

На правах рукописи

ФРИД Анна Эдуардовна

КОМБИНАТОРНЫЕ СЛОЖНОСТНЫЕ
ХАРАКТЕРИСТИКИ БЕСКОНЕЧНЫХ СЛОВ,
ЯЗЫКОВ И ПЕРЕСТАНОВОК

Специальность 01.01.09 – дискретная математика
и математическая кибернетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
доктора физико-математических наук

Новосибирск – 2012

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте математики им. С. Л. Соболева Сибирского отделения РАН

Официальные оппоненты: доктор физико-математических наук
Д. С. Кротов
доктор физико-математических наук,
профессор В. Л. Селиванов
доктор физико-математических наук
А. М. Шур

Ведущая организация: Учреждение Российской академии
наук Санкт-Петербургское отделение
Математического института
им. В.А. Стеклова РАН

Защита состоится ?? ??бря ???? г. в ?? час. 00 мин. на заседании диссертационного совета Д 003.015.01 при Учреждении Российской академии наук Институте математики им. С. Л. Соболева Сибирского отделения РАН по адресу: 630090 г. Новосибирск, пр. Академика Коптюга 4, к. 417.

С диссертацией можно ознакомиться в библиотеке Учреждения Российской академии наук Института математики им. С. Л. Соболева Сибирского отделения РАН.

Автореферат разослан ” ____ ” _____ 2012 г.

Ученый секретарь
диссертационного совета
д.ф.-м.н.

Ю. В. Шамардин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В диссертации рассматриваются свойства и характеристики бесконечных слов и родственных им объектов, а именно бесконечных перестановок и факторных языков, зависящие только от множества входящих в слово (бесконечную перестановку, язык) фрагментов — подслов или факторов. Вопросы, частичные или полные ответы на которые даются в работе, в общих терминах формулируются следующим образом: какими свойствами может и какими не может обладать множество подслов данного слова (перестановки)? Сколько определенных заданным образом фрагментов данного размера может содержаться в бесконечном слове (перестановке), как может или не может расти их количество? Как факторный язык можно разложить в произведение нескольких факторных языков, и как решать уравнения над факторными языками?

Все эти вопросы, а следовательно, и результаты диссертации относятся к комбинаторике на словах — области, тесно связанной как с классической комбинаторикой и алгеброй, так и с символической динамикой. История комбинаторики на словах начинается с работ Туэ [36, 35], а также Морса и Хедлунда [29]. Туэ задался вопросом: существует ли бесконечное слово над конечным алфавитом, в котором никогда не встречаются *квадраты*, то есть слова, состоящие из двух последовательных вхождений одинаковых подслов: например, квадратом является слово *abcabc*, представляющее собой два подряд идущих вхождения подслова *abc*? Нетрудно проверить, что над бинарным алфавитом таких слов не бывает, но Туэ удалось построить первый пример такого слова над алфавитом из трех символов. Со временем его результаты стали основой теории избегаемости [2, 3, 12, 17, 19, 21, 26], тесно связанной с классической комбинаторикой и алгеброй. Так, существование слова, избегающего квадратов, было использовано Новиковым и Адяном в решении проблемы Бернсайда [1, 7].

С другой стороны, Морс и Хедлунд [29] занимались символической динамикой и рассматривали слова как коды траекторий точки в динамической системе. Их интересовало, в частности, возможное число $p_w(n)$ подслов заданной длины n в бесконечном слове w , которое можно интерпретировать как одну из возможных мер

сложности бесконечного слова — его *комбинаторную*, или *подсловную* сложность. К их классическим результатам относятся лемма о том, что сложность $p_w(n)$ ограничена тогда и только тогда, когда $p_w(n) \leq n$ для некоторого n , или когда слово w периодически (возможно, с предпериодом). Они же описали непериодические слова с наименьшей сложностью $p_w(n)$, равной $n + 1$ для всех n . Эти слова, названные Морсом и Хедлундом словами Штурма, в настоящий момент хорошо изучены и представляют собой одно из классических семейств бесконечных слов [14].

Теория подсловной сложности активно развивается по сей день. Совершенствуются как методы подсчета комбинаторной сложности конкретных примеров слов (которым была посвящена, в частности, кандидатская диссертация автора), так и методы, позволяющие получать более общие результаты о том, какой может или не может быть сложность бесконечного слова [15]. Как пример одного из лучших результатов такого рода можно привести теорему Кассе-ня [16], утверждающую, что если подсловная сложность бесконечного слова растет линейно, то ее первые разности $p_w(n+1) - p_w(n)$ ограничены сверху константой. Тем не менее, в теории подсловной сложности существует множество нерешенных задач: в частности, до сих пор неизвестна удобная характеристизация слов с линейной сложностью. Одно из последних продвижений в этом направлении принадлежит Ж. Леруа [28].

Известный нерешенный вопрос теории чисел о том, любая ли последовательность цифр встречается в q -ичном разложении данного иррационального числа (например, в десятичном разложении числа π), также является вопросом о подсловной сложности: равна ли подсловная сложность соответствующего слова q^n ? Самым сильным из известных результатов о комбинаторной сложности q -ичных разложений чисел является результат Адамчевского и Бюжо [9], утверждающий, что комбинаторная сложность всякого q -ичного разложения алгебраического иррационального числа растет быстрее, чем линейно.

Как в идущей от Туэ теории избегаемости, так и в идущей от Морса и Хедлунда теории сложности активно применяются одни и те же стандартные способы построения бесконечных слов с нетривиальными свойствами. Кроме того, наряду со словами изучаются *факторные* языки, то есть языки, замкнутые относительно взятия

подслов. Факторными являются как языки подслов бесконечного слова или семейства бесконечных слов, так и языки всех слов, избегающих подслов заданного вида; изучается их подсловная сложность [4, 11, 13, 34] и другие свойства, например, возможная частота символов [31].

