На правах рукописи

Петрова Елена Александровна

О комбинаторных свойствах бесповторных языков

01.01.09 — дискретная математика и математическая кибернетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Екатеринбург 2016 г.

Работа выполнена на кафедре алгебры и дискретной математики Института математики и компьютерных наук Уральского федерального университета имени первого Президента России Б. Н. Ельцина, г. Екатеринбург.

Научный руководитель: доктор физико-математических наук, профессор Шур Арсений Михайлович

Официальные оппоненты: Протасов Владимир Юрьевич, доктор физико-математических наук, профессор механико-математического факультета МГУ им М. В. Ломоносова, г. Москва.

Куликов Александр Сергеевич, кандидат физико-математических наук, научный сотрудник ФГБУН «Санкт-Петербургское отделение Математического института им. В. А. Стеклова РАН» г. Санкт-Петербург.

Ведущая организация: ФГБУН «Институт математики им. С. Л. Соболева СО РАН», г. Новосибирск.

Защита диссертации состоится «25» мая 2016 г. в 16:00 на заседании диссертационного совета Д 002.202.02 на базе ФГБУН «Санкт-Петербургское отделение Математического института им. В. А. Стеклова РАН» (191023, г. Санкт-Петербург, наб. реки Фонтанки, 27).

С диссертацией можно ознакомиться в библиотеке и на сайте ФБГУН «Санкт-Петербургское отделение Математического института им. В. А. Стеклова РАН», http://www.pdmi.ras.ru/

Автореферат разослан «   » ________ 2016 г.

Ученый секретарь
диссертационного совета
доктор физ.-мат. наук Малютин Андрей Валерьевич

Общая характеристика работы

Актуальность темы. Комбинаторика слов, относительно молодая и активно развивающаяся дисциплина, возникшая на рубеже XIX и XX веков, представляет собой важное направление на стыке современной дискретной математики и теоретических компьютерных наук, изучающее свойства символьных последовательностей (слов). Методы и результаты, разработанные и полученные в комбинаторике слов, широко применимы во многих областях, например, в теории групп, теории кодирования, теории игр, биоинформатике, сжатии данных.

Одним из основных объектов исследований в комбинаторике слов являются *бесповторные языки* — множества слов, не содержащих внутри себя повторяющихся фрагментов, или, как говорят, *избегающих* определённые структурные элементы. Норвежский математик Аксель Туэ в начале XX века занимался конструированием и изучением нескольких конкретных бесповторных языков, и принято считать, что именно с этих работ комбинаторика слов берёт своё начало как самостоятельная дисциплина. Туэ рассматривал, например, бескубный язык над двухбуквенным (бинарным) алфавитом и бесквадратный язык над трёхбуквенным (тернарным) алфавитом — эти языки состоят из слов, не содержащих трёх и двух идущих подряд одинаковых фрагментов, соответственно, — и получил ряд нетривиальных результатов, касающихся, в первую очередь, вопроса о конечности или бесконечности того или иного языка. Со времён Туэ тема бесповторных языков получила весьма широкое развитие во многих направлениях: было обобщено понятие степени, изучены алфавиты большей мощности, обобщено понятие избегаемости с простых повторов до шаблонов, абелевых степеней и т.д., получены результаты о комбинаторной сложности бесповторных языков, рассмотрены бесповторные слова с дополнительными ограничениями. Тем не менее, в отношении языков, рассмотренных Туэ, до сих пор остаётся множество открытых проблем. Некоторые из них решены в данной диссертации.

Приведём определения, необходимые для дальнейшего изложения. *Алфавитом* называется непустое конечное или счётное множество, элементы которого — *буквы*. Произвольная последовательность символов над алфавитом называется *словом*. Мы рассматриваем конечные и бесконечные слова и языки над основными алфавитами $\Sigma_2 = \{a, b\}$ (бинарный), $\Sigma_3 = \{a, b, c\}$ (тернарный) и некоторыми вспомогательными алфавитами. Слово, не содержащее букв, или *пустое* слово, обозначается как λ. Множество всех слов над алфавитом Σ вместе с λ образует свободный моноид, обозначаемый как Σ^*. Произвольное подмножество Σ^* называется *языком* над Σ.

Длина слова w — это количество содержащихся в нём букв, обозначается как $|w|$. Слово v называется *подсловом* w (обозначается $v \leqslant w$), если $w = pvs$ для некоторых слов p, s, причём, v является префиксом [суффиксом] w, если $p = \lambda$ [$s = \lambda$, соответственно].

Натуральное число $p \leqslant |w|$ называется *периодом* слова w, если $w[i] = w[i+p]$ для всех $i \in \{1, \dots |w| - p\}$. *Экспонента* слова w (обозначается $\exp(w)$) — это отношение $|w|$ к минимальному периоду w. *Локальной экспонентой* слова будем называть число $\operatorname{lexp}(w) = \sup\{\exp(v) | v \leqslant w\}$. Локальная экспонента характеризует степень повторяемости фрагментов внутри слова. Локальные экспоненты бесконечных слов также называют *критическими экспонентами*; в этом случае супремум в определении может не достигаться. Слова экспоненты 2 называются *квадратами*, экспоненты 3 — *кубами*.

