S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 6, стр. 166–181 (2009) УДК 512.31
MSC 68R15

СИЛЬНО БЕСКУБНЫЕ СЛОВА И СВОБОДНАЯ БЕРНСАЙДОВА ПОЛУГРУППА С ТОЖДЕСТВОМ $x^2 = x^3$ И ДВУМЯ ОБРАЗУЮЩИМИ

А. Н. ПЛЮЩЕНКО

Abstract. In this paper we consider the free Burnside semigroup with two generators satisfying $x^2 = x^3$. Elements of this semigroup are classes of equivalent words. We prove that each such equivalence class contains at most one overlap-free word.

Keywords: group, Burnside semigroup, overlap-free word.

Введение

Свободной бернсайдовой полугруппой называется относительно свободная полугруппа с тождеством $x^n = x^{n+m}$, где $n \geq 0$, $m \geq 1$. Её элементами являются классы эквивалентных слов (строгое определение будет дано ниже). Важнейшая проблема — *проблема равенства слов* — заключается в том, чтобы по двум произвольно заданным словам определить, эквивалентны они или нет.

Проблеме равенства слов в свободных бернсайдовых полугруппах посвящено немало исследований, прояснивших многие особенности внутренней структуры классов эквивалентности (более подробно с историей проблемы можно ознакомиться в [5, 6]). К настоящему времени, благодаря работам МакКаммода [8], де Лука и Варриккио [7], до Лаго [6] и В. С. Губы [2, 3], проблему равенства слов удалось решить для случая $n \geqslant 3$ и произвольного m. Результаты работ Грина и Риса [1], Кадурека и Полака [4] и др. позволили существенно продвинуться в решении проблемы для $n = 1$ и произвольного m и свести её к проблеме равенства слов для $n = 0$ при том же самом m. Как легко

Plyushchenko, A.N., Overlap-free words and free Burnside semigroup with two generators satisfying $x^2 = x^3$.

Поступила 12 сентября 2007 г., опубликована 29 июня 2009 г.

видеть, свободные бернсайдовы полугруппы с тождеством $x^m = 1$ являются периодическими группами, которые называются свободными бернсайдовыми группами. Проблема равенства слов для таких групп является классической проблемой теорий периодических групп и ещё далека от полного решения. Наконец, для случая $n = 2$ проблема равенства слов также остается открытой. О структуре бернсайдовых полугрупп с тождеством $x^2 = x^{2+m}$ при $m \geq 2$ известно немного. До Лаго показал (см. [5]), что такие полугруппы содержат достаточно сложные подгруппы (а при $n \geqslant 3$ все подгруппы бернсайдовых полугрупп являются циклическими). Но даже в случае тождества $x^2 = x^3$, когда все подгруппы являются тривиальными, решение проблемы равенства слов пока не получено. Данная работа является шагом в направлении решения проблемы слов для случая $n = 2$ и $m = 1$, при этом мы ограничимся изучением свободной бернсайдовой полугруппы с тождеством $x^2 = x^3$ и двумя образующими.

Один из возможных путей решения проблемы равенства слов состоит в том, чтобы в каждом классе эквивалентности выбрать подходящим образом канонического представителя. Тогда решение проблемы сводится к нахождению по данному слову его представителя. Естественно выбирать в качестве канонических представителей *бескубные* слова, однако, как показал К. В. Костоусов (результат не опубликован), в одном классе эквивалентности может содержаться сколь угодно много бескубных слов. Следовательно, круг «кандидатов» в канонические представители необходимо сузить. В первую очередь, интересен вопрос о том, как распределены по классам эквивалентности *сильно бескубные* слова. Е. В. Сухановым была выдвинута гипотеза о том, что *никакие два сильно бескубных слова не эквивалентны.* Косвенно эта гипотеза подтверждалась результатами работ [9] и [10]. В [10] доказывается, что в одном классе эквивалентности не может содержаться более одного слова Туэ-Морса. В то же время, согласно [9], произвольное сильно бескубное слово в некотором смысле очень похоже на подходящее слово Туэ-Морса.

Целью данной работы является доказательство истинности сформулированной выше гипотезы.

Работа состоит из трёх разделов. В первом разделе вводятся основные определения и соглашения об обозначениях. Второй раздел содержит необходимые вспомогательные утверждения. В третьем разделе формулируется и доказывается основной результат.

1. Основные определения и обозначения

Зафиксируем двухбуквенный алфавит $\Sigma = \{a, b\}$. Слова над Σ мы будем обозначать заглавными латинскими буквами из конца алфавита: $P, Q, \ldots, Z$; пустое слово — буквой λ. Моноид всех слов над Σ обозначается через Σ^*, а полугруппа всех непустых слов — через Σ^+. Длину слова U мы будем обозначать через $|U|$, i-ую букву слова U — через $U[i]$, причём нумерация начинается с единицы, т. е. $U = U[1]U[2]\ldots U[|U|]$.

Напомним определения основных понятий. Слово Y называется *подсловом* слова U (обозначается $Y \leqslant U$), если существуют такие слова $X, Z \in \Sigma^*$, что $U = XYZ$. Подслово Y слова U называется *собственным* (обозначается $Y < U$), если при этом $XZ \neq \lambda$. В случае, когда $X \neq \lambda$ и $Z \neq \lambda$, подслово Y

будем называть *внутренним* и писать $Y \ll U$. Слово Y называется *префиксом* (*суффиксом*) слова U, если существует слово $Z \in \Sigma^*$ такое, что $U = YZ$ (соответственно, $U = ZY$). Если при этом слово Z непустое, префикс (соответственно, суффикс) называется *собственным*.

Отметим, что подслово может встречаться в слове несколько раз, и нас могут интересовать конкретные вхождения подслова. О вхождении подслова Y в слово U начиная с k-ой позиции последнего мы говорим как о «подслове Y в k-ой позиции». Слово U называется *сильно бескубным*, если оно не имеет подслов вида $XYXYX$ ни при каких $X \in \Sigma^+, Y \in \Sigma^*$.

Свободная бернсайдова полугруппа с двумя образующими с тождеством $x^2 = x^3$ является факторполугруппой полугруппы Σ^+ по соответствующей *вербальной* конгруэнции, которую мы обозначим через $\sim$. Эта конгруэнция определяется следующим образом. Два слова $U, V \in \Sigma^+$ называются *соседними*, если верно одно из следующих условий:

$$\begin{array}{ll} 1) & U = V; \\ 2) & U = XY^2Z, \quad V = XY^3Z; \\ 3) & U = XY^3Z, \quad V = XY^2Z \end{array}$$

для некоторых $X, Z \in \Sigma^*$ и $Y \in \Sigma^+$. Факт соседства слов U и V будем обозначать так: $U \longleftrightarrow V$. Конгруэнция $\sim$ является транзитивным замыканием отношения $\longleftrightarrow$. Далее слова U и V, состоящие в отношении $\sim$, мы называем эквивалентными. Класс эквивалентности слова W будем обозначать через $[W]$.

Для записи языков над Σ мы часто будем использовать рациональные (регулярные) выражения. Как обычно, итерацию языка L будем обозначать через L^*, а сумму всех положительных степеней L — через L^+.

Введём ещё два важных понятия. Зафиксируем слово U, пусть V — соседнее с U слово и $V \neq U$. Тогда V получается из U заменой вхождения Y^3 на Y^2 или Y^2 на Y^3 для некоторого Y. Это конкретное вхождение в слово U подслова Y^3 или Y^2 мы будем называть *основой* (*базой*) слова U и обозначать через $\mathrm{Base}_V(U)$. Само слово Y мы называем *производящим элементом основы* или просто *производящим* и обозначаем через $\mathrm{Gen}_V(U)$. В дальнейшем индекс V мы будем опускать и писать просто $\mathrm{Base}(U)$ или $\mathrm{Gen}(U)$, так как из контекста будет понятно, какую пару соседних слов мы рассматриваем. Далее, пусть $\mathrm{Gen}(U) = Y$. Тогда через $Y^{(1)}$ обозначим первое вхождение Y в основу, через $Y^{(2)}$ — вхождение Y в $(|Y| + 1)$-ой позиции основы, а через $Y^{(3)}$ — вхождение Y в $(2|Y| + 1)$-ой позиции основы (если основа имеет вид Y^3).

