y = x_1 x x_2$. Тогда x называется *подсловом* слова y. В этом случае будем использовать обозначение $x \leqslant y$ (соответственно $x < y$, если $x \leqslant y$ и $x \neq y$) и говорить, что x *входит* в y. При этом под *вхождением* слова x

понимается его конкретное расположение в y. *Морфизмом* называется отображение $h : \Sigma^* \longrightarrow \Delta^*$, где Σ и Δ — алфавиты, удовлетворяющее условию $h(xy) = h(x)h(y)$. Другими словами, h — это гомоморфизм свободных моноидов. Произвольное подмножество из Σ^* будем называть (*формальным*) *языком.*

Перейдем к рассмотрению последовательности Аршона. Определим ее, следуя [1]. В качестве основного алфавита берется алфавит $\Sigma = \{1, 2, 3\}$. Систему *слов Аршона* $\{w_k\}_{k=1}^{\infty}$ определим индуктивно. По определению полагаем $w_1 = 1$. Пусть w_k уже построено. Тогда w_{k+1} получается из w_k в результате посимвольной замены:

для символов, находящихся в нечетной позиции:	для символов, находящихся в четной позиции:
$1 \longrightarrow 123$	$1 \longrightarrow 321$
$2 \longrightarrow 231$	$2 \longrightarrow 132$
$3 \longrightarrow 312$	$3 \longrightarrow 213.$

Таким образом, $w_2 = 123$, $w_3 = 123\ 132\ 312$ и т. д. Образы символов будем называть *блоками*. Будем различать четные и нечетные блоки в зависимости от четности позиции символа, из которого получается блок. Так, например, блок 123 является нечетным, а блок 213 — четным. Отметим, что по двум символам четность блока восстанавливается однозначно, а при известной четности блок единственным образом определяется по любому одному его символу.

Ясно, что каждое слово w_i является конкатенацией некоторого числа блоков. Слова, обладающие таким свойством, будем называть *целыми*. При этом разбиение слова на составляющие его блоки будем называть *естественным*. В дальнейшем нам потребуется знать конкретное расположение некоторого слова v в w_i относительно естественного разбиения слова w_i. Для этого определим следующие две величины: $\mathrm{lsp}(v)$ (соответственно $\mathrm{rsp}(v)$) — количество тех букв в левом (правом) блоке, пересекающемся со словом v, которые не принадлежат слову v. Из определения следует, что обе величины могут принимать лишь значения 0, 1 или 2. Смысл этих величин иллюстрирует следующая схема *):

$$\ldots\ \star\star\star\ \ \star\star\star\ \ \underbrace{\underline{\star}}_{\mathrm{lsp}(v)}\overbrace{\underbrace{\star\star\ \ \star\cdots\star\ \ \star}_{|v|}}^{v}\underbrace{\underline{\star\star}}_{\mathrm{rsp}(v)}\ \ \star\star\star\ \ldots$$

Здесь $\mathrm{lsp}(v) = 1$ и $\mathrm{rsp}(v) = 2$.

Легко понять, что каждое w_k является началом (*префиксом*) всех слов w_i при $i \geqslant k$. Следовательно, можно говорить о «пределе» последовательности слов $\{w_k\}$, т. е. о бесконечной вправо цепочке символов

*) Здесь и далее пробелами выделено естественное разбиение слова на блоки.

такой, что множеством ее префиксов является множество $\{w_k\}_{k=1}^{\infty}$. Такие бесконечные вправо цепочки будем называть *сверхсловами.* Сверхслово, порожденное системой $\{w_k\}_{k=1}^{\infty}$, является последовательностью Аршона; оно обозначается через A_w. Таким образом, A_w имеет вид

$$A_w = \underbrace{\overbrace{\underbrace{123}_{w_2}\ 132\,312}^{w_3}\ 321\,312\,132\,312\,321\,231}_{w_4}\ 213\,231\,321\,312\,321\,231\ \ldots$$

Сверхслово A_w обладает следующими свойствами:

Лемма 1. 1) *Любой нечетный блок в A_w начинается с нечетной позиции, а четный — с четной позиции.*

2) *Четности любых соседних блоков в A_w различны.*

3) *Последовательность A_w не содержит квадратов, т. е. подслов вида xx, где $x \in \Sigma^+$.*

4) *Каждое подслово из A_w повторяется в A_w счетное число раз.*

5) *Слово длины не более 3^k входит в A_w тогда и только тогда, когда оно входит в w_{k+5}.*

Доказательство. Свойства 1 и 2 очевидны. Бесквадратность A_w (свойство 3) доказана Аршоном [2].

Докажем свойство 4. Действительно, пусть $v < A_w$. Тогда $v \leqslant w_k$ для подходящего*) $k \in \mathbb{N}_0$. Слово w_k получается из $w_1 = 1$ путем применения $k-1$ раз подстановки (1). Но слово $w_1 = 1$ встречается в нечетных позициях последовательности A_w счетное число раз (тривиально проверяется с использованием свойств 1–3). Следовательно, все «образы» этого слова встречаются в A_w счетное число раз.

Докажем свойство 5. Пусть $|v| \leqslant 3^k$. Достаточно проверить, что если $v \leqslant A_w$, то $v \leqslant w_{k+5}$. Легко понять, что при указанных условиях $v \leqslant u$, где u получается из одной буквы применением $k+1$ раз подстановки (1). Так же непосредственно проверяется, что слово w_4 содержит все буквы из Σ как в четных, так и в нечетных позициях. Значит, $v \leqslant u \leqslant w_{k+1}$. Лемма 1 доказана.