Слово w *β-свободно* [β^+-свободно], если $\operatorname{lexp}(w) < \beta$ [соответственно, $\operatorname{lexp}(w) \leqslant \beta$]. Говорят также, что слово *избегает* степень β [соответственно, β^+]. Язык называется *бесповторным*, если он состоит из слов, избегающих некоторую фиксированную экспоненту β или β^+. Экспонента β называется k-избегаемой [k-неизбежной], если существует ω-слово над k-буквенным алфавитом, избегающее β [соответственно, множество слов над k-буквенным алфавитом, избегающих β, конечно]. Основные бесповторные языки, изучаемые в данной диссертации:

- избегающий экспоненту 2 над тернарным алфавитом, или тернарный бесквадратный язык (обозначается как SF, от англ. “square-free“);

- избегающий экспоненту 3 над бинарным алфавитом, или бинарный бескубный язык (обозначается как CF, от англ. “cube-free“);

В исследованиях, связанных с бесповторными языками, можно выделить несколько основных направлений.

1. Бесповторные морфизмы. Морфизм $f : \Sigma^+ \to \Sigma^+$ *избегает экспоненту* β, если из того, что $\operatorname{lexp}(u) < \beta$ следует, что $\operatorname{lexp}(f(u)) < \beta$ для любого слова u. Морфизмы, избегающие экспоненты, называют бесповторными. Такие морфизмы являются самым эффективным способом порождения бесконечных бесповторных последовательностей. Построением бесповторных морфизмов, в частности, занимался и Туэ в своих работах. Он доказал, что морфизм Туэ–Морса является сильно бескубным и, более того, любой сильно бескубный морфизм построен из блоков Туэ–Морса [51].

В [50] Туэ построил бесквадратный морфизм для четырёхбуквенного алфавита и затем нашёл условия, при которых произвольный морфизм для четырёхбуквенного алфавита является бесквадратным. В [51] он сделал это для тернарного алфавита.

В настоящее время, наилучшая с точки зрения вычислительной сложности характеризация бесквадратных морфизмов над произвольными алфавитами получена Крошмором в [13].

Для автоматизированной проверки морфизмов на бесповторность удобно иметь некоторый конечный набор слов такой, что если образы всех слов набора не содержат подслов запрещённой степени, то можно заключить, что образ любого слова под действием морфизма не содержит таких подслов. Такие наборы называются *тестовыми множествами* морфизмов. Характеризация тестовых множеств бинарных бескубных морфизмов дана Ришомом и Влазинским в [37]. В этой же работе доказано, что для k-свободных морфизмов над алфавитами мощности 3 и более, где $k \geqslant 3$, не существует конечных тестовых множеств. Мы используем результаты из [37] для доказательства результатов первой главы диссертации.

2. Граница повторяемости. Граница повторяемости $\mathsf{RT}(k)$ определяется как инфимум множества экспонент β таких, что существует бесконечное слово над k-буквенным алфавитом, не содержащее подслов экспоненты β. Ясно, что в однобуквенном алфавите никакая экспонента не является избегаемой. Туэ показал, что $\mathsf{RT}(2) = 2$. В 1972 году Ф. Дежан выдвинула гипотезу, что таблица значений функции $\mathsf{RT}(k)$ выглядит следующим образом [18]:

k	1	2	3	4	5	$\ldots$	n	$\ldots$
$\mathsf{RT}(k)$	∞	2	$7/4$	$7/5$	$5/4$	$\ldots$	$n/(n-1)$	$\ldots$

и доказала её для $k = 3$. На полное же доказательство гипотезы потребовались усилия многих исследователей в течение почти сорока лет. Подход, использованный Туэ и Дежан для бинарного и тернарного алфавитов, принципиально не годится для алфавитов большей мощности([9]).

Выход из этой ситуации нашёл Пансьё, предложив идею бинарного кодирования слов над k-буквенным алфавитом, в которых любые $(k-1)$

подряд идущие буквы различны [32]. Сам Пансьё доказал гипотезу Дежан для четырёхбуквенного алфавита, а кроме того, свёл задачу нахождения k-буквенного граничного ω-слова к нахождению бинарного кода, соответствующего $\mathsf{RT}(k)^+$-свободному ω-слову. Такие коды могут быть построены методом итерации морфизмов.

На коды Пансьё опираются все дальнейшие работы по доказательству гипотезы Дежан, см. [10, 15, 16, 29, 30, 34].

За долгое время доказательства гипотеза Дежан породила множество задач, среди которых существуют усиления гипотезы (например, исследование конечной границы повторяемости [4–6] и экспоненциальная гипотеза [25, 31, 52]), обобщения гипотезы [19, 22], исследование границы повторяемости для других видов степеней [23, 28, 38] и других видов слов [1, 2, 47].

В диссертации активно используются коды Пансьё, а также слова Туэ–Морса и Аршона, являющиеся граничными для бинарного и тернарного алфавитов, соответственно.