В последние годы развиваются также направления, основанные на модификациях определений комбинаторной сложности. Оставаясь в рамках подхода, при котором оценивается не сложность построения последовательности символов, а количество фрагментов заданного размера в этой последовательности, можно вводить и исследовать разнообразные функции, родственные комбинаторной сложности, но не равные ей. Так, Иванов в 1987 г. ввел понятие d -сложности [23]; Рестиво и Салеми в 2002 г. ввели шаблонную сложность [32] как количество слов над алфавитом переменных, вхождения которых появляются в данном бесконечном слове; Накашима, Тамура и Ясутоми ввели еще одно понятие модифицированной сложности [30]. Лучше всего исследованы и наиболее перспективны в данный момент две нестандартных комбинаторных функции сложности: максимальная шаблонная сложность, введенная в 2002 г. Камаэ и Замбони [24, 25] и арифметическая сложность, введенная в 2000 г. С. В. Августиновичем, Д. Г. Фон-Дер-Флаассом и автором [41].

Свойствам арифметической сложности полностью посвящена глава 2 диссертации. Удалось получить нетривиальные результаты об этой функции сложности. Так, наименьшую арифметическую сложность среди равномерно рекуррентных непериодических слов имеют не слова Штурма, как чаще всего бывает, а специфические слова, порожденные схемами Теплица. С помощью схем Теплица полностью описываются и все равномерно рекуррентные слова, арифметическая сложность которых растет линейно, сложность же слов Штурма оказывается равной $O(n^3)$. Заметим, что арифметическая сложность является в своем роде продолжением исследований, начатых классическими теоремами Ван дер Вардена и Семереди. О других известных исследованиях арифметических прогрессий можно прочесть в обзоре [8].

Новым научным направлением, развиваемым в диссертации, является установление комбинаторных сложностных характеристик бесконечных слов и родственных им объектов, наиболее ярко де-

монстрирующих нетривиальность свойств слов, перестановок и языков, дающих красивые и нетривиальные результаты о низкой или линейной сложности и предоставляющих новые техники для работы с соответствующими объектами. Кроме арифметической сложности бесконечных слов, к новым понятиям, описываемым и изучаемым в диссертации, относится понятие бесконечной перестановки, которому посвящена глава 3. Бесконечная перестановка определяется аналогично бесконечному слову, но роль ее «подслов» играют обычные конечные перестановки. Представляет интерес изучение свойств перестановок, аналогичных свойствам слов, и установление тех из них, которые наиболее органично описывают их структуру. Одновременно с автором и ее соавторами изучение бесконечных перестановок в данном контексте начали М. А. Макаров [5, 6] и Уидмер [37]. Среди более ранних работ, сходных по тематике, стоит отметить статью Дэвиса и др. [20].

Наконец, глава 4 диссертации посвящена свойствам произведений факторных языков и решению уравнений над ними. В отличие от уравнений над словами, решение которых возможно с помощью алгоритма Маканина [22], уравнения над языками в общем случае демонстрируют множество неожиданных свойств [18, 27, 33] и крайне редко могут быть решены полностью. Однако если ограничить рассматриваемый класс факторными языками, ситуация резко меняется, и уравнения над языками в большой степени начинают сводиться к уравнениям над словами. Таким образом, в работе разработана новая методика, позволяющая решать над факторными языками уравнения, решение которых над произвольными языками весьма сложно.

Цели работы: исследовать новые комбинаторные сложностные характеристики бесконечных слов; распространить как известные, так и новые функции от бесконечных слов на родственные им объекты, в частности, факторные языки и бесконечные перестановки, и исследовать, какие именно из этих функций являются естественными и важными для этих объектов; выработать новые методы исследования факторных языков слов и перестановок.

Методы исследования. В диссертации применяются как классические комбинаторные методы, в частности, методы комбинаторики на словах, так и методы алгебры, теории чисел, дискретной

геометрии и теории графов. В статьях, написанных в соавторстве, соавторы автора диссертации использовали также аналитические и динамические методы; соответствующие результаты, принадлежащие соавторам по работам [38, 44], приведены в диссертации без доказательств.

Хочется отметить, что многие методы исследования, применяемые в данной работе, разработаны автором и применены впервые. К самым нетривиальным из них можно отнести, прежде всего,

- метод бесконечных специальных ветвей, применяемый в главе 2 для характеристики слов с низкой арифметической сложностью;
- все примененные в главе 3 методы работы с бесконечными перестановками;
- метод выжимок до канонического разложения, применяемый в главе 4 для решения уравнений над факторными языками.

Научная новизна. Все результаты, представленные в диссертации, носят теоретический характер и являются **новыми**.

Практическая и теоретическая ценность. Работа носит теоретический характер. Полученные в ней результаты могут иметь применение в комбинаторике на словах и смежных областях алгебры и символической динамики.

Основными результатами работы являются следующие.

1. Охарактеризованы равномерно рекуррентные бесконечные слова, арифметическая сложность которых растет линейно.
2. Описаны равномерно рекуррентные бесконечные слова, арифметическая сложность которых минимальна для непериодических слов.
3. Найдены верхняя и нижняя оценки, а в некоторых случаях (включая случай слова Фибоначчи) точные формулы для арифметической сложности слов Штурма: оказалось, что она

растет как $O(n^3)$. Этот же результат может быть переформулирован следующим образом: найдены оценки, а в некоторых случаях точные формулы для количества всех вращательных слов с фиксированной шириной интервала.

4. Введено понятие периодичности бесконечных перестановок; установлено, что существует бесконечное количество периодических бесконечных перестановок с данной длиной периода; доказано, что периодичность перестановки эквивалентна ограниченности ее сложности; исследован возможный медленный рост сложности непериодических перестановок.
5. Охарактеризованы бесконечные перестановки, максимальная шаблонная сложность которых минимальна для непериодических перестановок и равна n ; как оказалось, они порождаются тем же способом, что и слова Штурма, и поэтому названы перестановками Штурма.
6. Доказана теорема существования и единственности канонического разложения факторного языка в конкатенацию факторных языков.
7. Полностью решено уравнение коммутирования $XY = YX$ на множестве бинарных факторных языков.