Через LA обозначается множество всех подслов, содержащихся в A_w, т. е. $LA = \{v \mid v \leqslant A_w\}$. Несмотря на достаточно простое правило порождения, последовательность A_w имеет достаточно сложную локальную структуру. Поэтому прежде чем перейти к рассмотрению интересующих нас характеристик языка LA, в следующем параграфе сформулируем и докажем несколько вспомогательных утверждений, позволяющих прояснить локальное строение последовательности A_w.

*) Здесь и далее $\mathbb{N}$ — множество натуральных чисел; $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$; $\mathbb{Z}$ — множество целых чисел.

§ 2. Вспомогательная конструкция

Пусть v и u — произвольные вхождения в A_w произвольных слов. Будем говорить, что v и u *блочно синхронны слева* (*справа*), если вхождения v и u начинаются (кончаются) в блоках одной четности. Блочно синхронные слева и справа вхождения будем называть *блочно синхронными*. Будем говорить, что v и u *синхронны слева* (*справа*), если вхождения v и u начинаются (кончаются) в одной и той же позиции (возможно различных блоков) одной четности. Вхождения, синхронные слева и справа, будем называть *синхронными*.

Например, следующие вхождения слов $u = 123$ и $v = 232$ блочно синхронны слева:

$$\underline{123}\ 132\ 31\underline{2\ 32}1,$$

но не блочно синхронны справа. Обратим внимание, что понятия синхронизации относятся к конкретным вхождениям слов. Например, следующие вхождения слова $v = 231$ не являются блочно синхронными ни слева, ни справа:

$$1\underline{23\ 1}3\underline{2\ 31}2.$$

Однако в некоторых случаях различные вхождения одного и того же слова могут оказаться синхронными:

$$123\ 132\ \underline{312}\ 321\ \underline{312}.$$

Оказывается, синхронизация различных вхождений одного и того же слова в A_w непосредственно связана со структурой последовательности A_w.

Лемма 2. *Пусть $v \in LA$ и $|v| \geqslant 8$. Тогда все вхождения слова v в A_w блочно синхронны.*

Доказательство. Рассмотрим возможные случаи. Пусть v_1 и v_2 — различные вхождения слова $v = a_1a_2a_3a_4a_5\ldots$ в A_w. Покажем, что v_1 и v_2 блочно синхронны слева. Для этого разберем все варианты расположения v_1 и v_2 относительно естественного разбиения.

1. Если $\mathrm{lsp}(v_1) \in \{0,1\}$ и $\mathrm{lsp}(v_2) \in \{0,1\}$, то слова a_1a_2 однозначно определяют одну и ту же четность блоков, в которых начинаются v_1 и v_2.
2. Пусть $\mathrm{lsp}(v_1) = 1$ и $\mathrm{lsp}(v_2) = 2$, т. е.

$$\begin{aligned} v_1 &= a_1a_2\ \underline{a_3a_4a_5}\ a_6\ldots \\ v_2 &= \quad a_1\ \underline{a_2a_3a_4}\ a_5a_6\ldots \end{aligned}$$

Выделенные блоки должны иметь одну и ту же четность, поскольку они содержат общее двухбуквенное подслово a_3a_4. Но в A_w четности соседних блоков всегда различны. Следовательно, вхождения v_1 и v_2 блочно синхронны.

Это же рассуждение остается верным, когда $\operatorname{lsp}(v_1) \in \{1,2\}$ и $\operatorname{lsp}(v_2) \in \{1,2\}$.

3. Ситуация, когда $\operatorname{lsp}(v_1) = 0$ и $\operatorname{lsp}(v_2) = 2$ (симметричный случай $\operatorname{lsp}(v_1) = 2$ и $\operatorname{lsp}(v_2) = 0$), в условиях леммы невозможна. Действительно, в этом случае расположение v_1 и v_2 относительно естественного разбиения A_w имеет вид

$$\begin{aligned} v_1 &= a_1a_2a_3\ \ldots \\ v_2 &= \qquad a_1\ \underline{a_2a_3}\ldots \end{aligned}$$

Заметим, что по выделенному слову a_2a_3 однозначно определяется следующий символ в v, поскольку $a_1a_2a_3$ образуют блок, а следовательно, $a_1 \neq a_2 \neq a_3 \neq a_1$ и a_2a_3 — начало блока. Значит, $v = a_1a_2a_3a_1\ldots$ (т. е. $a_4 = a_1$) и мы имеем

$$\begin{aligned} v_1 &= a_1a_2a_3\ \underline{a_1}\ldots \\ v_2 &= \qquad a_1\ a_2a_3a_1\ldots \end{aligned}$$

Учитывая смену четности у соседних блоков, выделенный символ a_1 однозначно определяет остаток блока ($a_5a_6 = a_3a_2$). Значит, $v = a_1a_2a_3a_1a_3a_2\ldots$.

Продолжая рассуждать подобным образом и учитывая условие $|v| \geqslant 8$, получаем, что v должно иметь вид $v = a_1a_2a_3a_1a_3a_2a_1 a_2\ldots$, а для его вхождений v_1 и v_2 имеем

$$\begin{aligned} v_1 &= a_1a_2a_3\ a_1a_3a_2\ a_1a_2\ldots \\ v_2 &= \qquad \underline{a_1}\ a_2a_3a_1\ a_3a_2a_1\ \underline{a_2}\ldots \end{aligned}$$

Вновь по выделенным символам с учетом смены четности однозначно определяются остатки блоков. Таким образом, вхождение v_2 в A_w имеет вид

$$\overbrace{a_3a_2\,\underbrace{\underline{a_1}\ a_2a_3a_1}}^{p}{}_{}\overbrace{\underbrace{a_3a_2a_1\ \underline{a_2}}_{v_2}a_3a_1}^{p} < A_w.$$

Но квадрат pp противоречит бесквадратности последовательности Аршона. Значит, этот случай невозможен.