3. Комбинаторная сложность. Для оценки количества слов языка используется понятие *комбинаторной сложности* языка. Она определяется как функция $C_L(n) : \mathbb{N} \to \mathbb{N}$, значение которой равно количеству слов длины n в языке L. Комбинаторной сложности как естественной характеристике языка посвящено большое количество работ, особенно в области бесповторных языков; см. обзор [49].

Скорость роста комбинаторной сложности характеризуется *индексом роста* языка $\alpha(L)$, который определяется следующим образом:

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

при этом в случае факториального языка верхний предел превращается просто в предел. Индекс роста, больший 1, говорит об экспоненциальном росте количества слов в языке, $\alpha(L) = 1$ соответствует субэкспоненциальному росту, а индекс роста конечного языка равен нулю.

Рестиво и Салеми в [35] установили, что язык OF бинарных слов, избегающих экспоненту 2^+, имеет полиномиальную комбинаторную сложность. Кассень [12] доказал, что функция C_{OF} не может быть выражена в виде полинома, т.е. что числа $\alpha = \sup\{r \mid \exists C > 0 \quad \forall n \quad C_{\mathsf{OF}} \geqslant Cn^r\}$ и $\beta = \inf\{r \mid \exists C > 0 \quad \forall n \quad C_{\mathsf{OF}} \leqslant Cn^r\}$ не равны друг другу, и построил верхнюю оценку для α и нижнюю для β, получив в итоге $1.155 < \alpha < 1.276 < 1.332 < \beta < 1.587$. Точные значения $\alpha = 1.27355\ldots$ и $\beta = 1.33224\ldots$ были недавно вычислены Гулиельми и Протасовым [21] в качестве приложения разработанного ими метода точного вычисления совместных спектральных характеристик конечных семейств матриц.

В работе, посвящённой бесповторным морфизмам [9], Бранденбург показал, в частности, что языки CF и SF имеют экспоненциальную комбинаторную сложность и получил некоторые оценки индексов роста этих языков. Наилучшие на данный момент оценки $1.457573 \leqslant \alpha(\mathsf{CF}) \leqslant 1.457577$, $1.301759 < \alpha(\mathsf{SF}) < 1.301762$ получены с помощью алгоритмов из [24, 26, 42, 43, 45].

В [44] Шуром получены асимптотические формулы для индексов роста $\alpha(k, \beta)$ β-свободных языков над произвольным k-буквенным алфавитом, где $\beta \geqslant 2$.

Значение $\alpha(\mathsf{SF})$ используется для оценки результатов третьей главы диссертации.

4. Бесповторные слова с дополнительными ограничениями. Плодотворной темой в исследовании бесповторных языков является построение и изучение бесповторных слов с дополнительными ограничениями. Как правило, дополнительное ограничение заключается в том, что помимо повторов в конструируемом бесконечном слове требуется избежать ещё какие-то подслова определённого вида либо ограничить количество их вхождений в слово.

Первым этой темой в комбинаторике слов занимался, опять же, Туэ. В [51] он не только доказал, что квадраты избегаемы над тернарным алфавитом, но и построил бесквадратные слова с дополнительными ограничениями. Например, он показал, что такое слово может не содержать максимум двух подслов длины 3 и перечислил все пары таких подслов. Доказательству близкого по духу результата посвящена вторая глава диссертации.

Много работ посвящено избегаемости шаблонов в бинарных словах. В [20] доказано, что для любого 2-избегаемого бинарного шаблона p язык, избегающий p, имеет экспоненциальную комбинаторную сложность. Классификацию неизбежных бинарных шаблонов завершил Кассень в [11]. В [27] рассмотрены шаблоны, избегаемые в бинарных бескубных словах, и дано их полное описание.

В [33] производится построение бескубного слова, избегающего квадраты с периодами $\leqslant 4$. Позднее возник естественный вопрос: можно ли в $RT(k)^+$-свободных бесконечных словах над k-буквенным алфавитом обойтись конечным числом подслов экспоненты $RT(k)$? В серии работ [5–7] вводится и исследуется понятие *конечной границы повто-*

ряемости $FRT(k)$ для k-буквенного алфавита. Это наименьшее рациональное число такое, что существует бесконечное слово w над k-буквенным алфавитом, локальная экспонента которого не превышает $FRT(k)$ и w содержит лишь конечное число подслов экспоненты $RT(k)$. В этих терминах из результата, полученного в [39], следует, что $FRT(2) = 7/3$. В [5] доказано, что $FRT(3) = RT(3) = 7/4$, а также выдвинута гипотеза $FRT(4) = 7/5$. Наконец, в [7] было завершено исследование конечной границы повторяемости доказательством того, что $FRT(k) = RT(k), k \geqslant 4$. Также в [52] Тунев и Шур доказали, что для всех нечётных $k \leqslant 101$ число k-арных граничных слов, содержащих только однобуквенные повторы, эспоненциально.

