S©MR

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ
Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru
ISSN 1813-3304

Том 22, № 1, стр. 343-353 (2025) УДК 510.567
https://doi.org/10.33048/semi.2025.22.023 MSC 03D05

БАЗОВЫЕ КАТЕГОРИАЛЬНЫЕ ГРАММАТИКИ С ОГРАНИЧЕНИЕМ НА КОЛИЧЕСТВО ПРИСВАИВАЕМЫХ КАТЕГОРИЙ

М.Е. Вишникин

Представлено В.В. Пржиялковским

Abstract: A categorial grammar is a classical formalism, in which each symbol of the alphabet is assigned one of several syntactic categories. The category of a string is then reduced from the categories assigned to its symbols, using two reduction rules. This paper investigates a special class of categorial grammars, in which the number of categories assigned to each symbol is bounded by arbitrary constant k and categories have a special form. It is proved that an infinite hierarchy exists in the class of languages defined by such grammars. Subsequently, the case of $k = 2$ is studied.

Keywords: formal grammars, categorial grammars, unique category assignment, hierarchy of rigid languages

1 Введение

Категориальная грамматика — понятие, восходящее к работам Айдукевича [1] и позднее Бар-Хиллела [2]. Известно, что данный формализм порождает в точности класс всех контекстно-свободных языков без пустого слова [3].

M.E. Vishnikin, Basic categorial grammars with restrictions on the number of assignments.

Поступила 15 октября 2024 , опубликована 15 апреля 2025.

Рассмотрение условия ограничения на количество присваиваемых категорий одному символу встречалось ранее, в особенности стоит отметить работы [4] и [5], в которых рассмотрены “rigid grammars” и “k-valued” грамматики, в которых одному символу присвоена не более чем одна или не более чем k категорий соответственно. Основной упор рассмотрения в данных работах делается на деревья вывода слов, и малое внимание уделяется свойствам порождаемых языков, в частности, большинство нетривиальных алгоритмических свойств данных классов остаются неизвестными. Как следствие, естественной является попытка рассмотреть более узкий класс грамматик.

Одним из вариантов сужения является рассмотрение подкласса категориальных грамматик, полученных ограничением на синтаксические типы, которые могут быть назначены одному символу. Ранее автором было предложено понятие базовой категориальной грамматики [6], сокращенно БКГ, отличающейся от категориальных грамматик отсутствием назначений “сложных” типов. Несмотря на это, БКГ обладают такой же выразительной возможностью, как и обычные категориальные грамматики, а именно, они могут генерировать все контекстно-свободные языки без пустого слова.

Для данных грамматик естественным образом вводится условие однозначности присвоения категорий. В отличие от ограничения на присваиваемые типы, это ограничение значительно уменьшает класс порождаемых языков, что также было доказано в [6].

В данной работе представлено естественное обобщение класса БКГ с однозначным присвоением категорий. Вводится класс $\mathcal{G}^k$, где количество категорий, которое можно присвоить одному символу, ограничено константой k.

Для данных классов возникает естественный вопрос соотношения классов языков, порождаемых грамматиками, который полностью решается во втором разделе. Говоря точнее, доказывается, что для любых классов $\mathcal{G}^m$ и $\mathcal{G}^k$, где $m < k$, существует язык, порождаемый грамматикой из $\mathcal{G}^k$, но не порождаемый никакой грамматикой из $\mathcal{G}^m$. Тем самым, будет получен аналог результата из работы [5], где была доказана теорема об иерархии для категориальных грамматик с ограничением на количество присваиваемых категорий.

Заключительный же раздел посвящен исследованию класса $\mathcal{G}^2$. Хотя данный класс не может задавать даже очень простые языки, его выразительной силы достаточно для того, чтобы “закодировать” произвольную грамматику. Говоря точнее, каким бы ни был язык, порожденный произвольной грамматикой G, существует язык, порожденный грамматикой из класса $\mathcal{G}^2$, полученный из языка $L(G)$ при помощи гомоморфизма h, и более того существует обратный гомоморфизм g к h. Тем самым, большинство алгоритмических свойств языков, порождаемых грамматиками из класса $\mathcal{G}^2$, эквивалентны свойствам всего класса контекстно-свободных языков.

2 Основные определения

Напомним основные определения.