Апробация: результаты диссертации докладывались, в частности, на четырех приглашенных докладах на международных конференциях, а именно на Рабочей встрече об избегаемости слов, сложности и морфизмах (Турку, Финляндия, 2004); на Летней сессии Канадского математического общества (Ватерлоо, Канада, 2005); на Международной алгебраической конференции (Екатеринбург, 2005); на конференции WORDS 2011 (Прага, Чехия).

Кроме того, результаты были доложены на Международном конгрессе по словам, языкам и комбинаторике (Киото, Япония, 2000); на конференциях WORDS 2003 (Турку, Финляндия), WORDS 2005 (Монреаль, Канада), WORDS 2009 (Салерно, Италия); на Европейской конференции по комбинаторике (Берлин, Гер-

мания, 2005); на Конференции по комбинаторике, автоматам и теории перечисления (Льеж, Бельгия, 2006); на Международном симпозиуме по теоретической информатике в России (Санкт-Петербург, 2006); на Международной конференции по теории формальных языков (Турку, Финляндия, 2007); на Школе по комбинаторике на словах (Монреаль, Канада, 2007); на Монсовских днях (Амьен, Франция, 2010); на Рабочей группе по комбинаторике на словах (Банфф, Канада, 2012); на Мальцевских чтениях (Новосибирск, 2006); на Лаврентьевских чтениях (Новосибирск, 2003 и 2007); в Санкт-Петербургском Computer Science Club (2011); на семинарах в университетах Марселя, Лиона, Нанси и Турку (2009–2012); на семинарах «Факторные языки», «Теория кодирования», «Теория графов» в Институте математики им. С. Л. Соболева СО РАН (1999–2011).

Публикации и личный вклад. По теме диссертации опубликовано 17 работ [38–54], из них 12 (все, кроме [39], [41], [43], [48,49]) — в изданиях из списка ВАК. Работы [40–43,45,46,54] написаны в неразделимом соавторстве с С. В. Августинovichем, Ж. Кассенем, Д. Г. Фон-Дер-Флаассом, Л. Замбони. В статьях [38] и [44] присутствуют леммы из математического анализа и символической динамики, полностью принадлежащие соавторам автора, в диссертации они приводятся без доказательств.

Структура и объем работы. Диссертация состоит из введения, четырех глав и списка литературы (105 наименований, включая 17 работ автора по теме диссертации, приведенных в конце списка). Текст работы изложен на 245 страницах и содержит 27 рисунков.

Содержание работы

В главе 1 работы вводятся понятия, применяемые в следующих главах: периодичность и рекуррентность бесконечного слова, комбинаторная сложность, максимальная шаблонная сложность, а также несколько способов построения непериодических бесконечных слов. Кроме того, там же описывается техника специальных

и биспециальных слов для вычисления комбинаторной сложности слова или языка.

Глава 2 посвящена определению и свойствам арифметической сложности бесконечных слов. Эта функция, введенная в работе С. В. Августиновича, Д. Г. Фон-Дер-Флаасса и автора в 2000 г. [41], определяется как количество всех слов данной длины, встречающихся в бесконечном слове $w = w_1 \cdots w_n \cdots$ по арифметическим прогрессиям:

$$a_w(n) = \#\{w_k w_{d+k} \cdots w_{(n-1)d+k} \mid k, d \geq 1\}.$$

Свойства арифметической сложности оказываются весьма нетривиальными и дают новую, часто неожиданную информацию о свойствах известных классов слов. Так, в разделе 2.2 найдена арифметическая сложность неподвижных точек симметрических морфизмов — стандартного семейства, слова из которого часто используются как примеры бесконечных слов с заданными свойствами. В частности, именно в нем содержится слово Туэ-Морса, классический пример бинарного слова, появляющегося в самых разных контекстах и строящегося как неподвижная точка $0110100110010110 \cdots$ морфизма $0 \mapsto 01, 1 \mapsto 10$. Комбинаторная сложность таких слов растет линейно, а арифметическая, как устанавливается в разделе 2.2, растет экспоненциально (теорема 2.2.3). В частности, сложность слова Туэ-Морса оказывается равна 2^n , то есть является максимально возможной.

Этот результат позволяет задаться следующим вопросом: а существуют ли бесконечные слова, арифметическая сложность которых растет достаточно медленно, например, линейно? Положительный ответ на этот вопрос дается в разделе 2.3, где описываются равномерно рекуррентные слова с наименьшей арифметической сложностью. Оказывается, что для непериодичных равномерно рекуррентных слов медленнее всего растет арифметическая сложность слов Теплица, порожденных чередованием схем $P_{a,p} = a^{p-1}\diamond$ и $P_{b,p} = b^{p-1}\diamond$, где число p простое, а a и b — различные символы алфавита. Эти слова строятся как предел следующей итеративной конструкции: начав со слова, состоящего из одних пробелов, то есть из символов $\diamond$, мы последовательно заменяем все оставшиеся пробелы сначала на слово $P_{a,p}^\omega$, потом на слово $P_{b,p}^\omega$, потом снова на

$P_{a,p}^\omega$, и так далее. В частности, при $p = 3$ это последовательно дает слова $\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond\diamond\dots$; $aa \diamond aa \diamond aa \diamond aa \dots$; $aabaabaa \diamond aa \dots$; $aabaabaaaaa \dots$.

Арифметическая сложность $a_p(n)$ слова, порождаемого таким образом с помощью схем $P_{a,p}$ и $P_{b,p}$, удовлетворяет неравенствам

$$\frac{2p+2}{p} \leq \frac{a_p(n)-2}{n} \leq \frac{6p-2}{2p-1},$$

а арифметическая сложность всех других равномерно рекуррентных непериодических слов не может расти медленнее, чем одна из функций $a_p(n)$. В этом и состоит основной результат раздела 2.3, совместный с С. В. Августиновичем и Ж. Кассенем и впервые опубликованный в [40].

В следующем разделе 2.4 техника бесконечных специальных ветвей, впервые примененная в разделе 2.3, усложняется и модифицируется для получения следующей характеристики всех непериодических равномерно рекуррентных слов, арифметическая сложность которых растет линейно.