2. Вспомогательные утверждения

Прежде, чем приступить к доказательству основного результата, мы напомним некоторые важные свойства сильно бескубных слов и конгруэнции $\sim$.

Морфизм Туэ-Морса — это эндоморфизм моноида Σ^*, определяемый равенствами

$$\varphi(a) = ab, \quad \varphi(b) = ba.$$

Слова из языка $\varphi(\Sigma^*)$ будем называть *φ-образами*. Очевидно, морфизм φ инъективен. Обратное отображение из $\varphi(\Sigma^*)$ в моноид Σ^* обозначается через φ^{-1}. В дальнейшем нам понадобится следующая характеризация слов из $\varphi(\Sigma^*)$ (см., например, [9]).

Предложение 1. *Язык $\varphi(\Sigma^*)$ состоит в точности из всех слов чётной длины, в которых квадраты букв встречаются только в чётных позициях.*

Важную роль в последующих рассмотрениях играет результат работы [10]. Приведём сначала несколько определений, взятых из той же работы. Мы рассмотрим три операции редуцирования, целью которых является приведение эквивалентных слов к более простому виду путём удаления некоторых специальных подслов с сохранением отношения эквивалентности между словами.

Преобразование r_1 сокращает степени букв в слове:

$$c^k \to c^2, \quad k \geqslant 3\,. \tag{1}$$

Результат преобразования r_1 слова U обозначим через $r_1(U)$. Очевидно, значение $r_1(U)$ не зависит от порядка применения сокращений (1). Назовём слово U *r_1-редуцированным*, если $U = r_1(U)$. В [11] доказано следующее предложение.

Предложение 2. *Для любых двух слов U, V из Σ^* если $U \longleftrightarrow V$, то $r_1(U) \longleftrightarrow r_1(V)$.*

Следствие 1. *Если $U \sim V$ и U, V — r_1-редуцированы, то найдутся r_1-редуцированные слова $R_0, R_1, \ldots, R_n$ такие, что*

$$U = R_0 \longleftrightarrow R_1 \longleftrightarrow \ldots \longleftrightarrow R_{n-1} \longleftrightarrow R_n = V.$$

Таким образом, даже рассматривая отношение соседства *только* для r_1-редуцированных слов, мы можем получить все пары эквивалентных r_1-редуцированных слов. Далее класс всех r_1-редуцированных слов, эквивалентных слову W, мы будем обозначать через $[W]_{r_1}$.

Назовём слово U *буквочередующимся*, если оно не содержит квадратов букв. Пусть $\widetilde{B} = b(ab)^+$ и $\widetilde{A} = a(ba)^+$ — языки буквочередующихся слов нечётной длины (за исключением однобуквенных слов a и b). Обозначим через $r_A(U)$ слово, полученное из U применением (пока возможно) преобразования

$$a\widetilde{A}a \quad \to \quad aa, \tag{2a}$$

а через $r_B(U)$ — слово, полученное из U применением (пока возможно) преобразования

$$b\widetilde{B}b \quad \to \quad bb. \tag{2b}$$

Отметим, что мы будем применять операции r_A и r_B только к r_1-редуцированным словам. В работе [10] доказывается, что сокращения (2a) и (2b) в этом случае можно производить в любом порядке. Слово U называется *r_A-редуцированным*, если $r_A(U) = U$, и называется *r_B-редуцированным*, если $r_B(U) = U$. Наконец, слово U называется *вполне редуцированным*, если оно r_1-, r_A- и r_B-редуцировано. Примерами вполне редуцированных слов являются φ-образы. Действительно, в подсловах aaa (bbb) и $a(a(ba)^i)a = aa(b(ab)^{i-1})aa$ (соответственно, $b(b(ab)^i)b = bb(a(ba)^{i-1})bb$) позиции первого и последнего вхождения подслова aa (соответственно, bb) имеют разную чётность. В силу предложения 1, слова из $\varphi(\Sigma^*)$ не могут содержать подслов указанного вида, что и доказывает вполне редуцированность φ-образов.

Сформулируем основное техническое утверждение из работы [10].

Предложение 3. *Пусть $\varphi(W) \sim S$, где S — вполне редуцированное слово. Тогда найдётся слово V такое, что $\varphi(V) = S$ и при этом $W \sim V$.*

В работе [10] вводятся ещё два важных понятия, которые нам понадобятся в дальнейшем. Слово U назовём *$\widetilde{A}$-целым*, если для любого вхождения в U подслова вида $a\widetilde{A}a$ этому вхождению предшествует ab, а после него следует ba. Таким образом, если подслово вида $\widetilde{A}$ слова U с обеих сторон окружено буквой a, то оно входит в U вместе с подсловом $aba\widetilde{A}aba$. Двойственным образом определяется *$\widetilde{B}$-целое* слово. Примерами $\widetilde{A}$- и $\widetilde{B}$-целых слов являются вполне редуцированные слова. Следующее утверждение также взято из [10].

Предложение 4. *Пусть U и V — соседние r_1-редуцированные слова. Тогда если U — $\widetilde{A}$-целое ($\widetilde{B}$-целое) слово, то слово V также является $\widetilde{A}$-целым (соответственно, $\widetilde{B}$-целым).*

Некоторые свойства сильно бескубных слов, установленные в [9], позволят нам применять предложение 3. Мы приведём эти свойства с доказательствами, чтобы получить некоторые следствия, которые вытекают именно из доказательств этих свойств, а не из их формулировок.

Как мы видели (предложение 1), слова из $\varphi(\Sigma^*)$ если и содержат квадраты букв, то только в чётных позициях. Обобщим ситуацию. Слово U мы будем называть *регулярным*, если подслова a^2 и b^2 входят в U либо всегда в чётных, либо всегда в нечётных позициях. Буквочередующиеся слова также считаем регулярными. Несложно показать, что это определение эквивалентно следующему: слово U называется регулярным, если оно не содержит подслова c^3 и подслов вида $c^2U'd^2$, где $c, d \in \Sigma$, $U' \in \Sigma^+$ и длина U' нечётна. Очевидно, любое подслово регулярного слова само регулярно. Следующее утверждение устанавливает связь между регулярностью и сильно бескубными словами.

Предложение 5. *Любое внутреннее подслово сильно бескубного слова является регулярным.*

Доказательство. Пусть W — сильно бескубное слово и $V \ll W$. Предположим, что V нерегулярно и V' — минимальное по длине нерегулярное подслово в V. Слово W сильно бескубно, поэтому W не содержит кубов букв. Тогда по определению регулярности слово V' содержит подслово c^2Rd^2, причём $|R| = 2k-1$. Так как слово c^2Rd^2 само нерегулярно, получаем (в силу минимальности V') $V' = c^2Rd^2$. Рассмотрим два случая.

$k = 1$. Тогда $V' = c^2ed^2$, причём $e \neq c$ и $e \neq d$. Так как в алфавите всего две буквы, получаем $c = d$, откуда $V' = aabaa$ или $V' = bbabb$. Слово V' — подслово в V, а V — внутреннее подслово в W, поэтому перед V' и после него есть по крайней мере по одной букве. Но тогда либо $baabaab$, либо $abbabba$ содержится в W, что противоречит его сильной бескубности.