Определение 1. *Пусть $Pr = \{p, q, r, \dots\}$ произвольное множество, его элементы будем называть "примитивными категориями". Множество "базовых категорий" BC определенно индуктивно:*

- *если p — примитивная категория, то $p \in BC$;*
- *если A — базовая категория, а p — примитивная категория, то (A/p) и $(p\backslash A)$ — базовые категории.*

Определение 2. *Базовая категориальная грамматика — это упорядоченная тройка $G = \langle \Sigma, S, \triangleright \rangle$, где Σ — конечный алфавит, $S \in Pr$ — произвольная примитивная категория и $\triangleright$ — произвольное конечное бинарное соответствие $\triangleright \subset \Sigma \times BC$, сопоставляющее каждой букве алфавита один или несколько синтаксических типов. Слово $a_1 \dots a_n$ порождается БКГ G, если существуют такие базовые категории $B_1, \dots, B_n$, что для любого $i \leq n$ выполняется $a_i \triangleright B_i$, и последовательность $B_1, \dots, B_n$ может быть приведена к S цепочкой преобразований (редукций) вида:*

$$A/q, q \to A \qquad \text{и} \qquad q, q\backslash A \to A$$

Язык, задаваемый G, определяется как множество всех непустых слов $a_1 \dots a_n$ в алфавите Σ, порождаемых грамматикой G, и обозначается $L(G)$.

Пусть G — произвольная категориальная грамматика, вместо того, чтобы выписывать все элементы множества $\triangleright$, будем писать все категории для одного символа в строчку через символ $|$, аналогично тому, как данное делается для правил контекстно-свободной грамматики. Например, обозначение $a \ \triangleright S \mid S/p \mid p$ значит, что символу a присвоены категории S, S/p и p. Или формально, что пары (a, S), $(a, S/p)$ и (a, p) принадлежат множеству $\triangleright$.

Каждой категории сопоставляется её числитель, а также левый и правый знаменатель.

Определение 3. *Для категории A определим числитель $\xi(A) \in Pr$, а также левый и правый знаменатели $\varphi(A), \psi(A) \in Pr^*$ индуктивно:*

- *если $A = p \in Pr$, то $\xi(A) = p$, $\varphi(A) = \psi(A) = \varepsilon$;*
- *если $A = q\backslash B$, то $\xi(A) = \xi(B)$, $\varphi(A) = \varphi(B)\, q$ и $\psi(A) = \psi(B)$;*
- *если $A = B/q$, то $\xi(A) = \xi(B)$, $\varphi(A) = \varphi(B)$ и $\psi(A) = q\, \psi(B)$.*

Естественное преобразование произвольной БКГ G в форму контекстно-свободной грамматики $CF(G)$ может быть найдено в [6]. Не повторяя формальное индуктивное построение, приведём пример.

Пример 1. *Пусть G задана набором присвоений:*

$$\begin{aligned} a &\rhd q\backslash(S/p) \mid S/S \\ b &\rhd (q\backslash(S/p))/p \mid p \\ c &\rhd q \end{aligned}$$

Тогда эквивалентная грамматика $CF(G)$ имеет следующие правила:

$$\begin{aligned} S &\to qap \mid aS \mid qbpp \\ p &\to b \\ q &\to c \end{aligned}$$

Теорема 1 (Теорема 1 [6]). *Языки, порождаемые БКГ G и контекстно-свободной грамматикой $CF(G)$, совпадают.*

Отметим несколько допущений, позволяющих сильно упростить обозначения. Везде, где не указано обратное, будем обозначать результирующую категорию грамматики S, тем самым уберём необходимость явно указывать результирующую категорию. Более того, если результирующая категория S — то БКГ однозначно восстанавливается по набору присвоений категорий буквам данной грамматики, следовательно, отождествим БКГ и множество присвоений категорий в данной грамматике. Например, под грамматикой $a \rhd S \mid (S/S)$ будем понимать следующую грамматику:

$$G = \langle \{a\}, S, \{(a, S), (a, (S/S))\} \rangle$$

Также отметим, что язык, порождаемый БКГ, не меняется от порядка скобок в присвоенных категориях, что является прямым следствием теоремы выше. Говоря точнее, для произвольных грамматик G_1 и G_2, которые отличаются лишь порядком скобок в присваиваемых категориях, $CF(G_1)$ и $CF(G_2)$ — это контекстно-свободные грамматики с одинаковым набором правил. Таким образом, в записи категорий будут опускаться все скобки. Например, категорию $q\backslash(S/p)$ будем писать как $q\backslash S/p$.

