

## Отзыв официального оппонента

на диссертацию Петюшко Александра Александровича  
«Биграммные языки» на соискание ученой степени  
кандидата физико-математических наук  
по специальности 01.01.09 –  
дискретная математика и математическая кибернетика

**Актуальность темы.** Из всех проблем синтаксиса формальных языков вопросы разнообразия таких языков, пожалуй, самые актуальные. Поэтому исследования новых формальных языков с точки зрения классификации Хомского и на предмет объемности таких языков являются безусловно актуальным для структуризации необходимого разнообразия языков. В диссертационной работе Петюшко А.А. изучаются и классифицируются формальные языки, заданные матрицами кратностей биграмм, т.е. частот пар соседних букв слова изучаемого языка. Еще в начале 20 века выдающимся русским ученым А.А. Марковым (старшим) был создан аппарат цепей (цепей Маркова). Предсказание и распознавание при условии максимизации вероятности события – вот основное применение цепей Маркова. Последовательность данных, или событий, аппроксимируется с помощью вероятностей переходов в цепи Маркова. Аналогом вероятности переходов в дискретном случае являются частоты (или кратности) появления пар соседних элементов изучаемой последовательности. Поэтому задача описать формальный язык, в котором все слова имеют одну матрицу кратностей биграмм, является в некотором роде обратной к традиционной вероятностной задаче нахождения случайных процессов по заданной последовательности вероятностей переходов.

В пятидесятых годах XX века американским лингвистом Н. Хомским была создана т.н. иерархия формальных языков, впоследствии названная в его честь. Обычно изучение любого формального языка начинается с его классификации согласно иерархии Хомского, а именно, определения, к какому классу принадлежит язык: регулярному, контекстно-свободному, контекстно-зависимому или рекурсивно-перечислимому. Как следствие, классификация биграммных языков согласно иерархии Хомского является классической и общепринятой практикой в теории изучения формальных языков.

Начало изучению биграммных языков было положено в семидесятых годах XX века, главным образом благодаря работам по подсчёту (Дж. Хатчинсон и Х. Уилф, 1975) точного числа ДНК-последовательностей по набору кратностей биграмм и одиночных символов, а также аппроксимации этого числа (верхняя асимптотическая оценка, К. Ким и

Ф. Рауш, 1979). Однако указанные работы посвящены исключительно конечным языкам. Это позволяет заключить, что изучение бесконечных биграммных языков, в чем и состоит основная часть диссертационной работы, относится к актуальной теме исследования.

**Научная новизна.** Все теоремы в диссертации являются новыми и получены автором самостоятельно. Диссертация состоит из введения, трех глав, заключения, списка цитируемой литературы и публикаций автора. Во введении дается обзор полученных ранее результатов, относящихся к исследуемой теме, обосновывается актуальность работы, приводятся формулировки основных понятий и полученных в диссертации результатов, а также излагается краткое содержание диссертации по главам.

В первой главе введены основные понятия, касающиеся матрицы кратностей биграмм, а именно конечные (простейшие) и (потенциально) бесконечные биграммные языки. Показаны основные свойства матрицы кратностей биграмм. Приводится связь между изучаемыми языками и их графовым представлением, для чего строится ориентированный граф, в котором допускаются петли и параллельные ребра. Получены условия непустоты, конечности и счетности биграммных языков в зависимости от того, будет ли соответствующий граф полуэйлеровым или эйлеровым.

Основной результат первой главы – это классификация счетных биграммных языков согласно иерархии Н. Хомского. Выясняется, что основным условием является количество элементарных циклов в разложении соответствующего графа, а также единственность разложения на элементарные циклы. Автор показывает, что наличие одного элементарного цикла в соответствующем ориентированном графе говорит о регулярности биграммного языка, наличие двух элементарных циклов (при единственности разложения в них) говорит о контекстно-свободности биграммного языка, в остальных же случаях язык является контекстно-зависимым. Доказано, что языков общего вида (рекурсивно-перечислимых) среди биграммных языков нет.

Во второй главе изучаются различные объемные, т.е. количественные свойства биграммных языков. Так, получена точная формула мощности множества слов простейшего биграммного языка, в которой не используется знание вектора униграмм (букв языка), как это было сделано в работе Дж. Хатчинсона и Х. Уилфа.

Надо отметить, что ввиду большой вычислительной сложности полученная автором формула мало пригодна для практических расчетов, и для исправления этого недочета автором была выведена асимптотически точная оценка мощности языка, в то время как в работе К. Кима и Ф. Рауша была получена только верхняя асимптотическая оценка.