Пусть теперь $k > 1$. Слово R не содержит квадратов букв. В самом деле, если $R = R'e^2R''$, то одно из слов R' и R'' имело бы нечётную длину, и тогда одно из подслов $c^2R'e^2$ и $e^2R''d^2$ являлось бы нерегулярным, что противоречит минимальности V'. Таким образом, слово R буквочередующееся, причём $c \neq R[1]$ и $d \neq R[|R|]$. Поэтому слово cRd также буквочередующееся. Но $|cRd| \geqslant 5$, следовательно, W содержит подслово $ababa$ или $babab$, что снова противоречит сильной бескубности W. Предложение доказано.

Непосредственно из доказательства предложения 5 вытекает следующее утверждение.

Следствие 2. *Сильно бескубное слово W нерегулярно тогда и только тогда, когда оно начинается или оканчивается на $aabaa$ или $bbabb$.*

Теперь установим взаимосвязь регулярных слов и морфизма φ.

Предложение 6. *Если слово W регулярно, то можно приписать не более одной буквы к началу слова W и не более одной буквы к его концу так, чтобы полученное слово лежало в $\varphi(\Sigma^*)$.*

Доказательство. Если W не является буквочередующимся и все квадраты букв находятся в W на нечётных позициях, добавим в начало слова W букву, не совпадающую с $W[1]$, в противном случае оставляем W без изменений. Полученное слово обозначим через W'. Далее, если длина слова W' нечётна, добавим в конец W' букву, отличную от его последней буквы, иначе W' не изменяем. Результат обозначим через W''. Тогда по предложению 1 слово W'' лежит в $\varphi(\Sigma^*)$, что и требовалось.

Фактически в доказательстве описан алгоритм построения φ-образа из регулярного слова W. Результат всех преобразований над W, применяемых в этом алгоритме, обозначим через $\xi(W)$. Из доказательства легко получить

Следствие 3. *Пусть слово W регулярно и не является буквочередующимся. Тогда если $cWd \in \varphi(\Sigma^*)$ и $eWf \in \varphi(\Sigma^*)$, где $c, d, e, f \in \{a, b, \lambda\}$, то $c = e$ и $d = f$.*

Следующее несложное, но красивое утверждение перебрасывает «мостик» между работами [9] и [10] и служит основой для дальнейших построений.

Предложение 7. *Слово U регулярно тогда и только тогда, когда оно вполне редуцировано.*

Доказательство. Выше мы показали, что все φ-образы являются вполне редуцированными. Эти же рассуждения остаются в силе для произвольных регулярных слов, т. е. любое регулярное слово является вполне редуцированным. С другой стороны, если слово W вполне редуцировано, то регулярность W получается аналогично доказательству предложения 5. Достаточно заметить, что если V' — минимальное по длине нерегулярное подслово в W и $V' = c^2Rd^2$, то слово cRd буквочередующееся и, следовательно, V' имеет вид $a\widetilde{A}a$ или $b\widetilde{B}b$, что невозможно, так как слово W вполне редуцировано.

Предложение 7 устанавливает равносильность двух понятий, которые раньше использовались независимо друг от друга, что позволяет по-новому взглянуть на прежние результаты в этой области.

В дальнейшем нам часто будут попадаться утверждения, при доказательстве которых сначала приходится рассматривать два соседних слова, а затем по индукции распространять результат на два произвольных эквивалентных слова. Чтобы каждый раз не повторять одни и те же рассуждения, будем ссылаться на известный факт из общей алгебры.

Предложение 8. *Пусть ρ, θ — бинарные отношения на множестве A, причём θ транзитивно. Тогда если $\rho \subseteq \theta$, то $\rho^+ \subseteq \theta$. (здесь ρ^+ обозначает транзитивное замыкание бинарного отношения ρ).*

Используя этот факт, легко доказать следующее предложение.

Предложение 9. 1) *Пусть $U \sim V$ и W — префикс (суффикс) слова U, не содержащий подслов вида YY, где $Y \in \Sigma^+$. Тогда W является также префиксом (соответственно, суффиксом) слова V.*

2) *Пусть U и V r_1-редуцированы, $U \sim V$ и W — префикс (суффикс) слова U, не содержащий подслов вида YY, где $|Y| > 1$. Тогда W является и префиксом (соответственно, суффиксом) слова V.*

Доказательство. Мы будем доказывать только первое утверждение, полагая при этом, что W — префикс U; в остальных случаях доказательство аналогично. Рассмотрим сначала пару соседних слов U и V. Если $U = V$, то W, очевидно, является префиксом V. Пусть $U \neq V$. Тогда одно из слов имеет вид $XYYZ$, а другое — вид $XYYYZ$ для некоторых $X, Z \in \Sigma^*$ и $Y \in \Sigma^+$. В любом случае U и V имеют общий префикс XYY. Понятно, что из двух префиксов одного слова один (более короткий) является префиксом другого. Так как и W, и XYY — префиксы слова U, а W по условию не может содержать подслово YY, то W является префиксом слова XYY и, следовательно, является префиксом слова V. Таким образом, для пары соседних слов доказываемое утверждение верно. В силу предложения 8, оно верно для любой пары эквивалентных слов (для этого в обозначениях предложения 8 нужно в качестве ρ взять отношение соседства, а θ определить так: пара слов лежит в отношении θ, если W либо является префиксом обоих слов, либо не является префиксом ни одного из них). Предложение доказано.

3. Формулировка и доказательство основного результата

Теорема 1. *В каждом классе эквивалентности конгруэнции $\sim$ содержится не более одного сильно бескубного слова.*

Доказательство. Предположим противное: пусть U и V — два сильно бескубных слова таких, что $U \sim V$ и $U \neq V$. Применим к словам U и V следующий алгоритм.

Алгоритм A.
Шаг 0. Пусть $U_0 = U$, $V_0 = V$, $k = 0$.
Шаг 1. Если $U_k \in \{ab, ba\}$, $V_k \in \{ab, ba\}$ или хотя бы одно из слов U_k и V_k нерегулярно, прекращаем работу.
Шаг 2. Полагаем $U_{k+1} = \varphi^{-1}(\xi(U_k))$, $V_{k+1} = \varphi^{-1}(\xi(V_k))$, $k = k + 1$ и возвращаемся к шагу 1.

Идея этого алгоритма позаимствована из работы [9]. Там же доказываются три его важных свойства (см. леммы 1–3 ниже). Лемма 1 обосновывает конечность алгоритма, а в леммах 2 и 3 устанавливаются некоторые характерные особенности слов U_k и V_k.

Лемма 1. *Если все слова U_k и V_k регулярны, то найдётся такое t, что $U_t \in \{ab, ba\}$ или $V_t \in \{ab, ba\}$.*

Назовём слово W *почти сильно бескубным*, если все его собственные подслова сильно бескубны.

Лемма 2. *Все слова U_k и V_k почти сильно бескубны.*

Заметим, что леммы 1 и 2 немного отличаются от соответствующих лемм работы [9] (там алгоритм применялся только к одному сильно бескубному слову), однако доказательство их полностью аналогично.

Итак, лемма 1 гарантирует, что, обрабатывая два сильно бескубных слова U и V, алгоритм A рано или поздно остановится. Далее мы будем считать, что алгоритм завершил работу при $k = t$, причём слово U_t нерегулярно или $U_t \in \{ab, ba\}$.

Лемма 3. *Пусть слово U_t нерегулярно. Тогда возможны три варианта (с точностью до переобозначения букв a и b):*
1) $U_t = aaa$;
2) $U_t = abbabba$;
3) $U_t = aabaaU'$ *или* $U_t = U'aabaa$ *для некоторого* $U' \in \Sigma^*$.

Докажем одну вспомогательную лемму.