Теперь дадим определение класса грамматик $\mathcal{G}^k$ с не более чем k присвоенными категориями для одного символа.

Определение 4. *Произвольная базовая категориальная грамматика G принадлежит классу $\mathcal{G}^k$ тогда и только тогда, когда для любого символа $a \in \Sigma$ выполнено $|\{A \mid a \rhd A\}| \leq k$. Обозначим через $\mathcal{L}^k$ — класс языков, порождаемых грамматиками из класса G^k.*

Отметим простое следствие определений.

Предложение 1. *Для произвольного k, если грамматика $G \in \mathcal{G}^k$, то $G \in \mathcal{G}^{k+1}$. Аналогично, если язык $L \in \mathcal{L}^k$, то $L \in \mathcal{L}^{k+1}$.*

Как уже отмечалось, любой контекстно-свободный язык без пустого слова порождается некоторой базовой категориальной грамматикой. Тем самым, любой контекстно-свободный язык без пустого слова принадлежит какому-то из классов $\mathcal{L}^k$.

Рассмотрим тривиальные примеры языков из различных классов $\mathcal{L}^k$.

Пример 2. *Язык $\{aa\}$ порождается грамматикой из класса $\mathcal{G}^2$, но не порождается грамматикой из класса $\mathcal{G}^1$.*

Доказательство. Чтобы было возможно хотя бы одно сокращение категорий в грамматике, букве a должна быть назначена хотя бы одна примитивная категория. Если a назначена категория S, то грамматика должна порождать слово из одного символа a, что не так. Иначе, a назначена другая категория p, но тогда слово aa не будет порождаться данной грамматикой. Тем самым, данный язык не в классе $\mathcal{L}^1$, но грамматика $a \triangleright p \mid S/p$ порождает данный язык. $\square$

Таким образом, мы доказали, что класс $\mathcal{L}^1 \subsetneq \mathcal{L}^2$. Подобным образом, попробуем "разделить" классы $\mathcal{L}^2$ и $\mathcal{L}^3$, что продемонстрируем следующим примером.

Пример 3. *Язык $L = \{a, aa\}$ не задаётся никакой грамматикой из $\mathcal{G}^2$, но задаётся грамматикой $a \triangleright S \mid p \mid p\backslash S$.*

Доказательство. Чтобы слово a принадлежало языку, в грамматике символу a должна быть присвоена категория S. Рассмотрим сокращение категорий для слова aa. Чтобы было возможно сокращение, есть две возможности: первой a присвоена произвольная примитивная категория p, а второй a присвоена категория $p\backslash S$ или второй a присвоена произвольная примитивная категория p, а первой a — категория S/p. Тем самым, единственные возможности, чтобы и слово a, и слово aa порождалось грамматикой из $\mathcal{G}^2$, необходимо, чтобы символу a были присвоены категории S/S и S, или $S\backslash S$ и S. Но обе эти грамматики порождают больший язык, в частности, слово aaa порождается обеими грамматиками. $\square$

Итого, мы показали, что $\mathcal{L}^1 \subsetneq \mathcal{L}^2 \subsetneq \mathcal{L}^3$.

3 Иерархия классов

Наша ближайшая цель состоит в том, чтобы обобщить примеры из конца предыдущей секции на случай грамматик с k категориями. Как мы уже видели, различия в классах $\mathcal{L}^1$, $\mathcal{L}^2$ и $\mathcal{L}^3$ появляются в достаточно простых языках. В действительности, конечных языков хватит, чтобы "различить" $\mathcal{L}^k$ и $\mathcal{L}^{k+1}$ при произвольном k. Для доказательства нам потребуются следующее техническое утверждение.

Предложение 2. *Пусть G — БКГ, задающая конечный язык. Тогда существует эквивалентная G БКГ G' и терминальный символ a, которому присвоена примитивная категория p такая, что никакая строка* *из двух и более категорий, встречающихся в присвоениях категорий грамматики G', не сокращается до p. Более того, если $G \in \mathcal{G}^k$, то $G' \in \mathcal{G}^k$.*

Доказательство. Сперва устраним в грамматике присвоения категорий, которые не могут участвовать ни в какой редукции, что делается стандартным образом, аналогично контекстно-свободным грамматикам.