5. Структура частично упорядоченного множества слов бесповторного языка. На множестве слов бесповторного языка можно естественным образом ввести отношения префиксного, суффиксного и подсловного порядков. Изучение структуры ч.у.м. слов бесповторных языков относительно этих порядков формирует широкий круг задач, решение которых проливает свет на внутреннее устройство бесповторного языка. Наибольший интерес представляет исследование префиксного (или суффиксного) дерева языка, которому посвящён ряд работ. В [8] показано, что в дереве любого бесповторного языка каждое поддерево имеет хотя бы один лист. В [14] Карри доказал, что для некоторых k-свободных языков над алфавитом Σ разрешим вопрос о конечности поддерева, порождённого данным словом, кроме того, деревья этих языков не содержат бесконечных изолированных ветвей, т.е. любая бесконечная ветвь ветвится бесконечно часто. Позднее Карри и Шелтон [17] распространили этот результат на все остальные k-свободные языки.

В [17] Карри и Шелтон доказали, что при любом k деревья k-свободных языков не содержат бесконечных изолированных ветвей. Так же они доказали (неконструктивно), что вопрос о конечности заданного порожденного поддерева разрешим.

Остальные известные результаты касаются конкретных языков. Самую простую структуру с точки зрения дерева имеет язык OF благодаря своей полиномиальной сложности и тесной связи с языком слов Туэ-Морса. Рестиво и Салеми в [36] предложили алгоритм проверки конечности поддерева, порождённого узлом в дереве языка OF, а также доказали существование в этом дереве поддеревьев произвольной конечной высоты. В [48] для этого же языка Шур показал разрешимость за линейное время вопроса об изоморфизме поддеревьев, порождённых двумя данными узлами и, кроме того, представил другой алгоритм проверки конечности поддерева, который к тому же вычисляет высоту поддерева в конечном случае.

В работах [40, 41] Шелтон и Сони доказали, что если в префиксном дереве языка SF все бесконечные пути из узла w проходят через один и тот же узел wv, то $|v| \leqslant C|w|^{2/3}$, где C — некоторая абсолютная константа.

Иными словами, здесь получена верхняя субполиномиальная оценка на размер (глубину) конечного поддерева, порождённого узлом в дереве SF. Естественно выдвинуть гипотезу о том, что на самом деле справедлива логарифмическая оценка. В пользу этого предположения говорит, например, экспоненциальная сложность языка. Ручная проверка для поддеревьев небольшого размера показывает, что можно выдвинуть ещё более сильную гипотезу: конечное поддерево, порождённое словом длины n, имеет $O(\log n)$ узлов. Однако улучшением оценки Шелтона и Сони ввиду сложности задачи с тех пор никто не занимался. Решению описанной проблемы посвящена третья глава диссертации.

Рестиво и Салеми в [36], помимо вопросов о конечности порождённого данным узлом поддерева и существовании конечных поддеревьев произвольной высоты, сформулировали следующую задачу: пусть даны два слова $u, v \in L$; существует ли такое слово w, что $uwv \in L$, и как, в случае положительного ответа, построить это слово? Эта задача, очевидно, тоже имеет связь со структурой ч.у.м. языка L. Авторы сформулировали все эти проблемы для языков OF и SF и привели решения для OF. Решение вопроса о конечности порождённого поддерева для SF следует из результатов [41]. Остальные проблемы, сформулированные Рестиво и Салеми, для языка SF до сих пор оставались открытыми. Открытая проблема о существовании конечных поддеревьев произвольной высоты в дереве SF также фигурирует в книге Аллуша и Шаллита [3]. Решение этой проблемы для языков CF и SF описано в первой главе диссертации.

За последние годы опубликовано значительное количество работ, посвященных решению подобных задач. Результаты, связанные с бесповторными последовательностями, регулярно докладываются на международных конференциях по дискретной математике и компьютерным наукам. Существует и специализированная международная конференция WORDS по комбинаторике слов.

Методы исследования. Для доказательства результатов, представленных в диссертации, в основном используются методы комбинаторики слов, основанные на свойствах периодических слов и граничных слов над бинарным и тернарным алфавитами (слова Туэ–Морса и Аршона, соответственно), слов Фибоначчи, кодировании бесповторных слов бинарными кодами Пансьё и «маршрутными кодами» (метод маршрутных кодов предложен Шуром в [46]). Также для доказательства одного из вспомогательных результатов используется метод решения одной из модификаций задачи составления расписаний.

Цели и задачи. Основная цель диссертации — развитие теории бесповторных языков в следующих направлениях: структурный анализ префиксных/суффиксных деревьев и подсловных ациклических графов, образованных множеством слов бесповторного языка; избегаемость буквенных шаблонов в бесповторных языках.

Основные задачи, рассматриваемые в диссертации:

1. конструктивно доказать существование поддеревьев произвольной конечной высоты в суффиксных деревьях языков бинарных бескубных слов и тернарных бесквадратных слов;

2. доказать логарифмическую от длины слова оценку длины его фиксированного контекста и найти нижнюю границу частоты ветвления префиксного (суффиксного) дерева тернарного бесквадратного языка;

3. классифицировать буквенные шаблоны по их избегаемости в тернарном бесквадратном языке.

Научная новизна. Все результаты диссертации являются новыми.