Лемма 4. *Пусть $U_k \sim V_k$ и $U_k \neq V_k$. Тогда оба слова U_k и V_k не являются буквочередующимися.*

Доказательство. Все буквочередующиеся слова распределены между 10 классами эквивалентных слов: $[a] = a$, $[ab] = ab$, $[aba] = aba$, $[abab] = (ab)^2(ab)^*$, $[ababa] = (ab)^2(ab)^*a$ и ещё 5 классов, соответствующих буквочередующимся словам, начинающимся с буквы b. Легко видеть, что каждый из перечисленных выше классов содержит только одно почти сильно бескубное слово, а по условию, учитывая лемму 2, таких слов должно быть не меньше двух: U_k и V_k. Поэтому U_k и V_k не могут быть буквочередующимися.

Следующая лемма является ключевой: в ней доказывается, что алгоритм A сохраняет отношение $\sim$ и отношение неравенства.

Лемма 5. *Для любого k, $0 \leqslant k \leqslant t$, выполняется $U_k \sim V_k$ и $U_k \neq V_k$.*

Доказательство. Доказывать будем индукцией по k. База ($k = 0$), очевидно, выполняется. Пусть утверждение леммы выполнено для некоторого $k < t$. Рассмотрим слова U_k и V_k. По предположению индукции $U_k \sim V_k$ и $U_k \neq V_k$. Оба слова регулярны, следовательно, вполне редуцированы (предложение 7). Пусть $\xi(U_k) = cU_kd$, где $c, d \in \{a, b, \lambda\}$. Так как $U_k \sim V_k$, то $cU_kd \sim cV_kd$. Докажем, что слово cV_kd вполне редуцировано. По определению преобразования ξ имеем $c \neq U_k[1]$ и $d \neq U_k[|U_k|]$, а в силу предложения 9 первые и последние буквы слов U_k и V_k совпадают, поэтому $c \neq V_k[1]$ и $d \neq V_k[|V_k|]$. Таким образом, подслова a^3, b^3, $a\widetilde{A}a$ и $b\widetilde{B}b$ в слове cV_kd могут встречаться только внутри V_k. Так как слово V_k вполне редуцировано, то и cV_kd вполне редуцировано.

Мы имеем следующее: $\xi(U_k) \sim cV_kd$, где cV_kd — вполне редуцированное слово, и $\xi(U_k) = \varphi(U_{k+1})$. Воспользуемся предложением 3. Существует слово V' такое, что $\varphi(V') = cV_kd$ и $V' \sim U_{k+1}$. В силу леммы 4, слова U_k и V_k не являются буквочередующимися, а отсюда, по следствию из предложения 6, получаем $cV_kd = \xi(V_k)$. Но тогда $V' = \varphi^{-1}(\xi(V_k))$, т. е. $V' = V_{k+1}$. Так как $U_k \neq V_k$, то и $cU_kd \neq cV_kd$, или $\xi(U_k) \neq \xi(V_k)$, откуда следует, что $U_{k+1} \neq V_{k+1}$. В результате получаем, что $U_{k+1} \sim V_{k+1}$ и при этом $U_{k+1} \neq V_{k+1}$. Индукцией по k лемма доказана.

Теперь мы можем уточнить ситуацию. В силу леммы 3, по окончании работы алгоритма слово U_t с точностью до переобозначения букв либо равно ab,

aaa или $abbabba$, либо имеет вид $aabaaU'$ или $U'aabaa$ для некоторого $U' \in \Sigma^*$. В первом случае слово U_t назовём «хорошим», так как можно в явном виде указать его класс эквивалентности. Следующая лемма исключает из рассмотрения все «хорошие» варианты.

Лемма 6. *Слово U_t с точностью до переобозначения букв начинается на $aabaabba$ или кончается на $abbaabaa$.*

Доказательство. Без ограничения общности можно считать, что U_t начинается на букву a. Рассмотрим все возможные случаи.

1) $U_t = ab$. Это противоречит лемме 5, так как $[ab] = ab$.

2) $U_t = aaa$. Так как слово aaa не является сильно бескубным, алгоритм A сделал хотя бы одну итерацию, т. е. $t > 0$. Тогда $\xi(U_{t-1}) = ababab$, следовательно, U_{t-1} — буквочередующееся слово, что противоречит лемме 4. Отметим, что $V_t \neq aaa$ по той же причине; из леммы 2 можно сделать важный вывод о том, что слова U_t и V_t являются r_1-редуцированными.

3) $U_t = abbabba$. Покажем, что все r_1-редуцированные слова, эквивалентные U_t, имеют вид $(abb)^2(abb)^*a$. Ввиду предложения 8 достаточно доказать, что такой вид слов сохраняется при отношении соседства. Пусть $Q \longleftrightarrow R$ и Q имеет вид $(abb)^2(abb)^*a$. Тогда производящий элемент $\mathrm{Gen}(Q)$ может начинаться на a и кончаться на b — в этом случае $\mathrm{Gen}(Q) = (abb)^i$ для некоторого $i > 0$. Производящий элемент может начинаться на b и кончаться на a — тогда $\mathrm{Gen}(Q) = (bba)^i$ $(i > 0)$. Наконец, он может начинаться на b и кончаться на b: $\mathrm{Gen}(Q) = (bab)^i$ для $i > 0$. Начинаться на a и кончаться на a производящий элемент не может, ибо тогда основа содержала бы подслово aa, которого нет в Q. Во всех трёх случаях с учётом того, что основа содержит как минимум два вхождения производящего элемента, заключаем, что слово R, соседнее с Q, имеет тот же вид $(abb)^2(abb)^*a$.

Так как слово V_t является r_1-редуцированным и $V_t \neq U_t$, получаем $V_t = (abb)^2(abb)^i a$ для некоторого $i > 0$, что противоречит лемме 2.

Таким образом, мы доказали, что слово U_t с точностью до переобозначения букв начинается или кончается на $aabaa$. Предположим, что U_t начинается на $aabaa$ (остальные случаи аналогичны). Продолжим перебор вариантов.

4) $U_t = aabaa$. Тогда $[U_t]_{r_1} = U_t$, что противоречит лемме 5.

5) $U_t = aabaab$. Аналогично случаю 3, получаем $[U_t]_{r_1} = (aab)^2(aab)^*$, т. е. $V_t = (aab)^2(aab)^i a$ для некоторого $i > 0$, противоречие с леммой 2.

6) $U_t = aabaaba$. Тогда $[U_t]_{r_1} = (aab)^2(aab)^*a$ (доказательство, как в случае 3). Поэтому $V_t = (aab)^2(aab)^i a$, где $i > 0$, что противоречит лемме 2.

7) $U_t = aabaabb$. Как и в случаях 3, 5 и 6, достаточно показать, что $[U_t]_{r_1} = (aab)^2(aab)^*b$. Действительно, пусть Q — слово указанного вида. Так как основа слова содержит как минимум два производящих элемента, а подслово bb встречается в Q всего один раз (в конце), то производящий элемент не содержит bb и последняя буква слова Q не входит в основу. Таким образом, эта буква не влияет на отношение соседства; «зачеркнув» её, мы попадаем в случай 5.

Таким образом, U_t должно начинаться на $aabaa$ и $|U| \geqslant 8$. Поэтому, ввиду леммы 2, слово U_t необходимо начинается на $aabaabba$, что и требовалось доказать.