Обозначим полученную грамматику G'. Рассмотрим слово $w = a_1 \dots a_n$ максимальной длины в языке $L(G')$, таким образом, существуют присвоения категорий символам a_i, что строка $A_1 \dots A_n$ редуцируется в категорию S.

Если $w = a_1$, то в грамматике G' символу a_1 присвоена категория S, и так как это слово максимальной длины, то никакие другие категории из присвоений грамматики G' не могут сокращаться до S, тем самым получаем требуемое утверждение.

Теперь рассмотрим случай, когда w состоит из двух или более символов. Чтобы была возможна хотя бы одна редукция, одна из категорий A_i должна быть примитивной категорией p. Отметим, что p не может быть результирующей категорией (т.е. не равна S), иначе строка

$$a_1 \dots a_{i-1} a_1 \dots a_{i-1} a_i a_{i+1} \dots a_n a_{i+1} \dots a_n$$

принадлежит языку $L(G')$, что демонстрируется следующей последовательностью редукций:

$$A_1 \dots A_{i-1} A_1 \dots A_{i-1} A_i A_{i+1} \dots A_n A_{i+1} \dots A_n \Rightarrow \\ A_1 \dots A_{i-1} S A_{i+1} \dots A_n \Rightarrow S$$

что противоречит максимальности длины слова w.

Теперь предположим, что какая-то из строк категорий, встречающихся в G', может сократиться до p. Обозначим её α, а строку терминалов, из которой присвоениями получается α, обозначим u. Рассмотрим строку символов $a_1 \dots a_{i-1} u a_{i+1} \dots a_n$ и соответствующее присвоение категорий её символам $A_1 \dots A_{i-1} \alpha A_{i+1} \dots A_n$. Следующие редукции показывают, что $A_1 \dots A_{i-1} \alpha A_{i+1} \dots A_n$ редуцируется в S:

$$A_1 \dots A_{i-1} \alpha A_{i+1} \dots A_n \Rightarrow A_1 \dots A_{i-1} p A_{i+1} \dots A_n \Rightarrow S$$

значит, слово $a_1 \dots a_{i-1} u a_{i+1} \dots a_n$ принадлежит языку грамматики G', что противоречит максимальности длины слова w. $\square$

Также нам потребуются следующие функции.

Определение 5. *Для произвольного языка L определим значение функции λ как такое минимальное число k, что L может быть задан грамматикой из $\mathcal{G}^k$.*

Определение 6. *Для произвольного конечного языка L определим значение функции μ как такое минимальное число k, что существует конечный язык L' с условием $L \subseteq L'$, и L' может быть задан грамматикой из $\mathcal{G}^k$.*

Предложение 3. *Функции λ и μ удовлетворяют следующим свойствам:*

- *пусть L — произвольный конечный язык, тогда $\mu(L) \leq \lambda(L)$;*
- *пусть L и $\tilde{L}$ — конечные языки такие, что $L \subseteq \tilde{L}$, тогда $\mu(L) \leq \mu(\tilde{L})$.*

Доказательство. Отметим, что $\mu(L)$ — это минимальное необходимое количество категорий, чтобы задать конечный язык, содержащий L, следовательно, $\mu(L)$ меньше или равно количеству категорий, необходимых для задания L.

Для доказательства второго утверждения достаточно заметить, что минимум $\mu(L)$ берётся по большему множеству языков, чем $\mu(\tilde{L})$. □

Теперь с помощью функции λ удобно сформулировать основную теорему данной секции.

Теорема 2. *Для произвольного натурального числа k существует такой язык L, что $\lambda(L) = k$.*

Доказательство. Рассмотрим значение функции μ на языках вида $L_n = \{a^i \mid i \leq n\}$ при различных n.

Лемма 1. *Значения функции μ неограничены на множестве языков $\{L_n \mid n \in \mathbb{N}\}$.*

Доказательство. Доказательство от противного. Пусть l — наименьшая верхняя грань, которая достигается на языке $\{a^i \mid i \leq t\}$.