Доказательство правой блочной синхронности проводится точно так же. Лемма 2 доказана.

Итак, если рассматривать достаточно длинные слова из A_w, Для определенных слов из A_w эти закономерности позволят ввести понятия образа и прообраза для порождающей подстановки (1). Следующее утверждение является ключевым.

Лемма 3. *Если $v \in LA$ и $|v| \geqslant 9$, то все вхождения v в A_w синхронны.*

Доказательство. Действительно, из предыдущей леммы 2 следует, что все вхождения v в A_w блочно синхронны. Осталось доказать синхронность. Для этого рассмотрим все возможные случаи. Пусть v_1 и v_2 — различные вхождения слова v в A_w. Покажем, что v_1 и v_2 синхронны слева.

Достаточно проверить, что $\mathrm{lsp}(v_1) = \mathrm{lsp}(v_2)$. Для этого убедимся в невозможности следующих случаев:

(a) $\mathrm{lsp}(v_1) = 0$ и $\mathrm{lsp}(v_2) = 1$ (т. е. v_1 начинается с первой, а v_2 — со второй буквы блоков);

(b) $\mathrm{lsp}(v_1) = 0$ и $\mathrm{lsp}(v_2) = 2$ (т. е. v_1 начинается с первой, а v_2 — с третьей буквы блоков);

(c) $\mathrm{lsp}(v_1) = 1$ и $\mathrm{lsp}(v_2) = 2$ (т. е. v_1 начинается со второй, а v_2 — с третьей буквы блоков).

Ограничимся рассмотрением случая (a), поскольку именно он представляется наиболее интересным. Для разбора остальных случаев используются аналогичные рассуждения.

Итак, пусть v_1 начинается с первой, а v_2 — со второй буквы блоков, причем четности этих блоков совпадают. Пусть $v = a_1a_2a_3\ldots$. Повторяя рассуждения, аналогичные приведенным в доказательстве леммы 2, получаем, что v должно иметь следующий вид: $v = a_1a_2a_3a_2a_1a_3a_1a_2$ $a_3\ldots$, а

$$\begin{aligned} v_1 &= a_1a_2a_3\; a_2a_1a_3\; a_1a_2a_3\ldots \\ v_2 &= \quad\underline{a_1a_2}\; a_3a_2a_1\; a_3a_1a_2\; \underline{a_3}\ldots \end{aligned}$$

Опять с учетом четности по выделенным символам однозначно определяются остатки блоков и, следовательно,

$$\overbrace{a_3\,\mathrlap{\underbrace{}_{v_2}}a_1a_2\,a_3a_2a_1}^{p}\,\overbrace{a_3a_1a_2\,a_3\,a_2a_1}^{p} < A_w.$$

Получили противоречие с бесквадратностью последовательности Аршона.

Рассмотрение оставшихся случаев и доказательство правой синхронности проводятся аналогично. Лемма 3 доказана.

Обратим внимание, что требования к длине слов в леммах 2 и 3 не завышены, поскольку для слова $v = 1231321$ ($|v| = 7$) существуют его вхождения, не являющиеся блочно синхронными, а для $v = 12321312$ ($|v| = 8$) существуют его вхождения, не являющиеся синхронными.

Лемма 3 позволяет построить конструкцию, важную для исследования последовательности Аршона. Рассмотрим произвольное слово $v \in LA$ такое, что $|v| \geqslant 9$. Тогда в силу леммы 3 все вхождения слова v в A_w синхронны. В частности, это означает, что все вхождения слова v в A_w начинаются с позиций одинаковой четности (см. лемму 1). Но тогда при подстановке (1) все вхождения слова v в A_w будут иметь один и тот же образ. Будем обозначать его через $\varphi(v)$ и называть *естественным образом* v относительно подстановки (1). Ясно, что понятие естественного образа определено корректно лишь для слов из A_w длины не меньшей 9. Для слов меньшей длины мы не можем гарантировать единственность образа. Обозначение $\varphi(v)$ будем использовать лишь в тех случаях, когда оно корректно.

Аналогично если $v \in LA$, $|v| \geqslant 9$, и все вхождения слова v (они синхронны!) начинаются с первой буквы и кончаются на последней букве блоков ($\mathrm{lsp}(v) = \mathrm{rsp}(v) = 0$, т. е. v — целое), то для v однозначно определен *естественный прообраз* — слово $u < A_\omega$ такое, что v получается из некоторого вхождения слова u в результате применения подстановки (1). В этом случае будем использовать обозначение $u = \varphi^{-1}(v)$.

Например, для слова $v = 312321312$ имеем

$$\varphi(v) = 3123212312132313213123212 31$$

и $u = \varphi^{-1}(v) = 313$. Ясно, что путем применения подстановки (1) из слова $u = 313$ можно получить слово $v' = 213123213$, причем $v' \neq v$. Значит, в общем случае даже существование естественного прообраза $\varphi^{-1}(u)$ для произвольного слова $u \in LA$ не гарантирует наличие естественного образа $\varphi(\varphi^{-1}(u))$. Однако равенство $\varphi^{-1}(\varphi(u)) = u$ верно для произвольного $u \in LA$, для которого определен естественный образ $\varphi(u)$.