Назовем слово u *канонически p -регулярным*, если для всех $k > 0$ и $i \in \{1, \dots, p^k - 1\}$ его арифметические подпоследовательности $u_{p^k}^i$ периодичны. Канонически p -регулярные слова также могут быть определены с помощью схем Теплица, подобных описанным выше схемам $P_{a,p}$.

Основным результатом раздела 2.4 является следующая

Теорема 2.4.1. *Арифметическая сложность непериодического равномерно рекуррентного слова растет линейно тогда и только тогда, когда это слово принадлежит орбите некоторого канонически p -регулярного слова w , где число p просто, и $w_{p^a}^{p^a} = w_{p^b}^{p^b}$ для некоторых $a \neq b$.*

В 2007-м г. автор работы получил за этот результат Третью премию Эйлера в номинации “Молодые исследователи”.

В следующих двух разделах, 2.5 и 2.6, с помощью схем Теплица строятся семейства примеров бесконечных слов, арифметическая сложность которых растет как n^s для какой-либо степени $s \geq 1$. Основными результатами соответствующих разделов являются следующие две теоремы.

Теорема 2.5.1. Для всякого бесконечного слова u над конечным алфавитом Σ , комбинаторная сложность которого обозначается через $p_u(n)$, и для всякого простого $p \geq 3$ существует бесконечное слово над алфавитом Σ , арифметическая сложность которого растет как $\Theta(np_u(\lceil \log_p n \rceil))$.

Теорема 2.6.1. Пусть $Q = \{q_1, \dots, q_k\}$ — произвольное конечное множество простых чисел, $k \leq 2$. Тогда для любого $d > 1$ существует бесконечное слово $w = w(Q, d)$ над d -буквенным алфавитом, для арифметической сложности $a_w(n) = p_{Q,d}(n)$ которого выполняется свойство

$$\frac{p_{Q,d}(n)}{n^{\alpha(Q,d)+1}} \rightarrow C \text{ при } n \rightarrow \infty,$$

где C — некоторая положительная константа, а $\alpha(Q, d)$ — единственный положительный корень уравнения

$$\frac{d-1}{d} = \prod_{i=1}^k (1 - q_i^{-x}).$$

В обеих теоремах соответствующее слово строится с помощью схем Теплица, но если в доказательстве теоремы 2.5.1 применяются чисто комбинаторные методы, в том числе уже ставшая классической техника биспециальных слов, то в доказательстве теоремы 2.6.1, совместного результата с Ж. Кассенем и Ф. В. Петровым, для оценки асимптотики арифметической сложности используются, кроме комбинаторных, аналитические методы. Две используемые для этой оценки леммы принадлежат соавторам автора и поэтому приводятся в тексте диссертации без доказательства.

Из вышеописанных результатов следует, в частности, что у слов Штурма — слов с наименьшей комбинаторной сложностью, равной $n + 1$, — арифметическая сложность растет быстрее, чем линейно. Верхняя оценка и в ряде случаев точные формулы для комбинаторной сложности слов Штурма приводятся в разделе 2.7. Результаты этого раздела получены совместно с Ж. Кассенем [45]. Основной используемый в разделе метод впервые применен Берстелем и Поккьолой [13], нашедшими с помощью него совокупное

количество всех подслов слов Штурма заданной длины. Метод основан на том, что каждому слову, заданному двумя параметрами, — например, всякому арифметическому подслову слова Штурма — можно поставить в соответствие точку в двумерном пространстве, координаты которой равны этим двум параметрам, причем при правильном выборе параметров точки из некоторой связной области будут соответствовать одному и тому же слову. Таким образом, верхняя оценка количества всех арифметических подслов слова Штурма может быть получена как число таких связных областей. В данном случае, как и в задаче, рассмотренной Берстелем и Поккёлой, области являются гранями некоторого плоского графа, и их количество может быть найдено по формуле Эйлера.

Далее, однако, задача становится более трудоемкой, чем рассмотренная Берстелем и Поккёлой: если в их случае количество нужных слов было равно количеству граней и нахождение количества граней означало решение задачи, то в этом случае необходимо классифицировать грани, соответствующие одному и тому же слову. Как оказалось, вся схема расположения граней обладает центральной симметрией, что дает возможность разделить количество граней на два и получить следующую верхнюю оценку (единую для всех слов Штурма):

Лемма 2.7.8. *Для всех n и для всякого слова Штурма w верно неравенство*

$$a_w(n) \leq \frac{(n-1)n(n+1)}{6} + \sum_{p=1}^{n-1} (n-p)\varphi(p) + 2,$$

где $\varphi(p)$ — функция Эйлера, то есть количество натуральных чисел, не превосходящих p и взаимно простых с ним.

Заметим, что правая часть вышеприведенного неравенства растет как $(1/6 + 1/\pi^2)n^3 + O(n^2)$, то есть полученная оценка растет как $\Theta(n^3)$. Точные значения арифметической сложности зависят от того, какое слово Штурма рассматривается, а точнее, от *наклона* этого слова. Путем подробного рассмотрения случаев, когда разные грани соответствуют одному и тому же слову, удастся найти точные значения арифметической сложности слов Штурма с наклонами, близкими к $1/2$, а точнее, с наклонами из интервала

$[1/3; 2/3]$. В частности, в этот интервал попадает наклон слова Фибоначчи — самого известного из всех слов Штурма, неподвижной точки $abaababa \dots$ морфизма $a \mapsto ab, b \mapsto a$. Наклон слова Фибоначчи равен $\alpha = (3 - \sqrt{5})/2 = 0,381966 \dots$, и потому оно попадает в зону действия следующей теоремы.