Рассмотрим язык $\{a^i \mid i \leq (t+1)^{l+1}+1\}$ и категориальную грамматику G, задающую данный язык. Заметим, что $\mu(L_{(t+1)^{l+1}+1}) = l$ по предположению и второй части утверждения 3. По определению μ это значит, что существует конечный язык $\tilde{L}$, содержащий язык $L_{(t+1)^{l+1}+1}$, и $\lambda(\tilde{L}) = l$.

В языке $\tilde{L}$ содержится как минимум $(t+1)^{l+1}+1$ различных слов, так как $\tilde{L}$ содержит язык $L_{(t+1)^{l+1}+1}$. Более того, язык $\tilde{L}$ конечен и задаётся грамматикой из $\mathcal{G}^l$, обозначим эту грамматику G, а категории, присвоенные символу a в этой грамматике, $A_1, \ldots, A_l$.

Рассмотрим $CF(G)$. Глубина деревьев вывода в данной грамматике ограничена $l+1$, иначе существовала бы примитивная категория, повторяющаяся в одной ветви дерева вывода, а следовательно, повторением этой ветви можно было бы получить слово длины больше произвольной константы.

Предположим, что в каждом правиле грамматики $CF(G)$ с правой стороны не более чем t символов. Значит, на каждом шаге добавляется не более t новых вершин, а следовательно, в итоговом слове не более $(t+1)^l$ символов. Следовательно, существует правило $p \to u$, что в правой части минимум $t+1$ символ.

Правило $p \to u$ не может быть использовано в выводе слов длины меньше $t+1$. Следовательно, грамматика $CF(G)$ с удаленным правилом $p \to u$ обязана порождать все слова с длиной не более чем t. Без ограничения общности будем считать, что в изначальной грамматике G правило $p \to u$ соответствовало присвоению $a \triangleright A_1$. Рассмотрим G' — грамматику, полученную из грамматики G удалением присвоения $a \triangleright A_1$. По построению, $CF(G')$ отличается от $CF(G)$ только правилом $p \to u$, а следовательно, G' порождает все слова с длиной не более чем t, а значит, $\{a^i \mid i \leq t\} \subseteq L(G)$. Однако, в грамматике G' всего $l-1$ присвоение, следовательно, $\mu(\{a^i \mid i \leq t\}) \leq l-1$, что противоречит предположению. □

Лемма 2. *Для* $k \geq 2$ *верно* $\lambda(\{a^i \mid i \leq k+1\}) \leq \lambda(\{a^i \mid i \leq k\}) + 1$.

Доказательство. Пусть грамматика G порождает язык $\{a^i \mid i \leq k\}$. По утверждению 2 существует грамматика G', порождающая тот же язык, и примитивная категория p, присвоенная символу a в G', что никакая строка из двух и более категорий в данной грамматике не может сокращаться до p. Обозначим присвоения в G' следующим образом: $a \triangleright A_1 \mid \ldots \mid A_n$, где $n = \lambda(\{a^i \mid i \leq k\})$.

Теперь рассмотрим новую грамматику G'' со следующими присвоениями — $a \triangleright S/p \ldots /p \mid A_1 \mid \ldots \mid A_n$, где количество делений на категорию p в первой категории равно k. Так как эта грамматика содержит все категории из грамматики, задающей L_k, любое слово из L_k также принадлежит языку данной грамматики. Также заметим, что a^{k+1} принадлежит порождаемому языку, когда первой букве a присваивается категория $S/p \ldots /p$, а всем остальным — категория p.

Осталось доказать, что никаких других слов данная грамматика породить не может. Для этого достаточно рассмотреть $CF(G'')$ и $CF(G')$. Данные грамматики отличаются только одним правилом, а именно $S \to ap \ldots p$, которое, таким образом, обязано быть использовано при выводе нового слова a^t, где $t \geq k+1$. Заменим все поддеревья в выводе слова a^t, начинающиеся с указанного выше правила, на дерево вывода слова a^k. Итого, получим дерево вывода, в котором использовались только правила из $CF(G')$, и если правило $S \to ap \ldots p$ было использовано не в корне дерева вывода a^t, то получим слово длины больше k, в выводе которого использовались только правила из $CF(G')$, что невозможно, так как $L(CF(G')) = \{a^i \mid i \leq k\}$. Итого, правило $S \to ap \ldots p$ было использовано в корне дерева вывода a^t. Осталось отметить, что, так как никакая строка нескольких категорий в G'' не сокращается до p, значит, поддерево, соответствующее p, после применения указанного выше правила может состоять только из одного символа a, а это значит, что $t = k+1$. □