Итак, справедлива следующая

Лемма 4. *Любое слово v из LA, $|v| \geqslant 9$, обладает следующими свойствами:*

1) *для v определен естественный образ $\varphi(v)$;*

2) *если $\mathrm{lsp}(v) = \mathrm{rsp}(v) = 0$, то для v определен естественный прообраз $\varphi^{-1}(v)$;*

3) $\varphi^{-1}(\varphi(v)) = v$;

4) *если для v определен естественный прообраз $\varphi^{-1}(v)$ такой, что $|\varphi^{-1}(v)| \geqslant 9$, то $\varphi(\varphi^{-1}(v)) = v$.*

Тем самым показано, что на достаточно длинных словах подстановка (1) ведет себя как «обратимый» морфизм. Знание этого факта поможет нам в изучении свойств языка LA.

§ 3. Граница избегаемости языка LA

В предыдущих параграфах упоминалось, что последовательность A_ω не содержит подслов вида xx, т. е. избегает слово x^2. Более точно понятие избегаемости определяется следующим образом. Пусть Δ и Γ — алфавиты. Говорят, что слово $v \in \Delta^*$ *допускает* слово $u \in \Gamma^*$, если $\vartheta(u) \leqslant v$ для подходящего морфизма $\vartheta : \Gamma^* \longrightarrow \Delta^*$. В противном случае слово v *избегает* слово u.

Понятие избегаемости можно видоизменить следующим образом. Пусть $v \in \Delta^*$. Число p называют *периодом* слова v, если для каждого $k \in \{1, 2, \ldots, |v| - p\}$ символы $v[k]$ и $v[k+p]$ совпадают. *Наименьший период* слова v обозначается через $\mathrm{per}(v)$, а величина $\exp(v) = |v|/\mathrm{per}(v)$ называется *экспонентой* слова v. *Наследственной экспонентой* слова v называется величина $\mathrm{hexp}(v) = \max\{\exp(u) \mid u \leqslant v\}$. Для сверхслов наследственная экспонента вводится аналогично: $\mathrm{hexp}(v_\omega) = \sup\{\exp(u) \mid u \leqslant v_\omega, |u| < \infty\}$. Понятно, что это же определение годится и для произвольного языка L: $\mathrm{hexp}(L) = \sup\{\exp(u) \mid u < v \in L\}$.

Будем говорить, что слово (сверхслово, язык) *допускает экспоненту* k, если его наследственная экспонента не меньше k, и *избегает экспоненту* k в противном случае. Таким образом, наследственная экспонента делит множество экспонент на две части: допускаемые и избегаемые. Границу этого разбиения (т. е. наследственную экспоненту) будем называть *границей избегаемости.*

Легко понять, что в случае рационального k для введенных понятий можно дать эквивалентное определение: слово $v \in \Delta^*$ допускает экспоненту k, если существует число $n \in \mathbb{N}_0$ и слова x и y такие, что $|x| = a$, $|y| = b$, слово $(xy)^n x$ является подсловом слова v и $k = n + \frac{a}{a+b}$. В дальнейшем нам будет удобнее пользоваться именно вторым определением и при этом иметь дело с последовательностью Аршона, а не с LA. Поскольку последовательность A_ω не содержит квадратов, то $1 \leqslant \mathrm{hexp}(LA) \leqslant 2$. Более того, справедлива

Теорема 1. *Граница избегаемости языка* LA *равна* $7/4$ *и допускается словом* $3121312 < A_\omega$.

Доказательство. Из оценки $1 \leqslant \mathrm{hexp}(LA) \leqslant 2$ следует, что для изучения границы избегаемости языка LA достаточно рассматривать все возможные вхождения в A_ω слов вида vuv. Поэтому в дальнейшем будем иметь дело со словом $vuv < A_\omega$. При этом пусть $a = |v|$ и $b = |u|$.

Заметим, что язык LA не избегает экспоненту $7/4$, поскольку для $v = 312$ и $u = 1$ слово $vuv = 3121312$ принадлежит языку LA. Но тогда в силу определения наследственной экспоненты достаточно рассматривать лишь такие слова v и u, что $\frac{a}{a+b} > \frac{3}{4}$. Следовательно, можно считать, что $a > 3b$.

Далее, если в языке LA есть достаточно большое слово вида xyx, то, оказывается, всегда можно указать слово меньшей длины с не меньшей экспонентой.

Действительно, пусть $2a + b \geqslant 19$. Тогда из неравенства $a > 3b$ следует, что $\frac{7a}{3} \geqslant 19$. Поэтому $a \geqslant \frac{57}{7} > 8$. Поскольку $a \in \mathbb{N}$, то $a \geqslant 9$. Значит, в силу леммы 3 все вхождения слова v синхронны. Этим же свойством обладают и все вхождения слова vuv, так как его длина также не меньше 9. Поэтому дальнейшие рассуждения будут справедливы для любого вхождения слова vuv.

Если v начинается не с первой буквы блока ($\operatorname{lsp}(v) > 0$), то его можно продолжить влево до начала блока, т. е. найти такое слово $q(|q| = \operatorname{lsp}(v))$, что $qvuv < A_\omega$ и $\operatorname{lsp}(qvuv) = 0$. Далее, если $|u| \leqslant |q|$, то в силу синхронизации вхождений слова v получаем, что $q = q'u$ для подходящего $q' \in \Sigma^*$. Но тогда $qvuv = q'\underline{uvuv} < A_\omega$, что противоречит бесквадратности последовательности A_ω.