В дальнейшем слова $(aab)^*(aab)^2ba$ и $(bba)^*(bba)^2ab$ мы будем называть *левыми хвостами*, а слова $ab(baa)^2(baa)^*$ и $ba(abb)^2(abb)^*$ — *правыми хвостами.*

Хвосты $(aab)^*(aab)^2ba$ и $ab(baa)^2(baa)^*$ мы также будем называть A-*хвостами*, а $(bba)^*(bba)^2ab$ и $ba(abb)^2(abb)^*$ — B-*хвостами*. Наконец, слова, имеющие в качестве префикса левый хвост или в качестве суффикса правый хвост, будут называться *хвостатыми*. Наша задача состоит в том, чтобы «сократить» хвосты слов U_t и V_t, сохраняя при этом эквивалентность, после чего к «подправленным» U_t и V_t можно будет снова применить алгоритм A.

Предварительно разобьём все хвостатые слова на четыре группы.

К группе G_1 отнесём все слова, имеющие только левый хвост, т. е. слова вида $(aab)^*(aab)^2baW$ или $(bba)^*(bba)^2abW$, не содержащие $ab(baa)^2(baa)^*$ и $ba(abb)^2(abb)^*$ в качестве суффикса. В группу G_2 попадут все слова, имеющие только правый хвост. К группе G_3 отнесём все слова, имеющие и A-хвост, и B-хвост. Наконец, группа G_4 будет состоять из всех слов с двумя A-хвостами либо двумя B-хвостами.

Ввиду симметрии, достаточно рассмотреть только слова групп G_1, G_3 и G_4, начинающиеся на A-хвост. Начнем с утверждения о словах с левыми хвостами.

Лемма 7. *Пусть W_1, W_2 — два соседних r_1-редуцированных слова, $W_1 = P_1Q_1$, где $P_1 = (aab)^ka$ для некоторого $k \geqslant 0$, а $Q_1 = abaabbaR_1$, $R_1 \in \Sigma^*$, причём слово Q_1 является $\widetilde{A}$-целым. Тогда $W_2 = P_2Q_2$, где*

1) *$P_2 = (aab)^la$, $Q_2 = abaabbaR_2$ при некоторых $l \geqslant 0$ и $R_2 \in \Sigma^*$;*
2) *слово Q_2 является $\widetilde{A}$-целым;*
3) *$Q_1 \longleftrightarrow Q_2$.*

Доказательство. Докажем сначала пункты 1 и 3 леммы. Посмотрим, как может располагаться основа слова W_1. Для этого нам понадобится следующее представление слова W_1: $W_1 = ST$, где $S = (aab)^{k+2}$, а $T = baR_1$. Рассмотрим возможные случаи.

1) $\mathrm{Base}(W_1)$ целиком лежит в S. Тогда $W_2 = S'T$, где $S' \longleftrightarrow S$. Слово S эквивалентно $aabaab$, а $[aabaab]_{r_1} = (aab)^2(aab)^*$ (лемма 6, случай 5). Следовательно, существует $l \geqslant 0$ такое, что $S' = (aab)^{2+l}$. Положив $P_2 = (aab)^la$, $Q_2 = Q_1$, получим требуемое.

2) $\mathrm{Base}(W_1)$ целиком лежит в T. Тогда $\mathrm{Base}(W_1)$ целиком лежит в Q_1. Поэтому, очевидно, $W_2 = P_1Q_2$, где $Q_2 \longleftrightarrow Q_1$.

3) $\mathrm{Base}(W_1)$ начинается в S, а кончается в T. Пусть $Y = \mathrm{Gen}(W_1)$. Возможны два варианта расположения подслова $Y^{(1)}$ в слове W_1.

а) $Y^{(1)}$ начинается в S и кончается в S. Тогда подслово bb не встречается в Y, но встречается в основе, так как S кончается на b, а T на b начинается:

$$\underbrace{(aa\,\overbrace{b)^{k+1}aa(b}^{}}_{S}\overbrace{\underbrace{b)a\ldots\ldots}_{T}}^{\mathrm{Base}(W_1)}.$$

Такое возможно, только если bb образуется на стыке вхождений производящего элемента в основу, т. е. когда первая и последняя буквы Y есть b. Но $Y^{(1)}$ — подслово S, а $Y^{(1)}Y^{(2)}[1]$ — нет, так как оканчивается на bb. Отсюда следует, что $Y^{(1)}$ — суффикс S, а $Y^{(2)}$ — префикс T. Таким образом, $Y = (baa)^nb$ для некоторого $n \geqslant 1$ (в r_1-редуцированных словах производящий элемент не может быть однобуквенным).

Если $n > 1$, то слово Y имеет префикс $baabaa$. Так как $Y^{(2)}$ — префикс T, то и T начинается на $baabaa$. Но S кончается на букву b, поэтому Q_1 содержит bT.

Таким образом, в Q_1 входит подслово $bbaabaa$, т. е. подслово вида $bba\widetilde{A}a$, что противоречит тому, что Q_1 — $\widetilde{A}$-целое слово.

Если же $n = 1$, то $Y = baab$, $Y^{(1)}$ — суффикс S длины 4. Следовательно, вся основа целиком содержится в Q_1:

$$\underbrace{\overbrace{(aab)^k a}^{P_1}\overbrace{a(baab)baR}^{Q_1}}_{S}\ .$$

Поэтому $W_2 = P_2Q_2$, где $P_2 = P_1$, а $Q_2 \longleftrightarrow Q_1$.

б) $Y^{(1)}$ начинается в S, а кончается в T. Тогда Y начинается на некоторый суффикс слова S. Если длина этого суффикса больше 5 символов, то $aabaabb \leqslant Y$ ($aabaab$ — суффикс S, а последняя буква b — начало T). Но тогда $aabaabb \leqslant Y^{(2)} \leqslant Q_1$. Снова получили противоречие с тем, что Q_1 — $\widetilde{A}$-целое. Следовательно, длина суффикса S, покрываемого $Y^{(1)}$, не превосходит 5 и весь этот суффикс целиком попадает в Q_1. Опять получаем, что $W_2 = P_1Q_2$, где $Q_2 \longleftrightarrow Q_1$.

Тем самым мы доказали следующее: существуют такие слова P_2 и Q_2, что $W_2 = P_2Q_2$, $P_2 = (aab)^l a$ для некоторого $l \geqslant 0$, и $Q_2 \longleftrightarrow Q_1$. Из предложения 4 вытекает, что слово Q_2 также $\widetilde{A}$-целое. Наконец, из п. 2 предложения 9 следует, что Q_2 начинается на $abaabba$. Лемма доказана.

Для удобства ссылок сформулируем без доказательства симметричное лемме 7 утверждение для слов с правыми хвостами.

Лемма 8. *Пусть W_1, W_2 — два соседних r_1-редуцированных слова, $W_1 = Q_1P_1$, где $P_1 = a(baa)^k$ для некоторого $k \geqslant 0$, а $Q_1 = R_1abbaaba$, $R_1 \in \Sigma^*$, причём слово Q_1 является $\widetilde{A}$-целым. Тогда $W_2 = Q_2P_2$, где*

1) *$P_2 = a(baa)^l$, $Q_2 = R_2abbaaba$ при некоторых $l \geqslant 0$ и $R_2 \in \Sigma^*$;*
2) *слово Q_2 является $\widetilde{A}$-целым;*
3) *$Q_1 \longleftrightarrow Q_2$.*

Изучим теперь слова из группы G_3, т. е. с двумя разными хвостами.