Итого, доказано, что λ — неограниченная функция, так как μ неограничена, и на каждом шаге растёт не более чем на единицу. На языке

$\{a, aa\}$ значение λ равно 2. Тем самым, λ принимает все натуральные значения, что завершает доказательство. □

4 Класс языков с двумя присвоениями

Для начала разберём ранее встречавшийся пример языка $\{a, aa\}$, и категориальную грамматику с правилами $a \triangleright S \mid p \mid S/p$.

Пример 4. *Базовая категориальная грамматика:*

$$\begin{aligned} a_1 &\triangleright S/t_3/t_2 \mid t_1 \\ a_2 &\triangleright t_1\backslash p/t_3 \mid t_2 \\ a_3 &\triangleright t_2\backslash t_1\backslash S/p \mid t_3 \end{aligned}$$

порождает язык $\{a_1a_2a_3, a_1a_2a_3a_1a_2a_3\}$.

Неформально, каждая подстрока $a_1a_2a_3$ "кодирует" в себе все три категории, присвоенных букве a в изначальной грамматике.

Зафиксируем произвольную БКГ G и обобщим кодировку грамматики из примера.

Определение 7. *Построим грамматику G' по грамматике G, заменив присвоения одному символу $a \triangleright A_1 \mid \ldots \mid A_n$ на следующий набор правил:*

$$\begin{aligned} a_1 &\triangleright A_1/t_n^a/t_{n-1}^a/\ldots/t_2^a \mid t_1^a \\ a_2 &\triangleright t_1^a\backslash A_2/t_n^a/t_{n-1}^a/\ldots/t_3^a \mid t_2^a \\ &\ldots \\ a_{n-1} &\triangleright t_{n-2}^a\backslash\ldots\backslash t_1^a\backslash A_{n-1}/t_n^a \mid t_{n-1}^a \\ a_n &\triangleright t_{n-1}^a\backslash\ldots\backslash t_1^a\backslash A_n \mid t_n^a \end{aligned}$$

(t_i^a — новые примитивные категории, не встречающиеся в категориях изначальной грамматики). Обозначим A_i' и A_i'' категории, присвоенные a_i, соответственно. Для символа a определим образ гомоморфизма $h(a)$ как строку $a_1a_2\ldots a_n$.

Обратно, для символа a_1 определим образ гомоморфизма $g(a_1) = a$, для всех остальных символов a_i (где $i \neq 1$) положим $g(a_i) = \varepsilon$.

Перейдём к эквивалентным контекстно-свободным грамматикам для G и G'. Таким образом, грамматика $CF(G)$ задана следующим набором правил (выпишем правила с терминалом a с правой части):

$$\begin{aligned} &\xi(A_1) \to \varphi(A_1)a\psi(A_1) \\ &\xi(A_2) \to \varphi(A_2)a\psi(A_2) \\ &\ldots \\ &\xi(A_n) \to \varphi(A_n)a\psi(A_n) \end{aligned}$$

По определению $\xi(A_i) = \xi(A'_i)$, а строки левых и правых знаменателей следующие:

$$\begin{array}{ll}
\varphi(A'_1) = \varphi(A_1) & \psi(A'_1) = t^a_2 t^a_3 \dots t^a_n \psi(A_1) \\
\varphi(A'_2) = \varphi(A_2) t^a_1 & \psi(A'_1) = t^a_3 \dots t^a_n \psi(A_2) \\
\dots & \\
\varphi(A'_n) = \varphi(A_n) t^a_1 t^a_2 t^a_3 \dots t^a_{n-1} & \psi(A'_n) = \psi(A_n)
\end{array}$$

Для A''_i, положим $\xi(A''_i) = t^a_i$, а строки левых и правых знаменателей, соответственно, пустыми.