Значит, $|u| > |q|$. Вновь в силу синхронизации вхождений слова v получаем, что $u = u'q$ для подходящего $u' \in \Sigma^+$. Следовательно, $qvu'qv < A_\omega$

$$\frac{|qv|}{|qvu'|} = \frac{|q| + |v|}{|q| + |v| + |u'|} = \frac{|q| + a}{|v| + |u'q|} = \frac{|q| + a}{a + b} > \frac{a}{a + b}.$$

Таким образом, полученное слово $qvu'qv$ имеет бóльшую экспоненту, нежели исходное слово vuv. Поэтому в дальнейшем будем считать, что $\operatorname{lsp}(v) = 0$. Аналогично можно считать, что $\operatorname{rsp}(v) = 0$, так как в противном случае можно применить процедуру продолжения, описанную выше для наращивания слова v вправо.

Итак, достаточно рассматривать лишь целые слова v. А это значит, что слово vuv также будет целым. Но тогда для vuv определен естественный прообраз $\tilde{v}\tilde{u}\tilde{v} = \varphi^{-1}(vuv)$, причем $|\tilde{v}| = \frac{1}{3}|v|$ и $|\tilde{u}| = \frac{1}{3}|u|$. Следовательно,

$$\frac{|\tilde{v}|}{|\tilde{v}\tilde{u}|} = \frac{\frac{1}{3}a}{\frac{1}{3}(a + b)} = \frac{a}{a + b},$$

т. е. нашлось слово втрое меньшей длины с той же экспонентой, что и исходное слово. Таким образом, для произвольного слова $vuv < A_\omega$, удовлетворяющего условиям

$$|vuv| \geqslant 19 \quad \text{и} \quad \frac{|v|}{|vu|} > \frac{3}{4},$$

путем применения указанной процедуры можно построить слово длины не более 19, имеющее не меньшую экспоненту. Тем самым доказана следующая

Лемма 5. *Для нахождения границы избегаемости языка LA достаточно рассматривать слова vuv длины не более* 18.

Обратим внимание, что при доказательстве этой леммы было получено необходимое условие для длин слов v и u. Поэтому для нахождения наследственной экспоненты последовательности Аршона достаточно рассматривать такие слова $vuv \in \Sigma^*$, что $3 \leqslant 3b < a \leqslant 8$ (здесь $a = |v|$, $b = |u|$). При этих условиях все возможные варианты значений a и b приведены в следующей таблице:

$a = \|v\|$	$b = \|u\|$	$\exp(vuv) = \frac{2a+b}{a+b}$
4	1	$\frac{9}{5}$
5	1	$\frac{11}{6}$
6	1	$\frac{13}{7}$
7	1	$\frac{15}{8}$
7	2	$\frac{16}{9}$
8	1	$\frac{17}{9}$
8	2	$\frac{18}{10}$

Согласно лемме 1 остается лишь проверить, входят ли слова вида vuv с длинами $a = |v|$ и $b = |u|$ в w_9, т. е. в начальный отрезок длины 3^8 последовательности A_ω. Выполняя проверку этого условия полным перебором на компьютере, получаем следующий результат.

Лемма 6. *Граница избегаемости языка* $LA|_{18} = \{v \mid v \in LA, |v| \leqslant 18\}$ *равна* $7/4$.

Таким образом, из леммы 5 следует, что $\mathrm{hexp}(LA) = \mathrm{hexp}(LA|_{18})$, а в силу леммы 6 имеем $\mathrm{hexp}(LA|_{18}) = 7/4$. Следовательно, $\mathrm{hexp}(LA) = 7/4$, что и завершает доказательство теоремы 1.

Итак, мы нашли границу избегаемости языка LA и убедились в ее допустимости. В частности, можно утверждать, что если некоторое слово vuv принадлежит языку A_ω, то $|u| \geqslant \frac{1}{3}|v|$. Таким образом, два вхождения одного и того же слова должны находиться на определенном расстоянии друг от друга, причем это расстояние тем больше, чем больше длина слова.

Это перекликается с результатами работы [11]. С одной стороны, доказано существование последовательности над трехбуквенным алфавитом, в котором любое подслово вида vuv удовлетворяет неравенству $|u| \geqslant \frac{1}{3}|u|$. Несложно понять, что такая последовательность имеет границу избегаемости $7/4$. Эта последовательность порождается морфизмом, у которого образы всех букв имеют длину 19, в то время как у подстановки (1) образы букв имеют длину 3.

С другой стороны, известно [9], что никакое слово длины большей 39 не избегает экспоненты 7/4, т. е. эта экспонента является минимальной возможной наследственной экспонентой для последовательностей над алфавитом из трех букв.

§ 4. Синтаксическая конгруэнция языка LA

Синтаксическая конгруэнция σ_L формального языка L над алфавитом Σ определяется следующим образом: $(u, v) \in \sigma_L$ тогда и только тогда, когда для произвольных $p, q \in \Sigma^*$ слова́ puq и pvq одновременно принадлежат или не принадлежат языку L. Фактор полугруппа Σ^+/σ_L называется синтаксической полугруппой языка L. Нас будет интересовать σ_{LA}.