Теоретическая и практическая ценность. Диссертация носит теоретический характер. Полученные результаты могут быть использованы в дальнейших исследованиях по комбинаторике слов и при чтении специальных курсов для студентов математических специальностей.

Апробация результатов работы. Результаты диссертации были представлены на 8-й и 10-й международных конференциях по комбинаторике слов WORDS2011 и WORDS2015 (Прага, 2011; Киль, 2015), 1-м и 2-м российско-финских симпозиумах по дискретной математике (Санкт-Петербург, 2011; Турку, 2012), 37-й международной конференции по математическим основам компьютерных наук MFCS 2012 (Братислава, 2012), а также на семинарах «Алгебраические системы» и «Дискретная математика» (УрГУ/УрФУ, 2010-2015).

Публикации. Основные результаты диссертации опубликованы в работах [53–59], среди которых четыре работы [53–56] опубликованы в рецензируемых изданиях из списка, рекомендованного ВАК. В совместных работах [54–56, 58] А.М. Шуру принадлежат постановка задачи, общая методика исследований и оптимизация некоторых предложенных автором конструкций. Доказательства всех основных утверждений принадлежат автору. Работы [57, 59] представляют собой тезисы статей [55, 58].

Структура диссертации. Диссертация состоит из введения, трёх глав, заключения и списка литературы. Объём диссертации составляет 89 страниц, библиография включает 81 наименование.

Краткое содержание работы

Первая глава посвящена построению серий слов, порождающих конечные поддеревья произвольной высоты в суффиксных деревьях языков CF и SF. Глава состоит из трёх разделов.

Пусть L — язык над Σ и $w \in L$. Слово $u \in \Sigma^*$ такое, что $uw \in L$, называется *левым контекстом* w в L. Слово w называется *максимальным слева* [*предмаксимальным слева*], если оно не имеет левых контекстов [соответственно, имеет конечное число левых контекстов]. *Уровнем* предмаксимального слева слова w называется длина его самого длинного левого контекста (или высота поддерева, порождённого словом w в дереве суффиксного порядка на L). Понятия предмаксимального справа и максимального справа слова определяются симметрично. Говорят, что слово *предмаксимально*, если оно предмаксимально и слева, и справа. *Уровнем* предмаксимального слова w называется пара $(n, k) \in \mathbb{N}_0^2$ такая, что n и k – длины самого длинного левого и самого длинного правого контекстов w. Основные результаты первой главы сформулированы в следующих двух теоремах:

Теорема 1 ([54]). *В языке* CF *существуют*
а) предмаксимальные слева слова любого уровня $n \in \mathbb{N}_0$*;*
б) предмаксимальные слова любого уровня $(n, k) \in \mathbb{N}_0^2$*.*

Теорема 2 ([55]). *В языке* SF *существуют*
а) предмаксимальные слева слова любого уровня $n \in \mathbb{N}_0$*;*
б) предмаксимальные слова любого уровня $(n, k) \in \mathbb{N}_0^2$*.*

Раздел 1.1 содержит общее описание основной конструкции для построения односторонних предмаксимальных слов в языках CF и SF. Построение производится в два этапа. Главная идея первого этапа заключается в итеративном построении префиксной последовательности слов $\{w_n\}$ таких, что слово w_n имеет фиксированный левый контекст длины не менее n, путём запрета появления «нежелательных» букв. Для того, чтобы каждое w_n было соответственно бескубным или бесквадратным, необходимо при его построении из слов w_{n-1} вставлять в нужные места определённые «буферные» слова. Второй этап достройки слова с фиксированным контекстом до предмаксимального слева слова нужного уровня идейно не отличается от первого.

Раздел 1.2 посвящён предмаксимальным словам в бинарном бескубном языке. В подразделе 1.2.1 описано построение последовательности буферных слов, основанное на свойствах слов Туэ–Морса и бескубных морфизмах. В подразделе 1.2.2 производится доказательство бескубности построенных односторонних слов, то есть утверждения а) Теоремы 1. Наконец, в подразделе 1.2.3 доказана вторая часть Теоремы 1.

В разделе 1.3 идёт речь о предмаксимальных словах в тернарном бесквадратном языке. Подраздел 1.3.1 отведён под определение кодов Пансьё, основанных на них маршрутных кодов тернарных бесквадратных слов и описание их свойств, на которых основана конструкция предмаксимальных слов для SF. На самом деле строятся не сами слова, а их маршрутные коды. В подразделе 1.3.2 описаны особенности основной конструкции в связи с применением этих маршрутных кодов. В подразделе 1.3.3 вычислен маршрутный код ω-слова Аршона, реверс которого используется в качестве левого контекста конструируемых слов. Свойства запрещаемых на каждом шаге построения левых контекстов обсуждаются в подразделе 1.3.4. В подразделе 1.3.5 описано построение буферных слов и доказана корректность конструкции для слов с фиксированным левым контекстом. Наконец, в подразделе 1.3.6 производится доказательство Теоремы 2.