Итого, грамматика G' эквивалентна контекстно-свободной грамматике с правилами:

$$\begin{array}{l}
\xi(A_1) \to \varphi(A_1) a_1 t^a_2 t^a_3 \dots t^a_{n-1} t^a_n \psi(A_1) \\
\xi(A_2) \to \varphi(A_2) t^a_1 a_2 t^a_3 \dots t^a_{n-1} t^a_n \psi(A_2) \\
\dots \\
\xi(A_n) \to \varphi(A_n) t^a_1 t^a_2 t^a_3 \dots t^a_{n-1} a_n \psi(A_n) \\
t^a_1 \to a_1 \\
t^a_2 \to a_2 \\
\dots \\
t^a_n \to a_n
\end{array}$$

Теперь заметим, что каждый нетерминал t^a_i встречается в левой части грамматики ровно в одном правиле, а именно в i-м из последних n. Тем самым, можно везде в грамматике заменить t^a_i на a_i и получить эквивалентную грамматику.

Таким образом, получили грамматику (выписываем правила только полученные из a):

$$\begin{array}{l}
\xi(A_1) \to \varphi(A_1) a_1 a_2 a_3 \dots a_{n-1} a_n \psi(A_1) \\
\xi(A_2) \to \varphi(A_2) a_1 a_2 a_3 \dots a_{n-1} a_n \psi(A_2) \\
\dots \\
\xi(A_n) \to \varphi(A_n) a_1 a_2 a_3 \dots a_{n-1} a_n \psi(A_n)
\end{array}$$

которая отличается от правил грамматики G заменой терминала a на строку терминалов $a_1 \dots a_n$.

Тем самым были доказаны следующие теоремы.

Теорема 3. *Если слово w лежит в языке $L(G)$, то $h(w)$ лежит в языке $L(G')$.*

Аналогично для обратного гомоморфизма.

Теорема 4. *Если слово w лежит в языке $L(G')$, то $g(w)$ лежит в языке $L(G)$.*

Следствие 1. *Следующие алгоритмические задачи неразрешимы для языков из класса $\mathcal{G}^2$:*

- *по произвольной грамматике G из класса $\mathcal{G}^2$ проверить, является ли язык $L(G)$ существенно неоднозначным;*
- *по произвольной грамматике G из класса $\mathcal{G}^2$ проверить, является ли язык, порождаемый грамматикой G, регулярным;*
- *проверить по двум грамматикам из класса $\mathcal{G}^2$, равны ли порождаемые ими языки.*

Доказательство. В первых двух случаях свойства языков сохраняются под действием гомоморфизмов, таким образом, утверждаемое является следствием классических теорем.

Для доказательства третьего утверждения заметим, что в произвольных двух категориальных грамматиках G_1 и G_2 можно считать, что выбранному символу a в первой и во второй грамматике присвоенно одинаковое количество категорий. Иначе некоторые категории можно продублировать. Построим соответственно по G_1 и G_2 грамматики G'_1 и G'_2. Тогда для произвольного символа a верно, что $h_1(a) = h_2(a)$, где h_1 и h_2, соответствующие гомоморфизмы. Значит, $L(G'_1) = L(G'_2)$ тогда и только тогда, когда $L(G_1) = L(G_2)$. □

References

[1] K. Ajdukiewicz, *Die syntaktische Konnexität*, Stud. Philos., **1** (1935), 1–27. Zbl 0015.33702

[2] Y. Bar-Hillel, *A quasi-arithmetical notation for syntactic description*, Language, **29**:1 (1953), 47–58. Zbl 0156.25402

[3] Y. Bar-Hillel, H. Gaifman, E. Shamir, *On categorial and phrase structure grammars*, Bull. Res. Council. Israel, Sect. F, 9F (1960), 155–166. Zbl 0097.14202

[4] W. Buszkowski, *Solvable problems for classical categorial grammars* Bull. Pol. Acad. Sci., Math., **35**:1-10 (1987), 373–382. Zbl 0642.03022

[5] M. Kanazawa, *Identification in the limit of categorial grammars* J. Logic Lang. Inf., **5**:2 (1996), 115–155. Zbl 0855.03016

[6] M.E. Vishnikin, *Unique typed basic categorial grammars*, Mosc. Univ. Math. Bull., **77**:2 (2022), 93—96. Zbl 7584507

Maxim Evgenievich Vishnikin
Moscow State University,
GSP-1, Leninskie Gory,
119991, Moscow, Russia
Email address: maksim.vishnikin@math.msu.ru