Множество $I = \Sigma^+ \setminus LA$ является идеалом в свободной полугруппе Σ^+. Этому идеалу соответствует конгруэнция, которую мы обозначаем через ρ_I. Одним из ее классов является идеал I, а остальные классы одноэлементны. Конгруэнция ρ_I — это так называемая *рисовская конгруэнция* полугруппы Σ^+ по идеалу I.

Теорема 2. *Для языка LA справедливо равенство $\sigma_{LA} = \rho_I$.*

Прежде чем перейти к доказательству, необходимо провести ряд промежуточных рассуждений. Сначала обратим внимание на то, что язык LA в силу данных определений является объединением классов конгруэнции ρ_I. Известно, что синтаксическая конгруэнция языка является наибольшей конгруэнцией с таким свойством. Следовательно, $\rho_I \subset \sigma_{LA}$. Далее, если $(u, v) \in \sigma_{LA}$, то из определения синтаксической конгруэнции вытекают следующие импликации:

(a) $u \notin LA \Rightarrow v \notin LA \Rightarrow (u, v) \in \rho_I$;

(b) $u \in LA \Rightarrow v \in LA$.

Значит, для доказательства равенства конгруэнций достаточно установить, что если $u, v \in LA$ и $(u, v) \in \sigma_{LA}$, то $u = v$.

Сначала убедимся в равенстве длин слов u и v, предварительно доказав следующий простой факт.

Лемма 7. *Пусть $pvq \in LA$ и $puq \in LA$, причем $|p| \geqslant 9$ и $|q| \geqslant 9$. Тогда $|v| = |u| + 6k$ при подходящем $k \in \mathbb{Z}$.*

Доказательство. Действительно, по лемме 3 все вхождения слов p и q в A_ω синхронизированы. Значит, все вхождения p заканчиваются, а все вхождения q начинаются с одной и той же позиции блоков одинаковой чётности. Но тогда, вспоминая, что слова v и u находятся между какими-то вхождениями слов p и q, и учитывая, что соседние блоки в A_ω

имеют различную чётность, получаем равенство $|v| - |u| = 6k$ для некоторого $k \in \mathbb{Z}$. Лемма доказана.

Усилим полученный результат. Для этого убедимся в том, что достаточно длинные контексты однозначно определяют длину слов. Отметим, что посвященная этому следующая лемма носит технический характер. Поэтому оценки указаны в том виде, в котором они потребуются в доказательстве.

Лемма 8. *Пусть $pvq \in LA$, $puq \in LA$ и для некоторого $k \in \mathbb{N}_0$ выполнены условия:*

1) $\max\{|v|, |u|\} \leqslant 6 \cdot 2^k$;

2) $\min\{|p|, |q|\} \geqslant 9 \cdot 3^k + \sum\limits_{i=0}^{k} 4 \cdot 3^i$.

Тогда $|v| = |u|$.

Доказательство. Проведем индукцию по k.

База индукции. $k = 0$. Тогда $1 \leqslant |v| \leqslant 6$, $1 \leqslant |u| \leqslant 6$, $|p| \geqslant 9$ и $|q| \geqslant 9$. Но при таких ограничениях согласно лемме 7 $|v| = |u|$. База индукции доказана.

Шаг индукции. Пусть для всех $k < n$ $(n \geqslant 1)$ утверждение леммы справедливо. Покажем его истинность для $k = n$. Итак, пусть $|v|, |u| \leqslant 6 \cdot 2^n$ (в дальнейшем под v и u будем понимать конкретные их вхождения в A_ω в контекстах pvq и puq), а $|p|$ и $|q|$ не меньше $9 \cdot 3^n + \sum\limits_{i=0}^{n} 4 \cdot 3^i$. Тогда $|p|$ и $|q|$ не меньше 9, поскольку $n \geqslant 1$. Значит, в силу леммы 3 все вхождения слов p и q в A_ω синхронны. Это означает, что имеются такие слова α, β, γ, δ длины не более 2, что $p = \delta p_1 \alpha$, $q = \beta q_1 \gamma$, где p_1 и q_1 — целые.

$$\dots\ \star\star\star\ \overbrace{\star\star\star\ \underbrace{\star\cdots\star}_{p_1}}^{p}\ \underbrace{\overbrace{\star\star\star\ \star\cdots\star\ \star\star\star}^{v}}_{v_1}\ \overbrace{\underbrace{\star\cdots\star}_{q_1}\ \star\star\star}^{q}\ \star\star\star\ \dots$$

Рассмотрим слова $v_1 = \alpha v \beta$ и $u_1 = \alpha u \beta$. Легко понять, что в силу выбора α, β, γ и δ вхождения v_1 и v_2 также являются целыми. Следовательно, определены естественные прообразы $p_2 = \varphi^{-1}(p_1)$, $q_2 = \varphi^{-1}(q_1)$, $v_2 = \varphi^{-1}(v_1)$ и $u_2 = \varphi^{-1}(u_1)$. При этом $p_2 v_2 q_2 < A_\omega$, $p_2 u_2 q_2 < A_\omega$ и справедливы оценки

$$|p_2| = \frac{1}{3}|p_1| \geqslant \frac{1}{3}(|p| - 4) \geqslant \frac{1}{3}\Big(9 \cdot 3^n + \sum_{i=1}^{n} 4 \cdot 3^i\Big) = 9 \cdot 3^{n-1} + \sum_{i=0}^{n-1} 4 \cdot 3^i.$$