Лемма 9. *Пусть W_1, W_2 — два соседних r_1-редуцированных слова, $W_1 = P_1Q_1R_1$, где $P_1 = (aab)^{k_1}a$, $R_1 = b(abb)^{k_2}$ для некоторых $k_1 \geqslant 0$ и $k_2 \geqslant 0$, а Q_1 имеет префикс $abaabba$, суффикс $baabbab$ и является $\widetilde{A}$- и $\widetilde{B}$-целым. Тогда $W_2 = P_2Q_2R_2$, где*

1) *$P_2 = (aab)^{l_1}a$, $R_2 = b(abb)^{l_2}$ при некоторых $l_1, l_2 \geqslant 0$, а слово Q_2 имеет префикс $abaabba$ и суффикс $baabbab$;*
2) *слово Q_2 является $\widetilde{A}$- и $\widetilde{B}$-целым;*
3) *$Q_2 \longleftrightarrow Q_1$.*

Доказательство. Так же, как и в лемме 7, нам достаточно показать, что $W_2 = P_2Q_2R_2$, где $P_2 = (aab)^{l_1}a$, $R_2 = b(abb)^{l_2}$ и $Q_1 \longleftrightarrow Q_2$. Тогда из предложения 4 вытекает, что слово Q_2 является $\widetilde{A}$- и $\widetilde{B}$-целым, а из п. 2 предложения 9 следует, что Q_2 имеет префикс $abaabba$ и суффикс $baabbab$.

Рассмотрим расположение основы слова W_1. Если $\mathrm{Base}(W_1)$ целиком лежит в P_1Q_1 или в Q_1R_1, мы находимся в условиях леммы 7 или, соответственно, леммы 8, поэтому доказываемое утверждение справедливо. Остаётся единственный вариант: основа начинается в P_1, а кончается в R_1. Покажем, что это невозможно. Действительно, в этом случае $Q_1 \ll \mathrm{Base}(W_1)$, в частности,

$\mathrm{Base}(W_1)$ содержит подслова aa и bb. Ясно, что хотя бы одно из этих подслов содержится в производящем элементе. Пусть $Y = \mathrm{Gen}(W_1)$ и $bb \leqslant Y$ (случай $aa \leqslant Y$ разбирается аналогично). Тогда в Y входит подслово $aabaabb$, так как $Y^{(1)}$ начинается в P_1, а первое вхождение bb в W_1 находится в слове Q_1:

$$\underbrace{(aa\,b)^{k_1}}_{P_1}\underbrace{\overbrace{aabaabb}^{Y^{(1)}}\ldots\ldots b}_{Q_1}\underbrace{b(abb)^{k_2}}_{R_1}\ .$$

Получаем $aabaabb \leqslant Y^{(2)} \leqslant Q_1R_1$. Слово Q_1 имеет суффикс bab, поэтому Q_1R_1 заканчивается на $babb(abb)^{k_2}$. Легко видеть, что подслово $aabaabb$ может пересекаться с этим суффиксом только по одной букве b — последней для $aabaabb$ и первой для $babb(abb)^{k_2}$. Таким образом, $aabaabb$ целиком лежит в Q_1, а это противоречит тому, что Q_1 — $\widetilde{A}$-целое слово. Полученное противоречие доказывает лемму.

Осталась не рассмотренной только группа G_4 (слова с двумя однотипными хвостами). Здесь ситуация немного сложнее: один случай придётся рассматривать отдельно. Мы это сделаем позднее.

Лемма 10. *Пусть W_1, W_2 — два соседних r_1-редуцированных слова, $W_1 = P_1Q_1R_1$, где $P_1 = (aab)^{k_1}a$, $R_1 = a(baa)^{k_2}$ для некоторых $k_1 \geqslant 0$ и $k_2 \geqslant 0$, а слово Q_1 имеет префикс $abaabba$, суффикс $abbaaba$, не равно $abaabbaaba$ и является $\widetilde{A}$-целым. Тогда $W_2 = P_2Q_2R_2$, где*

1) *$P_2 = (aab)^{l_1}a$, $R_2 = a(baa)^{l_2}$ при некоторых $l_1, l_2 \geqslant 0$, а слово Q_2 имеет префикс $abaabba$ и суффикс $abbaaba$;*
2) *слово Q_2 является $\widetilde{A}$-целым;*
3) *$Q_1 \longleftrightarrow Q_2$.*

Доказательство. Как и в предыдущей лемме, достаточно доказать, что $P_2 = (aab)^{l_1}a$, $R_2 = a(baa)^{l_2}$ и $Q_2 \longleftrightarrow Q_1$. Если $\mathrm{Base}(W_1)$ целиком попадает в P_1Q_1 или Q_1R_1, доказываемое утверждение справедливо в силу лемм 7, 8. Покажем, что оставшийся случай, в котором $Q_1 \ll \mathrm{Base}(W_1)$, невозможен. В этом случае $\mathrm{Base}(W_1)$ содержит bb, так как $bb \leqslant Q_1$. Пусть $Y = \mathrm{Gen}(W_1)$. Если $bb \leqslant Y$, то, рассуждая так же, как и в лемме 9, получаем $aabaabb \leqslant Q_1$, что противоречит тому, что слово Q_1 — $\widetilde{A}$-целое. Если $bb \nleqslant Y$, то Y начинается и кончается на b, при этом все подслова bb в Q_1 образуются на стыке вхождений производящего элемента. Из условия леммы следует, что Q_1 содержит не менее двух вхождений bb, поэтому $\mathrm{Base}(W_1) = Y^3$, Y имеет вид $(baa)^2(baa)^*b$, причем вхождение $Y^{(2)}$ целиком лежит в Q_1:

$$\underbrace{(aa\,b)^{k_1}}_{P_1}\underbrace{aabaab\overbrace{b\ldots b}^{Y^{(2)}}baaba}_{Q_1}\underbrace{a(b\,aa)^{k_2}}_{R_1}\ .$$

Таким образом, на стыке $Y^{(1)}$ и $Y^{(2)}$ в Q_1 содержится подслово $bbaabaa$, что противоречит тому, что Q_1 — $\widetilde{A}$-целое слово. Полученное противоречие завершает доказательство леммы.

Теперь определим операцию *редуцирования хвостов* для произвольного слова W следующим образом. Если $W \in G_1$, то в W удаляется первая буква, если

$W \in G_2$ — то последняя. Если $W \in G_3 \cup G_4$, то удаляются и первая, и последняя буквы W. В остальных случаях слово оставляется без изменений. Результат операции обозначим через $r_T(W)$.

Операция сокращения общего префикса (или суффикса) в общем случае не сохраняет отношение эквивалентности между словами. Однако, как показывает следующая лемма, в нужном нам случае эквивалентность сохраняется.

Лемма 11. *Пусть* $U_t \not\sim aabaabbaabaa$ *и* $U_t \not\sim bbabbaabbabb$. *Тогда оба слова* U_t *и* V_t *лежат в одной группе* G_1, G_2, G_3 *или* G_4, *при этом выполняется* $r_T(U_t) \sim r_T(V_t)$ *и* $r_T(U_t) \neq r_T(V_t)$.

Доказательство. Предположим, что U_t принадлежит группе G_1, причём U_t начинается на $aabaabba$. В силу леммы 2 слова U_t и V_t являются r_1-редуцированными. U_t можно записать в виде $[aab]^0 a[abaabbaW]$ для некоторого $W \in \Sigma^*$. Пусть $P_1 = (aab)^0 a$, $Q_1 = abaabbaW$. Поскольку $Q_1 < U_t$, по лемме 2 слово Q_1 сильно бескубно, а в силу следствия из предложения 5 оно ещё и регулярно. Ввиду предложения 7, любое регулярное слово является вполне редуцированным, в частности, является $\widetilde{A}$-целым. Таким образом, все условия леммы 7 выполнены. Осталось вспомнить, что по предложению 8 всё, что утверждается в лемме 7 про отношение соседства, переносится и на отношение эквивалентности $\sim$. Поэтому $V_t = [aab]^l aQ_2$, где $Q_2 \sim Q_1$, при этом Q_2 начинается на $abaabba$. Тогда V_t начинается на $(aab)^{l+2}ba$, откуда в силу леммы 2 находим $l = 0$. Получаем $U_t = aQ_1$, $V_t = aQ_2$ и $Q_1 \sim Q_2$. Так как $U_t \neq V_t$, то $Q_1 \neq Q_2$. Слова U_t и V_t лежат в группе G_1, поэтому $r_T(U_t) = Q_1$ и $r_T(V_t) = Q_2$. Таким образом, $r_T(U_t) \sim r_T(V_t)$ и $r_T(U_t) \neq r_T(V_t)$, что и требовалось доказать.