Теорема 2.7.15. *Для всякого иррационального $\alpha \in (0.375, 0.4)$ верны равенства $a_\alpha(1) = 2, a_\alpha(2) = 4, a_\alpha(3) = 8, a_\alpha(4) = 16, a_\alpha(5) = 30, a_\alpha(6) = 52, a_\alpha(7) = 83, a_\alpha(8) = 128$ и*

$$a_\alpha(n+1) = \begin{cases} g(n) - 8, & \text{при четном } n, \\ g(n) - 9, & \text{при нечетном } n \end{cases}$$

$$\text{для } n+1 \geq 9, \text{ где } g(n) = \frac{n(n+1)(n+2)}{6} + \sum_{p=1}^n (n-p+1)\varphi(p) + 2.$$

Заметим, что для наклонов, не входящих в интервал $[1/3; 2/3]$, неизвестно, ограничена ли разность между имеющейся верхней оценкой арифметической сложности соответствующего слова Штурма и самой арифметической сложностью. Однако, как показывает основной результат раздела 2.8, арифметическая сложность всякого слова Штурма тем не менее растет как $\Theta(n^3)$. Справедлива следующая теорема, впервые опубликованная в [39].

Теорема 2.8.1. *Для всех $n > 0$ арифметическая сложность слова Штурма с наклоном $\alpha < 1/2$ удовлетворяет неравенству*

$$a_\alpha(2n-2) \geq \sum_{p=1}^n p\varphi(p) - \frac{1}{2} \sum_{p=1}^{\lfloor 2/\alpha \rfloor} p\varphi(p) - \frac{1}{2} \sum_{p=1}^{\lfloor 1/\alpha \rfloor} p\varphi(p),$$

где $\varphi(p)$ — функция Эйлера.

Заметим, что для $\alpha > 1/2$ имеем $a_\alpha(n) = a_{1-\alpha}(n)$, и потому данная теорема фактически дает верхнюю оценку арифметической сложности любого слова Штурма. Эта оценка удовлетворяет неравенству $a_\alpha(n) \geq \frac{n^3}{4\pi^2} + O(n^2) - \frac{9}{\pi^2\alpha^3} + O\left(\frac{1}{\alpha^2}\right)$, а значит, меньше верхней примерно в 10,58 раз. В отличие от верхней оценки она не претендует на точность, поскольку получена достаточно грубыми методами.

Заметим также, что арифметическое замыкание всякого слова Штурма с наклоном α совпадает с множеством всех вращательных слов с соответствующей шириной интервала. Таким образом, все вышеприведенные оценки и формулы для арифметической сложности слова Штурма могут быть интерпретированы и как оценки и формулы для количества всех вращательных слов с данной шириной интервала.

Глава 3 посвящена свойствам новых объектов, впервые введенных в работе [46], — бесконечных перестановок. Бесконечная перестановка может быть описана как множество натуральных или целых чисел, на которых, кроме стандартного линейного порядка, определен еще один, или, что равносильно, как последовательность (пронумерованная натуральными или целыми числами) попарно различных действительных чисел, в которой нас интересует порядок между элементами, но не их значения. Строгое определение бесконечной перестановки приводится в разделе 3.1. Важно, что аналогично подслову бесконечного слова, можно определить фактор длины n бесконечной перестановки, представляющий собой конечную перестановку длины n . Таким образом, бесконечные перестановки занимают промежуточное положение между классическими конечными перестановками и бесконечными словами.

В данной работе рассматриваются свойства бесконечных перестановок, аналогичные свойствам бесконечных слов, и демонстрируется, что, несмотря на сходство определений и структуры, свойства бесконечных перестановок часто оказываются более сложными и менее предсказуемыми, чем свойства бесконечных слов. Так, в разделе 2.2 устанавливается, что, в отличие от периодических слов, насчитывается счетное, а не конечное, число периодических перестановок с данным периодом. В разделе 2.3 рассматривается аналог для перестановок известной теоремы Файна — Уилфа:

Теорема (Файн, Уилф). *Если слово длины не меньше $p + q - (p, q)$ периодически с периодами p и q , то оно периодически с периодом (p, q) . Длина $p + q - (p, q)$ оптимальна для всех слов, кроме как над однобуквенным алфавитом, поскольку существуют бинарные слова длины $p + q - (p, q) - 1$, периодичные с периодами p и q , но не (p, q) .*

Оказывается, что теорема Файна — Уилфа может быть перенесена на перестановки лишь частично, а именно в случаях, когда периоды p и q взаимно просты.

Теорема 3.3.1. *Если перестановка α длины не меньше $p + q$ периодична с периодами p и q , где $(p, q) = 1$, то α периодична с периодом 1, то есть монотонна. Длина $p + q$ здесь наименьшая, поскольку всегда существуют перестановки длины $p + q - 1$, периодичные с периодами p и q , но не монотонные.*

Для произвольных периодов перестановок теорема Файна — Уилфа неверна, поскольку при $(p, q) \neq 1$ всегда можно построить перестановку, периодическую с периодами p, q , но не (p, q) . Тем не менее в качестве аналога теоремы Файна — Уилфа для перестановок общего вида доказана

Теорема 3.3.5. *Если перестановка α длины n является p - и q -периодической, то всякий ее фактор длины не более $n - p - q + 2(p, q) + 1$ является (p, q) -периодическим.*

В разделе 3.4 аналогично комбинаторной сложности слов определяется комбинаторная сложность бесконечных перестановок и доказывается, что, как и для слов, комбинаторная сложность перестановки ограничена константой тогда и только тогда, когда перестановка со временем периодична. Тем не менее, в отличие случая слов, комбинаторная сложность перестановки, определенной на натуральных числах, может расти сколь угодно медленно. Еще одно отличие от слов состоит в том, что сложность ведет себя по-разному в случае перестановок, бесконечных в одну или в обе стороны: если непериодическая перестановка определена на множестве $\mathbb{Z}$, то ее сложность ограничена снизу функцией $n - C$ для некоторой константы C .

Во всех вышеперечисленных результатах свойства перестановок оказывались более разнообразны, менее очевидны и несколько менее изящны, чем свойства слов. Ярким исключением оказываются результаты раздела 3.5, где на бесконечные перестановки переносится понятие максимальной шаблонной сложности. Известно [24], что максимальная шаблонная сложность непериодического бесконечного слова не может быть меньше, чем $2n$. Известно также, что сложность равна $2n$ как минимум для слов Штурма, слов Теплица, порожденных простыми схемами, и некоторых пере-

куррентных слов особого вида [25]. Однако до сих пор неизвестно, существуют ли другие примеры слов со сложностью $2n$. В то же время для непериодичных перестановок, как удалось установить в работе [44], совместной с С. В. Августинovichем, Т. Камаэ и П. В. Салимовым, наименьшая максимальная шаблонная сложность равна n , и, в отличие от слов, перестановки с максимальной шаблонной сложностью n удалось естественно охарактеризовать.