Аналогично

$$|q_2| \geqslant 9 \cdot 3^{n-1} + \sum_{i=0}^{n-1} 4 \cdot 3^i.$$

С другой стороны, при $n \geqslant 1$ имеем

$$|v_2| = \frac{1}{3}|v_1| \leqslant \frac{1}{3}(|v| + 4) = \frac{1}{3}(6 \cdot 2^n + 4) \leqslant 6 \cdot 2^{n-1}.$$

Аналогично

$$|u_2| \leqslant 6 \cdot 2^{n-1}.$$

Но тогда по индуктивному предположению имеем $|v_2| = |u_2|$. Следовательно,

$$\begin{aligned} |v| = |v_1| - |\alpha| - |\beta| = 3|v_2| - |\alpha| - |\beta| \\ = 3|u_2| - |\alpha| - |\beta| = |u_1| - |\alpha| - |\beta| = |u|. \end{aligned}$$

Шаг индукции, а вместе с ним и лемма доказаны.

Следствие. *Если $v, u \in LA$ и $(v, u) \in \sigma_{LA}$, то $|v| = |u|$.*

Доказательство. Действительно, можно выбрать такие $k \in \mathbb{N}_0$, чтобы выполнялось неравенство $\max\{|v|, |u|\} \leqslant 6 \cdot 2^k$. Рассмотрим произвольное вхождение v в A_ω такое, что слева от v находится не менее $9 \cdot 3^k + \sum_{i=0}^{k} 4 \cdot 3^i$ символов (v найдется согласно лемме 1). Пусть p и q — левый и правый контексты для этого вхождения v, т. е. $pvq < A_\omega$, причем длины слов p и q не меньше $9 \cdot 3^k + \sum_{i=0}^{k} 4 \cdot 3^i$. Такие p и q всегда найдутся, поскольку слева имеется достаточное количество символов в силу выбора вхождения v и A_ω — бесконечное вправо слово. Но тогда по определению синтаксической конгруэнции слово puq также принадлежит языку LA. Отсюда и из леммы 8 следует, что $|v| = |u|$. Следствие доказано.

Следующее утверждение проверяется непосредственно с использованием, например, компьютера, поскольку оно сводится к перебору не очень большого числа случаев.

Лемма 9. *Пусть $u, v \in LA$, $(u, v) \in \sigma_{LA}$ и $|u| = |v| = 9$. Тогда $u = v$.*

Доказательство. Достаточно перебрать все пары слов (u, v), удовлетворяющие условиям $u < A_\omega$, $v < A_\omega$, $u \neq v$ и $|u| = |v| = 9$. Для каждой такой пары (u, v) найдутся (это проверено с помощью компьютера) такие контексты p и q длины не более 5, что точно одно из слов pvq или puq будет лежать в LA. В силу леммы 1 для проверки этого факта достаточно просмотреть w_{10}, поскольку $|pvq|$ и $|puq|$ не превосходят 19.

Таким образом, за конечное число шагов утверждение леммы проверяется. Лемма доказана.

§ 5. Доказательство теоремы 2

Как уже отмечалось выше, достаточно показать, что если $u, v \in LA$ и $(u, v) \in \sigma_{LA}$, то $u = v$. Кроме того, в силу леммы 8 можно считать, что слова u и v имеют одинаковые длины.

Итак, пусть $u, v \in LA$, $(u, v) \in \sigma_{LA}$ и $|u| = |v| = k$. Докажем, что тогда $u = v$. Для этого индукцией по k покажем невозможность соотношения $u \neq v$.

В базе индукции рассмотрим следующие случаи:

1) $k \leqslant 25$. В этом случае существуют такие слова α и β, что $|a| = \mathrm{lsp}(u)$, $|b| = \mathrm{rsp}(u)$, $|\alpha u \beta| = 27$ и все вхождения слова $\alpha u \beta$ в A_ω являются целыми. Тогда из определения синтаксической конгруэнции имеем $(\alpha u \beta, \alpha v \beta) \in \sigma_{LA}$;

2) $k = 26$ и существует такая буква a, что все вхождения либо au, либо ua являются целыми. Снова либо $(au, av) \in \sigma_{LA}$, либо $(ua, va) \in \sigma_{LA}$ соответственно;

3) $k = 27$ и все вхождения слова u являются целыми.

Все три случая влекут наличие пары слов $(u, v) \in \sigma_{LA}$ такой, что $u \neq v$, $|u| = |v| = 27$ и все вхождения u в A_ω являются целыми.

Несложно понять, что этим же свойством обязаны обладать все вхождения v в A_ω. Действительно, для какого-нибудь вхождения слова u достаточно выбрать контексты длины большей 9. Тогда в этом контексте слово v должно входить в A_ω, а контексты обладают свойством синхронизации. По этим же соображениям слова u и v синхронизированы.

Но это означает, что для u и v определены естественные прообразы $u' = \varphi^{-1}(u)$ и $v' = \varphi^{-1}(u)$, причем по лемме 4 имеем $\varphi(u') = u$ и $\varphi(v') = v$. Теперь если предположить, что $u \neq v$, то $u' \neq v'$. Поскольку длины слов u' и v' равны 9, то в силу леммы 9 имеем $(u', v') \notin \sigma_{LA}$. Следовательно, по определению синтаксической конгруэнции существуют такие слова p' и q', что точно одно из слов $p'u'q'$ и $p'v'q'$ лежит в LA. Без ограничения общности будем считать, что $p'u'q' \in LA$ и $p'v'q' \notin LA$.