Случаи, когда U_t лежит в группах G_2, G_3 и G_4, разбираются аналогично, только вместо леммы 7 нужно применять соответственно леммы 8, 9 и 10. При этом условия $U_t \not\sim aabaabbaabaa$ и $U_t \not\sim bbabbaabbabb$ при $U_t \in G_4$ гарантируют, что $Q_1 \not\sim abaabbaaba$ и $Q_1 \not\sim babbaabbab$. Это необходимо для того, чтобы можно было распространить утверждение леммы 10 (с помощью предложения 8) не только на соседние, но и на произвольные эквивалентные слова.

Далее мы покажем, что случай $U_t \sim aabaabbaabaa$, «вычеркнутый» в двух предыдущих леммах, невозможен (вариант $U_t \sim bbabbaabbabb$ рассматривается симметрично). Пусть $\widetilde{A}_2 = (aab)^+ aa$, $\widetilde{A}_{02} = (aab)^* aa$. Слово $(aab)^i aa$ $(i \geqslant 0)$ будем обозначать через $\widetilde{A}_2\{i\}$. Сформулируем заключительную лемму нашего доказательства.

Лемма 12. *Язык всех* r_1*-редуцированных слов, эквивалентных слову* $aabaabbaabaa$, *равен*

$$\widetilde{A}_2 bb(\widetilde{A}_{02} bb)^* \widetilde{A}_2\,. \tag{3}$$

Доказательство. Докажем сначала, что все слова вида (3) эквивалентны слову $aabaabbaabaa = \widetilde{A}_2\{1\}bb\widetilde{A}_2\{1\}$. Действительно, пусть

$$W = \widetilde{A}_2\{i_0\}bb\widetilde{A}_2\{i_1\}bb \ldots \widetilde{A}_2\{i_n\}bb\widetilde{A}_2\{i_{n+1}\},$$

где все $i_k \geqslant 0$ $(1 \leqslant k \leqslant n)$, а $i_0, i_{n+1} > 0$. Заметим, что $\widetilde{A}_2\{l\} \sim \widetilde{A}_2\{2\}$ при $l \geqslant 2$. Далее, слово $\widetilde{A}_2\{1\}$ является собственным подсловом в W, поэтому (в силу r_1-редуцированности) для каждого вхождения $\widetilde{A}_2\{1\}$ либо перед ним, либо после него в W стоит буква b. Но $\widetilde{A}_2\{1\}b = (aab)^2 \sim (aab)^3 = \widetilde{A}_2\{2\}b$ и, аналогично, $b\widetilde{A}_2\{1\} \sim b\widetilde{A}_2\{2\}$. Таким образом, в слове W все вхождения $\widetilde{A}_2\{l\}$ при $l >$

0 можно заменить (сохраняя эквивалентность) на $\widetilde{A}_2\{1\}$. Поэтому мы будем считать, что все i_k равны 0 или 1, при этом $i_0 = i_{n+1} = 1$. Далее, преобразуем слово $\widetilde{A}_2\{1\}bb\widetilde{A}_2\{1\}$:

$$\widetilde{A}_2\{1\}bb\widetilde{A}_2\{1\} \sim \widetilde{A}_2\{2\}bb\widetilde{A}_2\{2\} = (aabaabaa)bb(aabaabaa) =$$
$$aa(baabaab)(baabaab)aa \longleftrightarrow aa(baabaab)(baabaab)(baabaab)aa =$$
$$\widetilde{A}_2\{2\}bb\widetilde{A}_2\{1\}bb\widetilde{A}_2\{2\} \sim \widetilde{A}_2\{1\}bb\widetilde{A}_2\{1\}bb\widetilde{A}_2\{1\}\,.$$

Мы показали, что от $aabaabbaabaa$ мы всегда сможем перейти (не нарушая эквивалентности) к слову вида $\widetilde{A}_2\{1\}bb(\widetilde{A}_2\{1\}bb)^*\widetilde{A}_2\{1\}$. Чтобы получить W, осталось в нужных местах вставить слово $\widetilde{A}_2\{0\}bb = aabb$. Это легко сделать:

$$\widetilde{A}_2\{1\}bbaab = (aabaa)bbaab =$$
$$aa(baab)(baab) \longleftrightarrow aa(baab)(baab)(baab) = \widetilde{A}_2\{1\}bb\widetilde{A}_2\{0\}bbaab\,.$$

Таким образом, после подслова $\widetilde{A}_2\{1\}bb$ можно вставить любое количество подслов $\widetilde{A}_2\{0\}bb$. Тем самым мы доказали, что слово W эквивалентно $aabaabbaabaa$.

Перейдем к доказательству обратного включения, т. е. того факта, что любое слово, эквивалентное $aabaabbaabaa$, имеет вид (3). В силу предложения 8 достаточно показать, что если слово V_1 имеет вид (3), то и любое соседнее с ним слово имеет тот же вид. Пусть $V_1 \longleftrightarrow V_2$ и $V_1 \neq V_2$. Рассмотрим соседние слова $W_1 = bV_1b$ и $W_2 = bV_2b$. Доказав, что W_2 имеет вид $b\widetilde{A}_2b(b\widetilde{A}_{02}b)^*b\widetilde{A}_2b$, мы получим требуемое условие для V_2.

Далее слова $b\widetilde{A}_2b$ и $b\widetilde{A}_{02}b$ будем называть *блоками*. Так как W_1 по условию не имеет ни префикса $baabb$, ни суффикса $bbaab$, то, по предложению 9, таких префикса и суффикса нет и у слова W_2. Поэтому достаточно показать, что W_2 имеет вид
$$(b\widetilde{A}_{02}b)^2(b\widetilde{A}_{02}b)^*\,. \tag{4}$$

Пусть одно из слов W_1 и W_2 равно $XYYZ$, а другое — $XYYYZ$, где X, $Z \in \Sigma^*$ и $Y \in \Sigma^+$. Тогда $X = (b\widetilde{A}_{02}b)^kP$, $Z = S(b\widetilde{A}_{02}b)^l$ для некоторых k, $l \geqslant 0$, причём P является префиксом, а S — суффиксом блока. Рассмотрим три случая.

1) Y содержит bb. Тогда $Y = Q(b\widetilde{A}_{02}b)^mR$, где $m \geqslant 0$, Q — суффикс блока, а R — префикс блока. При этом слова PQ, RQ и RS либо равны λ, либо образуют целые блоки:

$$W_1 = \underbrace{b\ldots\ldots b|P}_{X}\underbrace{Q|b\ldots\ldots b|R}_{Y}\underbrace{Q|b\ldots\ldots b|R}_{Y}\underbrace{S|b\ldots\ldots b}_{Z}$$

или

$$W_1 = \underbrace{b\ldots\ldots b|P}_{X}\underbrace{Q|b\ldots\ldots b|R}_{Y}\underbrace{Q|b\ldots\ldots b|R}_{Y}\underbrace{Q|b\ldots\ldots b|R}_{Y}\underbrace{S|b\ldots\ldots b}_{Z}\,.$$

Здесь вертикальная черта служит разделителем блоков. Отсюда видно, что если одно из слов $XYYZ$ и $XYYYZ$ имеет вид $(b\widetilde{A}_{02}b)^*$, то и другое слово имеет тот же вид. Ясно, что W_2 будет содержать не меньше двух блоков: если $Q \neq \lambda$, то это блоки RQ и PQ, если $R \neq \lambda$, то это блоки RQ и RS, наконец, если $RQ = \lambda$, то Y состоит из целого (ненулевого) числа блоков, а вхождений подслова Y в W_2 как минимум два. Поэтому слово W_2 имеет вид (4).