Теорема 3.5.2. *Максимальная шаблонная сложность $p_\alpha^*(n)$ бесконечной перестановки α равна n для всех n тогда и только тогда, когда α — перестановка Штурма.*

Перестановки Штурма, упоминаемые в этой формулировке, строятся с помощью слов Штурма или напрямую с помощью вращений единичной окружности, через которые задаются слова Штурма.

Наконец, в разделе 3.6 произведена попытка обобщить на бесконечные перестановки понятие автоматности бесконечных слов: слово называется k -автоматным, если его символ номер n может быть вычислен с помощью конечного автомата, на вход которого поступает k -ичное разложение числа n . Существует также по меньшей мере три других эквивалентных определения автоматного слова [10]. Оказывается, что эквивалентные определения автоматного слова при переносе на перестановки перестают быть эквивалентными. Основными результатами пункта 3.6 являются утверждения, описывающие строгую вложенность соответствующих классов. Самым нетривиальным из них является лемма 3.6.8, утверждающая, что всякая перестановка, порожденная автоматным словом, может быть порождена конечным автоматом.

Наконец, **глава 4** посвящена свойствам моноида факторных языков и решению уравнений над факторными языками. Напомним, что язык X называется факторным, если он замкнут относительно операции Fac взятия подслов всякого элемента: $X = \text{Fac}(X)$.

Назовем факторный язык X разложимым, если $X = YZ$ для некоторых факторных языков Y и Z , ни один из которых не равен X ; в противном случае язык X называется неразложимым. Разложение факторного языка X в произведение факторных языков $X_1, \dots, X_n$, $X = X_1 \cdots X_n$, назовем минимальным, если все языки X_i не равны $\{\lambda\}$, где λ пустое слово, и ни один из языков X_i не может быть без изменения результата заменен на свое собствен-

ное подмножество: $X \neq X_1 \cdots X'_i \cdots X_n$, где $X'_i \subsetneq X_i$. Минимальное разложение факторного языка в произведение неразложимых языков называется *каноническим* разложением языка X .

Важнейшим результатом главы 4 является следующая

Теорема 4.1.3. *У всякого факторного языка существует единственное каноническое разложение.*

Заметим, что теорема доказывается неконструктивными методами: в общем случае неизвестно, как по языку построить его каноническое разложение. В случае регулярного языка на этот вопрос получен ответ: в разделе 4.5 доказана теорема, согласно которой факторы канонического разложения всякого регулярного факторного языка регулярны и могут быть заданы некоторыми подавтоматами автомата, порождающего исходный язык.

Несмотря на неконструктивность доказательства, теорема 4.1.3 предоставляет технику, позволяющую сводить уравнения над факторными языками к уравнениям над словами. Метод решения уравнений основан на том, каким может быть каноническое разложение конкатенации факторных языков, каноническое разложение которых известно. Возможные варианты описаны в разделе 4.2. В разделе 4.4 описанная техника применена к решению над бинарными факторными языками уравнений $X^n = Y^n$, $XY = YX$ (коммутирование) и $XZ = ZY$ (сопряжение). Первое из них выполняется, только если $X = Y$; решения третьего слишком разнообразны, чтобы легко записываться; наконец, решения уравнения коммутирования $XY = YX$ над бинарными факторными языками полностью описаны и попадают в одну из пяти категорий. Чтобы перечислить их, для каждого факторного языка X введем подалфавиты

$$\Pi(X) = \{x \in \Sigma \mid Xx \subset X\} \text{ и } \Delta(X) = \{x \in \Sigma \mid xX \subset X\}.$$

Как нетрудно проверить, два бинарных факторных языка коммутируют в каждом из следующих пяти случаев.

Словесное коммутирование. $XY = YX$, если $X = Z^m$ и $Y = Z^n$ для некоторого факторного языка Z и неотрицательных целых n и m .

Поглощение. Пусть Σ_X — подалфавит всех букв, входящих в факторный язык X . Если $Y\Sigma_X \subseteq Y$, $\Sigma_X Y \subseteq Y$, то $XY = YX = Y = Y\Sigma_X^* = \Sigma_X^* Y$: язык Y *поглощает* язык X .

В бинарном случае поглощение значит, что либо $X = \{\lambda\}$, либо $X \subseteq x^*$ для некоторой буквы x и $\Pi(Y) = \Delta(Y) = \{x\}$, либо $Y = \{a, b\}^*$.

Неожиданное коммутирование I. Пусть Z — бинарный факторный язык с $\Delta(Z) = \{x\}$ и $\Pi(Z) = \{y\}$, $x \neq y$. Тогда для всех $r, p > 0$ язык Z^p коммутирует со всяким языком X , для которого верны вложения

$$Z^r \subseteq X \subseteq Z^r x^* \cap y^* Z^r.$$

Пример. Пусть $F_a = \text{Fac}(\{a, ab\}^*)$ и $F_b = \text{Fac}(\{b, ab\}^*)$, т.е. язык F_a содержит все слова, не содержащие двух стоящих рядом b , а язык F_b — все слова, не содержащие двух стоящих рядом a . Рассмотрим $Z = F_b \cdot F_a$. Тогда всякий язык $X = Z + S$, где S является факторным подмножеством языка a^*b^* , коммутирует со всякой степенью Z^p языка Z .

Неожиданное коммутирование II. Рассмотрим символ x алфавита $\{a, b\}$ и бинарный факторный язык Q , для которого $L_x(Q) = R_x(Q) = Q$ и $\Delta(Q), \Pi(Q)$ равны $\emptyset$ или $\{y\}$, $y \neq x$. Тогда для всех $p \geq 0$ и $r \geq 1$ язык $(x^*Q)^p x^*$ коммутирует со всяким языком A , удовлетворяющим включениям

$$(x^*Q)^r + (Qx^*)^r \subseteq A \subseteq (x^*Q)^r x^*.$$

Пример. Языки $A = a^*b^* + b^*a^*$ и $B = a^*$ коммутируют, поскольку $AB = BA = a^*b^*a^*$. Здесь $x = a$ и $Q = b^*$, так что в действительности B коммутирует со всяким языком, являющимся одновременно надмножеством языка $a^*b^* + b^*a^*$ и подмножеством языка $a^*b^*a^*$.