По лемме 4 для $p'u'q'$ определен естественный образ $\varphi(p'u'q') = puq$, причем $puq \in LA$. Здесь p и q — подходящие образы слов p' и q' относительно подстановки (1). Покажем, что $pvq \notin LA$. Действительно, предположим обратное, т. е. $pvq \in LA$. Тогда вследствие синхронизации слов u и v, а также в силу того, что по построению слова p и q являются конкатенациями блоков, определен естественный прообраз $\varphi^{-1}(pvq)$. При этом для этих конкретных вхождений слов p и q справедливы равенства $\varphi^{-1}(p) = p'$ и $\varphi^{-1}(q) = q'$. Следовательно, $p'v'q' \in LA$, что противоречит выбору p' и q'.

Но это означает наличие таких контекстов p и q, что $puq \in LA$, а $pvq \notin LA$. Поэтому $(u, v) \notin \sigma_{LA}$. Противоречие.

Итак, неравенство $u \neq v$ влечет противоречие. Значит, база индукции доказана.

Шаг индукции. Рассмотрим оставшиеся случаи $k = 26$, $k = 27$ и $k > 27$. Рассуждения будут одинаковыми.

Пусть $(u, v) \in \sigma_{LA}$, $|u| = |v| = k = n$. Предположим, что для $k \leqslant n$ утверждение доказано.

Покажем, что в этом случае равенство $u = v$ является единственной возможностью. Предположим, что $u \neq v$. Повторяя рассуждения, аналогичные рассуждениям из доказательства базы индукции, можно показать, что в этом случае существуют такие слова s и t ($|s| = \mathrm{lsp}(u)$, $|t| = \mathrm{rsp}(u)$), что $(sut, svt) \in \sigma_{LA}$ и слова sut и svt являются целыми. Отметим, что $|sut| = |svt| > 27$ в силу условий шага индукции. Так как $|sut| = |svt| > 9$, то определены естественные прообразы $u' = \varphi^{-1}(sut)$ и $v' = \varphi^{-1}(svt)$, причем $u' \neq v'$ (так как по предположению $u \neq v$), и справедливы соотношения

$$|u'| = |v'| = \frac{1}{3}|svt| > 9,$$

$$|u'| = |v'| = \frac{1}{3}|svt| \leqslant \left(\frac{k}{3} + \frac{4}{3}\right) \leqslant k - 1 \quad \text{для} \quad k \geqslant 4.$$

Следовательно, в силу индуктивного предположения $(u', v') \notin \sigma_{LA}$. Значит, аналогично доказательству базы индукции можно считать, что существуют такие контексты p' и q', что $p'u'q' \in LA$ и $p'v'q' \notin LA$. Но тогда, как и выше, определен естественный образ $\varphi(p'u'q') = psutq$ и вновь $psvtq \notin LA$ (доказательство в точности такое же). То есть $(psutq, psvtq) \notin \sigma_{LA}$, что противоречит выбору u и v.

Следовательно, если $(u, v) \in \sigma_{LA}$, то $u = v$. Шаг индукции обоснован. Теорема 2 доказана.

В заключение выражаем благодарность участникам семинара «Дискретная математика» в Уральском университете за полезные обсуждения. Мы также благодарны рецензенту, замечания которого способствовали улучшению текста работы.

ЛИТЕРАТУРА

1. **Адян С. И.** Проблема Бернсайда и тождества в группах. М.: Наука, 1985.
2. **Аршон С. Е.** Доказательство существования n-значных бесконечных асимметричных последовательностей // Мат. сб. 1937. Т. 2 (44), № 4. С. 769–779.

3. **Евдокимов А. А.** Полные множества слов и их числовые характеристики // Методы дискретного анализа в исследовании экстремальных структур: Сб. науч. тр. Новосибирск: Ин-т математики СО АН СССР, 1983. Вып. 39. С. 7–19.
4. **Евдокимов А. А., Крайнев В. А.** Задачи о полноте множеств слов // XXII обл. науч.-техн. конф. о-ва им. А. С. Попова. Тез. докл. Новосибирск, 1979. С. 105–107.
5. **Зимин А. И.** Блокирующие множества термов // Мат. сб. 1982. Т. 119, № 3. С. 363–375.
6. **Лаллеман Ж.** Полугруппы и комбинаторные приложения. М.: Мир, 1985.
7. **Саломаа А.** Жемчужины теории формальных языков. М.: Мир, 1986.
8. **Bean D. R., Ehrenfeucht A., McNulty G. F.** Avoidable patterns in strings of symbols // Pacific J. Math. 1979. V. 85, N 2. P. 261–294.
9. **Choffrut C., Karhumäki J.** Combinatorics of words // Handbook of formal languages. V. 1: Word, language, grammar. Berlin: Springer, 1997. P. 329–438.
10. **Crochemore M., Goralcik P.** Mutually avoiding ternary words of small exponent // Intern. J. Algebra Comput. 1991. V. 1, N 4. P. 407–410.
11. **Dejean F.** Sur un théorème de Thue // J. Comb. Theory. Ser. A. 1972. V. 13, N 1. P. 90–99.
12. **Pin J.-E.** Syntactic semigroups // Handbook of formal languages. V. 1: Word, language, grammar. Berlin: Springer, 1997. P. 679–746.

Адрес авторов:

Уральский государственный
университет,
мат.-мех. факультет,
кафедра алгебры
и дискретной математики,
пр. Ленина, 51,
620083 Екатеринбург, Россия.
E-mail:
evgeny.sukhanov@usu.ru,
alklepin@mail.utnet.ru

Статья поступила
6 мая 1998 г.