Во **второй** главе идёт речь о тернарных бесквадратных буквенных шаблонах длины 5 и 6 в языке тернарных бесквадратных слов (из результатов Туэ [51] и свойств бесквадратных кодов Пансьё следует, что все шаблоны меньших длин неизбежны в SF). Будем говорить, что слово w над k-арным алфавитом избегает *буквенный шаблон* u, если u — слово над k-арным *алфавитом переменных* Σ_k и w не содержит подслов вида $h(u)$, где для любого $x \in \Sigma_k$ $|h(x)| = 1$ и для любых

$x, y \in \Sigma_k \quad (x \neq y) \Rightarrow (h(x) \neq h(y))$. Например для алфавита переменных $\{x, y, z\}$ слово $w \in \Sigma_3$ избегает шаблон xyz, если оно не содержит следующих подслов: $abc, bca, cab, cba, acb, bac$. Основной результат второй главы —

Теорема 3 ([53]). *Следующие тернарные бесквадратные буквенные шаблоны избегаемы в языке* SF*:*
а) $xyxzx$, $xyzxy$;
б) $xyxzyz$ и все шаблоны длины 6, подсловами которых являются шаблоны из а).
Все другие шаблоны длины $\leqslant 6$ неизбежны в SF.

Доказательство того, что остальные 5- и 6-буквенные неизбежны, основано на свойствах кодов Пансьё тернарных бесквадратных слов. Доказательство Теоремы 3 опирается на маршрутные коды, описанные в подразделе 1.3.1. При ближайшем рассмотрении кодов Пансьё и маршрутных кодов шаблонов, фигурирующих в утверждении Теоремы 3, оказывается, что существование тернарных бесквадратных слов, избегающих эти шаблоны, равносильно существованию бесконечных бесквадратных маршрутных кодов с определённым свойством, а именно, кодов, построенных с использованием только двух букв из трёх возможных букв алфавита маршрутных кодов. Таким образом, для доказательства Теоремы 3 необходимо построить бинарные слова, удовлетворяющие некоторым условиям, причём условия одинаковы в случае всех трёх шаблонов. На роль такого бинарного слова подходит ω-слово Фибоначчи. Доказательство того, что на основе этого слова можно построить необходимые маршрутные коды, приведено в подразделе 2.1. Совмещая это с предварительными рассуждениями о неизбежных буквенных шаблонах, получаем утверждение Теоремы 3. В следующем подразделе 2.2 доказана теорема о локальных экспонентах тернарных бесквадратных слов, избегающих буквенные шаблоны длины $\leqslant 6$:

Теорема 4 ([53]). *Минимальная критическая экспонента тернарного бесквадратного слова, избегающего буквенный шаблон длины $\leqslant 6$*
(1) равна $15/8$ для шаблона $xyxzx$,
(2) равна $11/6$ для шаблона $xyzxy$, и
(3) не превосходит $1+\rho/2$ для шаблона $xyxzyz$, где ρ — золотое сечение.

В подразделе 2.3 обсуждается дальнейшая перспектива развития этой темы для буквенных шаблонов бóльших длин.

Третья глава посвящена исследованию структуры префиксного дерева тернарного бесквадратного языка и нахождению некоторых её комбинаторных характеристик. В случае, когда L — факториальный язык, *фиксированным левым контекстом* слова w называется такое слово v, что поддерево, порождённое узлом с меткой w в дереве суффиксного порядка, содержит узел vw и на пути от w к wv все узлы имеют ровно одного потомка. Основными результатами являются следующие утверждения:

Теорема 5 ([56]). *Любой фиксированный контекст тернарного бесквадратного слова w имеет длину $O(\log |w|)$.*

Теорема 6 ([56]). *Пусть $w \in \mathsf{SF}$ — ω-слово и $b(n)$ — количество ветвящихся узлов на пути $(\lambda, w[1..n])$ в дереве префиксного порядка языка SF. Тогда $\liminf_{n\to\infty} \frac{b(n)}{n} \geqslant \frac{2}{9}$.*

На основании этих теорем можно сделать вывод, что рассматриваемое префиксное дерево ветвится достаточно равномерно. Индекс роста языка SF говорит о том, что узел в префиксном дереве в среднем имеет приблизительно 1.3 сыновей; Теорема 6 утверждает, что вдоль любого бесконечного пути среднее количество сыновей не меньше $11/9$, а Теорема 5 — что любые два соседних ветвящихся узла на пути расположены близко. При этом результат Теоремы 5 асимптотически неулучшаем: примеры слов из SF с логарифмической длиной фиксированного контекста построены в Главе 1.

Первый раздел 3.1 отведён под доказательство Теоремы 5, которая гласит, что длина максимального начального неветвящегося сегмента в поддереве, порождённом узлом w, ограничена $C\log(|w|)$, где C — некоторая константа. Пусть $w \in \mathsf{SF}$ и v — некоторый фиксированный правый контекст w. Если заменить любую букву в суффиксе длины $|v|$ слова wv на отличную от неё и от предыдущей буквы, получится квадрат. Эти квадраты условно разделяются на две группы: короткие (с периодом $\leqslant 17$) и длинные (с периодом > 17), и для каждой группы по отдельности оценивается количество букв, фиксированных ими описанным выше образом. Подраздел 3.1.1 содержит доказательство оценки для коротких квадратов: не более $\frac{2}{3}l + O(1)$ позиций слова длины l фиксированы короткими квадратами.