Неожиданное коммутирование III. Рассмотрим бинарный факторный язык Z , для которого верно равенство $\overline{ZZ} \doteq \overline{Z} \overline{Z}$, и $\Delta(Z) = \{x\}$. Пусть B — факторный язык, для которого $Z^n \subseteq B \subseteq Z^n x^*$, $n > 0$. Тогда для всех $m > 0$ язык B коммутирует с $Z^m B$.

Симметрично, для всякого бинарного факторного Z с $\overline{ZZ} \doteq \overline{Z} \overline{Z}$ и $\Pi(Z) = \{x\}$ и для всякого бинарного факторного языка B с $Z^n \subseteq B \subseteq x^* Z^n$, язык B коммутирует с BZ^m , каковы бы ни были $n, m > 0$.

Пример. Рассмотрим языки $Z = a^*b^*$ и $B = \text{Fac}(a^*(bb)^*a^*b^* + a^*b(bb)^*a^*b^*a^*)$. Здесь $\Delta(Z) = \{a\}$ и $Z^2 = a^*b^*a^*b^* \subseteq B \subseteq a^*b^*a^*b^*a^* = Z^2 a^*$. Таким образом, B коммутирует со всеми множествами A вида $A = Z^m B$: $AB = BA = Z^{m+2} B$.

Следующая теорема утверждает, что выше перечислены все возможные случаи коммутирования бинарных факторных языков.

Теорема 4.4.9. Два бинарных факторных языка коммутируют тогда и только тогда, когда возникает одна из вышеописанных ситуаций: либо словесное коммутирование, либо поглощение, либо неожиданное коммутирование I, II или III.

Список литературы

- [1] С. И. Адян. *Проблема Бернсайда и связанные с ней вопросы* // Успехи мат. наук. 2010. Т. 65. № 5. С. 5–60.
- [2] А. А. Евдокимов. *О сильно асимметричных последовательностях, порожденных конечным числом символов* // Докл. АН СССР. 1968. Т. 179. № 6. С. 1268–1271.
- [3] А. И. Зимин. *Блокирующие множества термов* // Мат. Сб. 1982. Т. 119 (161), № 3 (11), С. 363–375.
- [4] Е. П. Липатов. *Об одной классификации двоичных наборов и свойствах классов однородности* // Проблемы кибернетики. 1982. М.: Наука. Т. 39. С. 67–84.
- [5] М. А. Макаров. *О перестановках, порожденных бесконечными бинарными словами* // Сиб. электрон. матем. изв. 2006. Т. 3. С. 304–311.
- [6] М. А. Макаров. *Антимонотонные перестановки* // Сиб. электрон. матем. изв. 2012. Т. 9. С. 346–359.
- [7] П. С. Новиков, С. И. Адян. *О бесконечных периодических группах. I, II, III* // Изв. АН СССР. Сер. матем. 1968. Т. 32. № 1. С. 212–244. № 2. С. 251–524. № 3. С. 709–731.
- [8] И. Д. Шкредов. *Теорема Семереди и задачи об арифметических прогрессиях* // Успехи мат. наук. 2006. Т. 61. № 6. С. 111–179.
- [9] B. Adamczewski, Y. Bugeaud. *On the complexity of algebraic numbers I. Expansions in integer bases* // Annals of Math. 2007. V. 165. P. 547–565.
- [10] J.-P. Allouche, J. Shallit. *Automatic sequences — theory, applications, generalizations*. Cambridge University Press, 2003.
- [11] P. Ambrož, A. Frid, Z. Masáková, E. Pelantová. *On the number of factors in codings of three interval exchange* // Discr. Math. Theoret. Comput. Sci. 2011. V. 13. P. 51–66.

- [12] D. R. Bean, A. Ehrenfeucht, G. F. McNulty. *Avoidable patterns in strings of symbols* // Pacific J. Math. 1979. V. 85. P. 261–294.
- [13] J. Berstel, M. Pocchiola. *A geometric proof of the enumeration formula for Sturmian words* // Int. J. Algebra and Comput. 1993. V. 3. P. 349–355.
- [14] J. Berstel, P. Séébold. Sturmian words, in: M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2002. P. 40–97.
- [15] J. Cassaigne. *Complexité et facteurs spéciaux* // Bull. Belg. Math. Soc. Simon Stevin. 1997. V. 4. P. 67–88.
- [16] J. Cassaigne. *Special factors of sequences with linear subword complexity* // Developments of Language Theory II. Singapore: World Sci. Publishing. 1996. P. 25–34.
- [17] J. Cassaigne. Unavoidable patterns. In: M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press. 2002.
- [18] C. Choffrut, J. Karhumäki. *On Fatou properties of rational languages* // C. Martin-Vide and V. Mitrana, Eds. Where Mathematics, Computer Science, Linguistics and Biology Meet. Dordrecht: Kluwer, 2000. P. 227–235.
- [19] J. D. Currie. *Open problems in pattern avoidance* // Amer. Mathem. Monthly 1993. V. 100. P. 790–793.
- [20] J. A. Davis, R. C. Entringer, R. L. Graham, and G. J. Simmons. *On permutations containing no long arithmetic progressions* // Acta Arithmetica 1977. V. 34. P. 81–90.
- [21] F. Dejean. *Sur un théorème de Thue* // J. Combin. Theory Ser. A 1972. V. 12. P. 90–99.
- [22] V. Diekert. Makanin’s Algorithm, in: M. Lothaire, Algebraic combinatorics on words, Cambridge Univ. Press, 2002. P. 387–442.
- [23] A. Iványi. *On the d -complexity of words* // Ann. Univ. Sci. Budapest. Sect. Comput. 1987. V. 8. P. 69–90.
- [24] T. Kamae, L. Zamboni. *Sequence entropy and the maximal pattern complexity of infinite words* // Ergodic Theory Dynam. Systems 2002. V. 22. P. 1191–1199.
- [25] T. Kamae, L. Zamboni. *Maximal pattern complexity for discrete systems* // Ergodic Theory Dynam. Systems 2002. V. 22. P. 1201–1214.

