

О Т З Ы В

официального оппонента, кандидата физико-математических наук
ХОЛОДЕНКО Александра Борисовича
на диссертацию ПЕТЮШКО Александра Александровича
«Биграммные языки», представленную на соискание ученой степени
кандидата физико-математических наук по специальности
01.01.09 — дискретная математика и математическая кибернетика.

Диссертационная работа А. А. Петюшко посвящена изучению формальных языков, заданных частотами встречаемости последовательностей символов фиксированной длины.

Актуальность темы исследования обусловлена широким распространением задач обработки информации, представленной линейными последовательностями символов, начиная от задач распознавания слитной человеческой речи и заканчивая задачами анализа последовательностей нуклеотидов при расшифровке генома. В большинстве практически важных и интересных задач объёмы информации таковы, что они практически не могут быть охвачены и проанализированы целиком, однако во многих случаях удаётся выявить локальные закономерности анализируемого потока данных. Одной из наиболее практически важных и хорошо изученных локальных особенностей потоков символов являются частоты встречаемости последовательностей символов фиксированной длины, так называемые n -граммы. Как известно, данный принцип был положен в основу аппарата марковских цепей, предложенных в начале XX века выдающимся российским учёным Андреем Андреевичем Марковым (старшим), и в настоящее время широко используется во многих задачах анализа данных. Другим важным примером использования частот последовательностей символов является огрубленное моделирование естественных языков в задачах распознавания слитной человеческой речи, используемое в настоящее время во многих системах распознавания речи.

Тем не менее, несмотря на широкое распространение исследований, связанных с использованием n -граммного подхода, подавляющая часть работ посвящена практическим аспектам, и только очень немногие работы изучают свойства таких языков как математических объектов. В работе не только

предпринята попытка изучить данные языки с точки зрения иерархии Хомского, но также рассмотрены некоторые обобщения данного подхода, связанные с рассмотрением так называемых «закольцованных» языков, то есть множеств слов, последняя буква в которых соединена с первой буквой того же слова. Таким образом, актуальность представленной диссертации не вызывает сомнений.

Диссертация состоит из введения, трёх глав, заключения и списка литературы.

Во введении даётся очень краткий обзор полученных ранее результатов, после чего описывается структура диссертации, а также вводятся основные определения и перечисляются наиболее существенные результаты, полученные автором в рамках своего диссертационного исследования.

В первой главе автор вводит понятие биграммы и биграммного языка и рассматривает вопросы существования языков, заданных набором кратностей биграмм, а также принадлежности таких языков к тем или иным классам в иерархии Хомского. В работе показано, что языков общего вида среди биграммных языков нет, а определение принадлежности к оставшимся классам языков является алгоритмически разрешимой задачей и сводится к анализу соответствующего биграммному языку мультиграфа. Более точно, для того, чтобы биграммный язык был непуст, необходимо наличие в мультиграфе, соответствующем анализируемому языку, эйлерова пути или эйлерова цикла, причём наличие эйлерова цикла также гарантирует бесконечность рассматриваемого языка. Анализ цикловой структуры мультиграфа также даёт ответ и на вопрос о расположении языка в иерархии Хомского. В случае единственного элементарного цикла соответствующий язык будет регулярным, в случае единственного разложения мультиграфа на два элементарных цикла – контекстно-свободным. Во всех остальных случаях язык будет контекстно-зависимым.

Вторая глава диссертации посвящена оценкам мощностей для биграммных языков. Основным результатом второй главы является точная формула

для вычисления мощности биграммного языка по кратностям его биграмм. Однако данная формула является слишком сложной в вычислительном плане, поэтому автор выводит также точную асимптотическую формулу для мощности.

Помимо этого, во второй главе автор исследует вопрос о мощностях различных подклассов биграммных языков, соответствующих иерархии Хомского. В работе показано, что доли конечных и бесконечных биграммных языков сравнимы, а среди бесконечных языков отношение доли более простых языков к доли более сложных (в смысле иерархии Хомского) стремится к нулю с ростом максимальной кратности биграммы.