2) Y не содержит bb, но bb входит в $\mathrm{Base}(W_1)$. Подслово bb могло образоваться в основе слова W_1 только на стыке вхождений Y, когда Y начинается и кончается на b: $Y = bQb$. С другой стороны, подслово bb в W_1 образуется только на стыке блоков. Следовательно, Y является одновременно суффиксом одного блока и префиксом другого, поэтому Q имеет вид $\widetilde{A}_{02}$, т. е. Y является блоком. Если $W_1 = XYYZ$, получаем:

$$W_1 = \underbrace{b\ldots\ldots b|P}_{X}\,\underbrace{bQb}_{Y}\,|\,\underbrace{bQb}_{Y}\,\underbrace{S|b\ldots\ldots b}_{Z}\,.$$

Тогда

$$W_2 = \underbrace{b\ldots\ldots b|P}_{X}\,\underbrace{bQb}_{Y}\,|\,\underbrace{bQb}_{Y}\,|\,\underbrace{bQb}_{Y}\,\underbrace{S|b\ldots\ldots b}_{Z}\,.$$

Таким образом, появляется новый блок. Если же $W_1 = XYYYZ$, то наоборот, один блок удаляется. В обоих случаях W_2 будет иметь вид (4).

3) $\mathrm{Base}(W_1)$ не содержит bb, а значит, целиком лежит в одном блоке. Легко проверить, что любое r_1-редуцированное слово, эквивалентное блоку, тоже является блоком. Поэтому W_2 имеет вид (4).

Лемма доказана.

Следствие 4. *$U_t \not\sim aabaabbaabaa$.*

Доказательство. Пусть $U_t \sim aabaabbaabaa$. Тогда по лемме 12

$$U_t = \widetilde{A}_2\{i_0\}bb\widetilde{A}_2\{i_1\}\ldots bb\widetilde{A}_2\{i_n\}bb\widetilde{A}_2\{i_{n+1}\},$$

где $i_k \geqslant 0$ $(1 \leqslant k \leqslant n)$, $i_0 > 0$ и $i_{n+1} > 0$. В силу леммы 2, все i_k равны 0 или 1. Пусть $n > 0$. Если $i_1 = 1$, то $(baa)^2b = bA_2\{i_1\}b < U_t$, что невозможно по лемме 2. Если же $i_1 = 0$, то $(aabb)^2aa < A_2\{i_0\}bbA_2\{i_1\}bbA_2\{i_2\}$, что невозможно в силу той же леммы 2.

Следовательно, $n = 0$. Таким образом, U_t с необходимостью равно $aabaabbaabaa$. Но те же самые рассуждения справедливы и для слова V_t, эквивалентного U_t. Получаем $U_t = V_t$, что противоречит лемме 5.

Теперь модифицируем алгоритм A на основании лемм 6 и 11.

Алгоритм A′.

Шаг 0. $U_0 = U$, $V_0 = V$, $k = 0$.

Шаг 1. Если слово $r_T(U_k)$ или слово $r_T(V_k)$ либо равно ab или ba, либо нерегулярно, завершаем работу.

Шаг 2. $U_{k+1} = \varphi^{-1}(\xi(r_T(U_k)))$, $V_{k+1} = \varphi^{-1}(\xi(r_T(V_k)))$, $k = k+1$; возвращаемся к шагу 1.

Понятно, что леммы 1–6 остаются в силе. Таким образом, обрабатывая два неравных эквивалентных сильно бескубных слова, алгоритм A' остановится (при $k = t$). По определению алгоритма он не может остановиться на хвостатых словах U_t и V_t, ибо если U_t и V_t лежат в группе G_1, G_2, G_3 или G_4, то $r_T(U_t)$ и $r_T(V_t)$ — регулярные слова, не равные ab и ba, т. е. условия прекращения работы (шаг 1) не выполняются. Однако для слов, не имеющих хвостов, условия остановки в алгоритмах A и A' совпадают. Следовательно, согласно лемме 6, алгоритм A' не может завершить работу на словах U_t и V_t без хвостов. Поскольку алгоритм завершает работу для любой пары сильно бескубных слов, значит, пары неравных эквивалентных сильно бескубных слов не существует. Доказательство теоремы завершено.

Отметим, что полученный результат — это лишь один из шагов на пути решения проблемы равенства слов. Во-первых, существуют классы, не содержащие ни одного сильно бескубного слова, такие, например, как класс $[ababa]$. Возникает вопрос: какие слова могли бы претендовать на роль канонических представителей таких классов?

Во-вторых, полученный результат не даёт ответа на вопрос, каким образом по произвольно заданному слову находить эквивалентное ему сильно бескубное или определять, что таких нет.

Имеющийся комбинаторный аппарат дает неплохие шансы построить алгоритм, отвечающий на второй вопрос (в существовании этого алгоритма автор не сомневается), что позволит решать проблему равенства для весьма широкого класса пар слов.

Автор выражает признательность А. В. Клепинину за редактирование первоначального варианта текста статьи, профессору Л. Н. Шеврину за ряд ценных замечаний и А. М. Шуру, благодаря которому статья приняла свой окончательный облик. Автор с благодарностью вспоминает Е. В. Суханова, который оказал большое влияние на круг научных интересов автора и познакомил его с этой интересной проблемой.

Список литературы

[1] Бакиров М. Ф., Суханов Е. В., *Слова Туэ-Морса и $\mathcal{D}$-строение свободной бернсайдовой полугруппы*, Известия Уральского гос. университета, серия «Математика и механика», **18**: 3 (2000), 5–19.

[2] Лаллеман Ж., *Полугруппы и комбинаторные приложения*, М.: Мир, 1985.

[3] Green J. A., Rees D., *On semigroups in which $x^r = x$*, Proc. Cambridge Philos. Soc., **48** (1952), 35–40.

[4] Guba V. S., *The word problem for the relatively free semigroups satisfying $t^m = t^{m+n}$ with $m \geqslant 3$*, International Journal of Algebra and Computation, **2**: 3 (1993), 335–348.

[5] Guba V. S., *The word problem for the relatively free semigroups satisfying $t^m = t^{m+n}$ with $m \geqslant 4$ or $m \geqslant 3$, $n = 1$*, International Journal of Algebra and Computation, **2**: 2, 125–140.

[6] Kaďourek L., Polák J., *On free semigroups satisfying $x^r \simeq x$*, Simon Stevin, **64**: 1 (1990), 3–19.

[7] Do Lago A. P., *Maximal groups in free Burnside semigroups*, In C. L. Lucchesi and A. V. Moura, editors, LATIN'98, Lecture Notes in Computer Science, , Berlin, Springer-Verlag, **1380** (1998), 70–81.

[8] Do Lago A. P., *On the Burnside semigroups $x^n = x^{n+m}$*, International Journal of Algebra and Computation, **6**: 2 (1996), 179–227.

[9] de Luca A., Varricchio S., *On non-counting regular classes*, Theoretical Computer Science, **100** (1992), 67–104.

[10] McCammond J., *The solution to the word problem for the relatively free semigroups satisfying $t^a = t^{a+b}$ with $a \geqslant 6$*, International Journal of Algebra and Computation, **1** (1991), 1–32.

[11] Shur A. M., *Overlap-free words and Thue-Morse sequences*, International Journal of Algebra and Computation, **6**: 3 (1996), 353–367.

Андрей Николаевич Плющенко
Уральский государственный университет им. А. М. Горького,
пр. Ленина 51,
620083, Екатеринбург, Россия
E-mail address: mathplush@yandex.ru