В подразделе 3.1.2 рассмотрены длинные квадраты. Доказывается верхняя оценка $l/9 + O(\log n)$ на количество позиций, фиксированных длинными квадратами в фиксированном контексте длины l слова длины n, причём показано, что $l \leqslant n$.

Объединяя результаты, полученные для коротких и длинных квадратов, получаем утверждение Теоремы 5.

В разделе 3.2 приведено доказательство Теоремы 6. Это следствие из доказательства Теоремы 5, где при суммировании долей букв, фиксированных короткими и длинными квадратами, остаётся «зазор»: грубо говоря, 2/3 букв фиксировано короткими квадратами и 1/9 — длинными. Отсюда имеем нижнюю оценку на количество ветвящихся узлов в дереве тернарного бесквадратного языка — 2/9.

Наконец, в разделе 3.3 обсуждается возможное усиление Теоремы 5.

В заключении кратко подведятся итоги работы отдельно по каждой главе диссертации и обсуждаются пути дальнейших исследований.

Положения, выносимые на защиту. Основными результатами диссертации являются решения сформулированных выше задач 1-3.

1. Доказано существование конечных поддеревьев произвольной высоты в суффиксном и префиксном деревьях, а также конечных подграфов специального вида в ациклическом подсловном графе языков бинарных бескубных слов CF и тернарных бесквадратных слов SF. Доказательство произведено конструктивным образом путём построения соответствующих серий предмаксимальных слева (справа) слов для любого натурального уровня n и двусторонних предмаксимальных слов для любого уровня $(n, k) \in \mathbb{N}^2$. Тем самым решена открытая проблема, сформулированная Рестиво и Салеми в [35] и Аллушем и Шаллитом в [3]. Результаты опубликованы в [54, 55, 58].

2. Доказана логарифмическая оценка на длину фиксированного контекста слова w в зависимости от $|w|$ в случае тернарного бесквадратного языка SF. Эта оценка асимптотически неулучшаема и значительно усиливает оценку Сони и Шелтона [41]. Получена нижняя оценка на частоту ветвления префиксного (суффиксного) дерева тернарного бесквадратного языка — 2/9, на основании которой, учитывая индекс роста языка SF, можно сделать вывод, что дерево языка SF имеет структуру, близкую к однородной. Результаты опубликованы в [56].

3. Доказано, что любой буквенный шаблон длины 4 и менее неизбежен в тернарном бесквадратном языке. Шаблоны длин 5 и 6 классифи-

цированы по избегаемости. Этот результат логически продолжает и обобщает результаты Туэ [51], опубликован в [53].

Автор выражает благодарность своему научному руководителю Шуру Арсению Михайловичу за постоянный интерес к работе и огромную помощь в оформлении диссертации и публикаций.

Список литературы

[1] Aberkane A., Currie J. D. Attainable lengths for circular binary words avoiding k-powers // Bull. Belg. Math. Soc. Simon Stevin. — 2005. — Vol. 12, no. 4. — P. 525–534.

[2] Aberkane A., Currie J. D. The Thue-Morse word contains circular $(5/2)^+$-power-free words of every length // Theoret. Comput. Sci. — 2005. — Vol. 332. — P. 573–581.

[3] Allouche J.-P., Shallit J. Automatic Sequences: Theory, Applications, Generalizations. — Cambridge University Press, 2003.

[4] Badkobeh G., Chairungsee S., Crochemore M. Hunting Redundancies in Strings // Proc. 15th Developments in Language Theory. DLT 2011. — Vol. 6795 of LNCS. — Berlin : Springer, 2011. — P. 1–14.

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

[6] Badkobeh G., Crochemore M. Fewest Repetitions in Infinite Binary Words // RAIRO Inform. Théor. App. — 2012. — Vol. 46, no. 1. — P. 1–31.

[7] Badkobeh G., Crochemore M., Rao M. Finite Repetition Threshold for Large Alphabets // Proc. 14th Mons Days of Theoretical Computer Science, 2012.

[8] Bean D. A., Ehrenfeucht A., McNulty G. Avoidable patterns in strings of symbols // Pacific J. Math. — 1979. — Vol. 85. — P. 261–294.

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

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

[11] Cassaigne J. Unavoidable binary patterns // Acta Informatica. — 1993. — Vol. 30. — P. 385–395.

[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] Crochemore M. Sharp characterizations of squarefree morphisms // Theoret. Comput. Sci. — 1982. — Vol. 18. — P. 221–226.

[14] Currie J. D. On the structure and extendibility of k-power free words // European J. Combinatorics. — 1995. — Vol. 16. — P. 111–124.

[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] Currie J. D., Shelton R. O. The set of k-power free words over Σ is empty or perfect // European J. Combinatorics. — 2003. — Vol. 24. — P. 573–580.

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

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

[20] Goralcik P., Vanicek T. Binary patterns in binary words // Internat. J. Algebra Comput. — 1991. — Vol. 1. — P. 387–391.