- [26] V. Keränen. *Abelian squares are avoidable on 4 letters* // Automata, Languages and Programming. Berlin: Springer, 1992. P. 41–52. (LNSC V. 623).
- [27] M. Kunc. *The power of commuting with finite sets of words* // Theory Comput. Syst. 2007. V. 40. P. 521–551.
- [28] J. Leroy. *Some improvements of the S-adic conjecture* // Advances in Appl. Math. 2012. V. 48. P. 79–98.
- [29] M. Morse, G. A. Hedlund. *Symbolic Dynamics II. Sturmian trajectories* // Am. J. Math. 1940. V. 62. P. 1–42.
- [30] I. Nakashima, J.-I. Tamura, Sh.-I. Yasutomi. **-Sturmian words and complexity* // J. Théorie Nombres Bordeaux 2003. V. 15. P. 767–804.
- [31] P. Ochem. *Binary words avoiding the pattern AABBCABBA* // Theoret. Informatics Appl. 2010. V. 44. P. 151–158.
- [32] A. Restivo, S. Salemi. *Binary patterns in infinite binary words* // Formal and Natural Computing. Berlin: Springer, 2002. P. 107–116. (LNCS Vol. 2300).
- [33] A. Salomaa, S. Yu. *On the decomposition of finite languages* // Developments in Language Theory. Foundations, Applications, Perspectives. World Scientific, 2000. P. 22–31.
- [34] A. M. Shur. *Growth properties of power-free languages* // Proc. DLT 2011. Berlin: Springer, 2011. P. 28–43. (LNCS Vol. 6795).
- [35] A. Thue. *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen* // Norske Vid. Skrifter I Mat.-Nat. Kl., Christiania 1912. V. 10. P. 1–67.
- [36] A. Thue. *Über unendliche Zeichenreihen* // Norske Vid. Skrifter I Mat.-Nat. Kl., Christiania 1906. V. 7 P. 1–22.
- [37] S. Widmer. *Permutation complexity of the Thue-Morse word* // Adv. Appl. Math. 2011. V. 47. P. 309–329.

Публикации автора по теме диссертации

- [38] Ж. Кассень, Ф. В. Петров, А. Э. Фрид. *О возможных скоростях роста языков Теплица* // Сиб. мат. журнал 2011. Т. 52. № 1. С. 81–94.
- [39] А. Фрид. *Нижняя оценка на комбинаторную сложность слов Штурма* // Сиб. электрон. матем. изв. 2005. Т. 2. С. 14–22.

- [40] S. V. Avgustinovich, J. Cassaigne, A. E. Frid. *Sequences of low arithmetical complexity* // Theoret. Informatics Appl. 2006. V. 40. P. 569–582.
- [41] S. V. Avgustinovich, D. G. Fon-Der-Flaass, A. E. Frid. *Arithmetical complexity of infinite words* // Proc. Words, Languages and Combinatorics III, 2000. Singapore: World Scientific, 2003. P. 51–62.
- [42] S. V. Avgustinovich, A. E. Frid. *A unique decomposition theorem for factorial languages* // Int. J. Algebra Comput. 2005. V. 15. P. 149–160.
- [43] S. V. Avgustinovich, A. E. Frid. *Canonical decomposition of a regular factorial language* // Proc. CSR 2006. Berlin: Springer, 2006. P. 18–22. (LNCS Vol. 3967).
- [44] S. V. Avgustinovich, A. E. Frid, T. Kamae, P. Salimov. *Infinite permutations of lowest maximal pattern complexity* // Theoret. Comput. Sci. 2011. V. 412. P. 2911–2921.
- [45] J. Cassaigne, A. E. Frid. *On arithmetical complexity of Sturmian words* // Theoret. Comput. Sci. 2007. V. 380. P. 304–316.
- [46] D. G. Fon-Der-Flaass, A. E. Frid. *On periodicity and low complexity of infinite permutations* // European J. Combinatorics 2007. V. 28. P. 2106–2114.
- [47] A. E. Frid. *Arithmetical complexity of symmetric D0L words* // Theoret. Comput. Sci. 2003. V. 306. P. 535–542.
- [48] A. Frid. *Canonical decomposition of a catenation of factorial languages* // Sib. Electron. Math. Reports 2007 V. 4. P. 12–19.
- [49] A. Frid. *Commutation of binary factorial languages* // Proc. DLT 2007. Berlin: Springer, 2007. P. 193–204. (LNCS Vol. 4588).
- [50] A. Frid. *Fine and Wilf’s theorem for permutations* // Sib. Electron. Math. Reports 2012. V. 9. P. 377–381.
- [51] A. E. Frid. *On possible growths of arithmetical complexity* // Theoret. Informatics Appl. 2006. V. 40. P. 443–458.
- [52] A. E. Frid. *Sequences of linear arithmetical complexity* // Theoret. Comput. Sci. 2005. V. 339. P. 68–87.
- [53] A. Frid. *Simple equations on binary factorial languages* // Theoret. Comput. Sci. 2009. V. 410. P. 2947–2956.
- [54] A. Frid, L. Zamboni. *On automatic infinite permutations* // Theoret. Informatics Appl. 2012. V. 46. P. 77–85.

Фрид Анна Эдуардовна

Комбинаторные сложностные характеристики бесконечных слов,
языков и перестановок

Автореферат диссертации
на соискание ученой степени
доктора физико-математических наук

Подписано в печать ??.??.???. Формат 60х84 1/16.
Усл. печ. л. 2,0. Уч.-изд. л. 2,0. Тираж 120 экз. Заказ № ??.

Отпечатано в ООО "Омега Принт"
пр. Ак. Лаврентьева, 6, 630090 Новосибирск