Третья глава посвящена в первую очередь специальному расширению понятия биграммных языков, которое в работе было названо «биграммными языками с закольцовыванием». Основная идея такого обобщения – рассматривать в дополнение к обычным биграммам также биграмму, составленную из последней и первой букв слова, таким образом, как будто бы слово задано на окружности (кольце). Данное расширение обладает рядом особенностей (например, в нём не могут существовать непустые конечные языки), однако большинство результатов главы 1 могут быть также перенесены и на языки с закольцовыванием. Так, непустота языка связана с наличием эйлера цикла в соответствующем ему мультиграфе, а принадлежность к классам регулярных, контекстно-свободных или контекстно-зависимых языков – со свойствами разложения мультиграфа в сумму простейших эйлеровых циклов, повторяющими соответствующую теорему из главы 1.

В конце третьей главы автор также обобщает основные результаты своей работы на случай n -грамм с произвольным $n > 2$. Он показывает, что используя конструкцию графов де Брёйна, можно свести задачу работы с n -граммами к задаче работы с $(n-1)$ -граммами.

Следует заметить, что оппонируемая работа не лишена недостатков.

Во-первых, необходимо отметить некоторую небрежность в оформлении работы, в том числе значительное количество опечаток и стилистических

недочётов в тексте, а также отсутствие формального обозначения глав (хотя в работе идут ссылки на главы, формально глав в тексте нет).

Во-вторых, хочется отметить излишнюю краткость в изложении предыстории вопроса. Несмотря на то, что n -граммный подход широко используется в самых разных исследованиях, и, в частности, биграммные языки уже становились предметом изучения разных исследователей, это практически не нашло своего отражения в тексте диссертации.

В-третьих, представляется целесообразным добавить в текст диссертации более подробные словесные пояснения ряда названий и понятий. Так, например, на стр. 26 Лемма 1.2 называется «условием неразрывности», хотя в тексте ни до данной леммы, ни после неё нет пояснений «физического смысла» упомянутой «неразрывности».

В-четвёртых, настоящее исследование, несомненно, украсило бы приведение какого-либо примера успешного применения предложенного математического аппарата на практике.

Однако перечисленные недостатки не являются принципиальными и не снижают общую положительную оценку работы. В целом, в диссертационной работе А. А. Петюшко получены новые интересные результаты, связанные с анализом специального класса формальных языков.

Достоверность и обоснованность положений диссертации обусловлена строгим математическим доказательством всех утверждений работы.

Диссертация представляет собой завершённую научно-исследовательскую работу, содержащую решение актуальной задачи теории формальных языков. Полученные автором результаты имеют теоретическую и практическую ценность и могут быть использованы при разработке систем распознавания речи, а также в других задачах анализа символьной информации.

Основные результаты диссертации изложены в одиннадцати публикациях, семь из которых – в рецензируемых журналах, входящих в перечень журналов, рекомендованный ВАК, и доложены на ведущих международных

и всероссийских семинарах и конференциях по тематике диссертационной работы. Опубликованные работы достаточно полно отражают содержащиеся в диссертации научные результаты.

Автореферат в полной мере отражает содержание диссертации.

Считаю, что диссертация «Биграммные языки» полностью удовлетворяет всем требованиям ВАК Министерства образования и науки РФ, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 01.01.09 – «Дискретная математика и математическая кибернетика», а ее автор, Петюшко Александр Александрович, заслуживает присуждения ему ученой степени кандидата физико-математических наук.

Официальный оппонент, кандидат физико-математических наук,
Холоденко Александр Борисович.

Дом. адрес: 117246, г. Москва, ул. Обручева, д. 35, кор. 3, кв. 16.

Телефон: +7 (916) 113-80-69.

e-mail: alexander@kholodenko.ru

Место работы: ООО «Центр прикладной соционики», заместитель генерального директора.

Адрес работы: 127055, Москва г, Новослободская ул, дом 31, стр. 2,
тел. +7 (495) 979-78-16.

Канд. физ.-мат. наук

Холоденко А.Б.

26.02.2016г

Подпись Холоденко А.Б. удостоверяю.

Генеральный директор ООО

«Центр прикладной соционики»



Бескова Л.А.