[21] Guglielmi N., Protasov V. Exact Computation of Joint Spectral Characteristics of Linear Operators // Foundations of Computational Mathematics. — 2013. — Vol. 13, no. 1. — P. 37–97.

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

[23] 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.

[24] Kolpakov R. Efficient lower bounds on the number of repetition-free words // J. Integer Sequences. — 2007. — Vol. 10. — P. 07.3.2.

[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] Kolpakov R. M. On the number of repetition-free words // J. Appl. Ind. Math. — 2007. — Vol. 1, no. 4. — P. 453–462.

[27] Mercaş R., Ochem P., Samsonov A., Shur A. Binary patterns in binary cube-free words: Avoidability and growth // RAIRO Inform. Théor. App. — 2014. — Vol. 48, no. 4. — P. 369–389.

[28] Mercaş 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.

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

[30] 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.

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

[32] 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.

[33] Rampersad N., Shallit J., Wang M. Avoiding large squares in infinite binary words // Theoret. Comput. Sci. — 2005. — Vol. 339. — P. 19–34.

[34] Rao M. Last Cases of Dejean's Conjecture // Theoret. Comput. Sci. — 2011. — Vol. 412. — P. 3010–3018.

[35] 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. — Vol. 192 of LNCS. — Springer-Verlag, 1985. — P. 198–206.

[36] Restivo A., Salemi S. Some decision results on non-repetitive words // Combinatorial algorithms on words / Ed. by A. Apostolico, Z. Galil. — Vol. F12 of NATO ASI series. — Springer-Verlag, 1985. — P. 289–295.

[37] Richomme G., Wlazinski F. Some results on k-power-free morphisms // Theoret. Comput. Sci. — 2002. — Vol. 273. — P. 119–142.

[38] Samsonov A. V., Shur A. M. On Abelian repetition threshold // Proc. 13th Mons Days of Theoretical Computer Science. — Univ. de Picardie Jules Verne, Amiens, 2010. — P. 1–11.

[39] 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.

[40] Shelton R. Aperiodic words on three symbols. II // J. Reine Angew. Math. — 1981. — Vol. 327. — P. 1–11.

[41] Shelton R. O., Soni R. P. Aperiodic words on three symbols. III // J. Reine Angew. Math. — 1982. — Vol. 330. — P. 44–52.

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

[43] Shur A. M. Two-sided bounds for the growth rates of power-free languages // Proc. 13th Int. Conf. on Developments in Language Theory. DLT 2009. — Vol. 5583 of LNCS. — Springer, 2009. — P. 466–477.

[44] Shur A. M. Growth of power-free languages over large alphabets // Proc. 5th International Computer Science Symposium in Russia. CSR 2010. — Vol. 6072 of LNCS. — Springer, 2010. — P. 350–361.

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

[46] Shur A. M. On ternary square-free circular words // Electronic J. Combinatorics. — 2010. — Vol. 17, no. # R140.

[47] Shur A. M. On the existence of minimal β-powers // Proc. 14th Int. Conf. on Developments in Language Theory. DLT 2010. — Vol. 6224 of LNCS. — Springer, 2010. — P. 411–422.

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

[49] Shur A. M. Growth properties of power-free languages // Computer Science Review. — 2012. — Vol. 6. — P. 187–208.

[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 Internat. Conf. on Mathematical Foundations of Computer Science. MFCS 2012. — Vol. 7464 of LNCS. — 2012. — P. 801–813.

Публикации автора по теме диссертации

Статьи, опубликованные в изданиях, рекомендованных ВАК

[53] Petrova E. A. Avoiding letter patterns in ternary square-free words // Electronic J. Combinatorics. — 2016. — Vol. 23, no. 1. — P. #P1.18.

[54] Petrova E. A., Shur A. M. Constructing premaximal binary cube-free words of any level // Internat. J. Found. Comp. Sci. — 2012. — Vol. 23, no. 8. — P. 1595–1609.

[55] Petrova E. A., Shur A. M. Constructing premaximal ternary square-free words of any level // Proc. 37th Internat. Conf. on Mathematical Foundations of Computer Science. MFCS 2012. — Vol. 7464 of LNCS. — 2012. — P. 752–763.

[56] Petrova E. A., Shur A. M. On the tree of ternary square-free words // Proc. 10th Internat. Conf. Words (WORDS 2015). — Vol. 9304 of LNCS. — 2015. — P. 223–236.

Другие публикации

[57] Petrova E. A., Shur A. M. Constructing premaximal binary cube-free words of any level // Proc. 1st Russian Finnish Symp. on Discrete Mathematics, Saint-Petersburg, 2011. — 2011. — P. 67–68.

[58] Petrova E. A., Shur A. M. Constructing premaximal binary cube-free words of any level // Proc. 8th Internat. Conf. Words (WORDS 2011). — Vol. 63 of EPTCS. — 2011. — P. 168–178.

[59] Petrova E. A., Shur A. M. Constructing premaximal ternary square-free words of any level // Proc. 2nd Russian Finnish Symp. on Discrete Mathematics, Turku, 2012. — TUCS Lecture Notes. — 2012. — P. 146–148.