Также в данной главе автор рассмотрел вопрос о доле матриц, которые задают пустые, конечные или счётные биграммные языки; также вопрос о доле матриц, задающих регулярные, контекстно-свободные или контекстно-зависимые биграммные языки. Было доказано, что при стремлении максимального элемента матрицы кратностей биграмм к бесконечности доля матриц, порождающих пустые языки, мажорирует остальные виды матриц; при этом доля матриц, задающих конечные биграммные языки, по порядку роста максимального элемента приближенно равна доле матриц, задающих счётные биграммные языки. При этом доля матриц, задающих регулярные биграммные языки, есть «о малое» от доли матриц, задающих контекстно-свободные биграммные языки, а в свою очередь доля матриц, задающих контекстно-свободные биграммные языки, есть «о малое» от доли матриц, задающих контекстно-зависимые биграммные языки. Как результат, получаем, что доля матриц, задающих конечные биграммные языки, по порядку роста максимального элемента приближенно равна доле матриц, задающих контекстно-зависимые биграммные языки.

В третьей, заключительной, главе автор рассматривает различные расширения понятия «биграммный язык», данного им в самом начале работы. Так, вводится понятие «биграммного языка с закольцовыванием», в котором по сравнению с биграммным языком к набору биграмм, который задает биграммный язык, добавляется биграмма, состоящая из последней и первой буквы рассматриваемого слова (что означает, что слово можно считать заданным на окружности, или кольце). Автор изучил свойства нововведенных биграммных языков с закольцовыванием, а также построил биекцию между словами биграммных и биграммных языков с закольцовыванием для случая одинаковой задающей матрицы биграмм.

В этой главе была построена классификация биграммных языков с закольцовыванием согласно иерархии Н. Хомского, схожая с классификацией биграммных языков. Тем не менее, несмотря на общую схожесть понятий биграммного и биграммного с закольцовыванием языка, есть и отличия (не считая разных доказательств): не бывает конечных биграммных языков с закольцовыванием.

В заключение третьей главы автор делает своеобразное обобщение: он показывает, что разобранные выше задачи (как качественные, так и количественные) в случае  $n$ -граммных языков (в этом случае задается не двухмерная матрица биграмм, а  $n$ -мерная матрица  $n$ -грамм) при  $n > 2$  сводятся к соответствующим задачам для биграммных языков, которые автором в полной мере были уже исследованы.

**Теоретическая и практическая значимость.** Полученные результаты имеют теоретический характер и могут быть использованы в дальнейших исследованиях по изучению формальных языков, заданных матрицей кратностей биграмм. Также результаты диссертации могут быть использованы в прикладных задачах поиска похожих фрагментов данных в системах хранения, например, с помощью быстрого нахождения аналога локального хэша – матрицы кратностей биграмм.

**Обоснованность положений и выводов.** Все результаты диссертации являются новыми, сопровождаются полными математическими доказательствами, получены автором самостоятельно и отражены в публикациях автора, неоднократно докладывались на конференциях и научных семинарах.

**Критический анализ диссертации.** Отметим ряд недостатков диссертационной работы:

1. В диссертации недостаточно внимания уделяется прикладной стороне использования биграммных языков.
2. Интересно было бы проследить связь биграммного языка с его семантикой.
3. Следствие 3.3.1 (Алгоритмическая разрешимость) не следует из леммы 3.3 так как лемма описывает только достаточное условие конструктивного определения непустоты языка.
4. В диссертации имеются редакционные неточности и описки (стр. 9 и др.), термин "граф" в диссертации используется по умолчанию в смысле "мультиграфа", формулировка теорем - критериев в работе имеет непривычный вид, отличный от стандартного "тогда и только тогда" или "необходимо и достаточно".

**Заключение.** Перечисленные недостатки не сказываются на достоверности полученных результатов и общей положительной оценке содержания работы.

Полученные в диссертации результаты могут быть использованы на практике в системах хранения данных для уменьшения избыточности, на спецсеминарах по теории автоматов и формальных языков, и для дальнейших исследований в Московском университете, Новосибирском, Иркутском, Уральском, Саратовском государственном университете, Южном федеральном университете.

Считаю, что диссертация Петюшко А.А. «Биграммные языки» представляет собой законченное научное исследование на актуальную тему, удовлетворяет требованиям п.п. 9, 10, 11, 13, 14 «Положения о порядке присуждения ученых степеней ВАК Министерства

образования и науки РФ» от 24 сентября 2013 г. № 842, предъявляемым к кандидатским диссертациям. Автореферат диссертации полностью отражает содержание работы.

Автор диссертации Петюшко Александр Александрович заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 01.01.09 – дискретная математика и математическая кибернетика.

Официальный оппонент, доктор физико-математических наук,  
профессор Чечкин Александр Витальевич.

Дом. адрес: 115419, Москва, ул. Академика Петровского, д. 3, кв. 17

Тел.: +7(916)164-01-96, E-mail: [a.chechkin@mail.ru](mailto:a.chechkin@mail.ru)

Место работы: профессор кафедры «Математика» ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации».

Адрес работы: 125993, Москва, Ленинградский проспект, 49.

Тел.: +7(499)277-21-54 E-mail: [mathfa@mail.ru](mailto:mathfa@mail.ru)

Доктор физ.-мат. наук, профессор

А. В. Чечкин

25.02.2016 г

