ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ НАУКИ
МАТЕМАТИЧЕСКИЙ ИНСТИТУТ им. В. А. СТЕКЛОВА
РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи
УДК 519.766.23



**Вишникин Максим Евгеньевич**

# Категориальные грамматики с ограничением на количество присваиваемых типов

Специальность 1.1.5 — математическая логика, алгебра,
теория чисел и дискретная математика

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва 2025

Работа выполнена в отделе математической логики Федерального государственного бюджетного учреждения науки Математического института им. В. А. Стеклова Российской академии наук.

| | |
|---|---|
| Научный руководитель: | **Кузнецов Степан Львович**<br>доктор физико-математических наук, ведущий научный сотрудник отдела математической логики ФГБУН Математического института им. В. А. Стеклова Российской академии наук (специальность — 1.1.5). |
| Официальные оппоненты: | **Селиванов Виктор Львович**<br>доктор физико-математических наук, профессор, профессор факультета математики и компьютерных наук ФГБОУ ВО Санкт-Петербургский государственный университет (специальность — 01.01.06). |
| | **Карлов Борис Николаевич**<br>кандидат физико-математических наук, доцент, доцент кафедры информатики факультета прикладной математики и кибернетики ФГБОУ ВО «Тверской государственный университет» (специальность — 01.01.06). |
| Ведущая организация: | ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)». |

Защита состоится 18 сентября 2025 г. в 15 часов 30 мин. на заседании диссертационного совета 24.1.167.03 при Математическом институте им. В.А. Стеклова Российской академии наук по адресу: 119333 г. Москва, ул. Губкина, д. 8.

С диссертацией можно ознакомиться в библиотеке Математического института им. В. А. Стеклова и на сайте МИАН: https://www.mi-ras.ru/dis/ref25/vishnikin/dis.pdf.

Автореферат разослан “ ” июля 2025 г.

Учёный секретарь
диссертационного совета
24.1.167.03 при МИАН,
д.ф.-м.н. профессор РАН Д. В. Осипов

## Общая характеристика работы

### Актуальность темы и степень её разработанности

Настоящая диссертация посвящена теории формальных языков, относящейся к пересечению математической логики и теории алгоритмов, теоретической информатики и алгебры. Сегодня эта теория представляет собой обширный раздел математики, в котором формируется целый спектр формализмов, восходящих к лингвистическим моделям.

Важнейшим источником для теории формальных языков является теория порождающих грамматик Н. Хомского[1]. В этой теории был предложен подход к описанию синтаксических структур естественного языка в виде формального исчисления, порождающего цепочки в некотором алфавите. Символами (буквами) алфавита являются слова естественного языка; используются также дополнительные символы, среди которых могут быть синтаксические типы, например: «группа подлежащего». С помощью такого формализма можно также естественно описывать синтаксис языков математической логики и языков программирования[2]. В диссертации мы используем термин «слово» как синонимичный термину «цепочка символов» в данном (конечном) алфавите. Хомским была предложена иерархия порождающих грамматик и, соответственно, порождаемых ими языков. Важнейшим является класс контекстно-свободных грамматик и порождаемых ими языков.

Другой подход, который можно охарактеризовать как «логический», заключается в том, чтобы приписать элементам алфавита $\Sigma$ некоторые «типы» – формулы специального построенного логического языка. Предложение же считают «корректно заданным», если существует такой выбор типов для каждого отдельного слова, что в некотором исчислении выводится импликация из данной последовательности типов в фиксированный результирующий тип. Такое описание синтаксиса хорошо согласуется с определением семантики по Монтегю[3], где «смысл» выражения определяется, исходя из семантики его частей (см. также монографию И.А. Мельчука[4]). Общее название для класса подобного рода формализмов – «категориальные грамматики».

Исторически первым вариантом категориальных грамматик были «AB-грамматики»

[1] Chomsky N. Three models for the description of language // IRE Transactions on Information Theory. – 1956. – Vol. 2, N. 3. – P. 113–124.

[2] Aho A., Sethi R., Ullman J. Compilers: Principles, Techniques and Tools. – Reading, Mass.: Addison-Wesley, 1985.

[3] Carpenter B. Type-Logical Semantics. – MIT Press, 1998.

[4] Мельчук И.А. Опыт теории лингвистических моделей «Смысл ⇔ Текст» – М.: Изд-во «Наука», Главная редакция восточной литературы, 1974. – 314 с.

(в честь К. Айдукевича и Й. Бар-Хиллела), восходящие к работе Айдукевича[5] 1935 года. В своей современной форме данное понятие появляется в работе Бар-Хиллела[6] 1953 года. В этих грамматиках типы построены из произвольного множества «примитивных типов» посредством двух операций – левого и правого деления. Позднее в работе Бар-Хиллела и др.[7] доказаны основные свойства формализма, наиболее интересным из которых является равенство класса порождаемых языков классу контекстно-свободных языков без пустого слова. Данный формализм играет ключевую роль, так как большинство других категориальных грамматик являются его расширением при помощи добавления новых операций и правил.

Приведём пример, как категориальные грамматики могут задавать множества корректных предложений. Рассмотрим фразу: «Иван встречает Марию». Синтаксическим единицам «Иван» и «Мария» присваиваются типы именной группы $np$. Глаголу «встречает» присваивается тип $(np\backslash(S/np))$, где $S$ — тип правильно построенного предложения. Таким образом, для слов в данной фразе существует такое присвоение типов $np; (np\backslash(S/np)); np$, что выводится соответствующая импликация: $np; (np\backslash(S/np)); np \Rightarrow S$.

Дополнительные возможности в описании синтаксиса естественного языка дают категориальные грамматики, предложенные И. Ламбеком[8,9]. Этот вид грамматик построен на основе AB-грамматик за счёт добавления правил введения делений справа и дополнительной операции — умножения. М.Р. Пентусом[10] доказано равенство класса языков, порождаемых категориальными грамматиками Ламбека, классу контекстно-свободных языков без пустого слова. Среди расширений AB-грамматик рассматривались также категориальные грамматики зависимостей (CDG), комбина́торные категориальные грамматики (CCG) и конъюнктивные

[5]Ajdukiewicz K. Die syntaktische Konnexität // Studia Philosophica. – 1935. – Bd. 1 – S. 1–27.

[6]Bar-Hillel Y. A quasi-arithmetical notation for syntactic description // Language. – 1953. – Vol. 29, N. 1. – P. 47–58.

[7]Bar-Hillel Y., Gaifman H., Shamir E. On categorial and phrase structure grammars // Bulletin of the Research Council of Israel. – 1960. – Vol. 9F – P. 155–166.

[8]Lambek J. The mathematics of sentence structure // The American Mathematical Monthly. – 1958. – Vol. 65, N. 3. – P. 154–170.

[9]Lambek J. On the calculus of syntactic types // In: R. Jakobson (Ed.), Structure of Language and Its Mathematical Aspects. / Proceedings of Symposia in Applied Mathematics – Providence, R.I. : AMS, 1961. – Vol. 12 – P. 166–178.

[10]Pentus M. Lambek grammars are context free // In: Proceedings Eighth Annual IEEE Symposium on Logic in Computer Science. – IEEE, 1993. – P. 429–433.

категориальные грамматики. Введенные М.И. Дехтярём совместно с А.Я. Диковским[11,12], а впоследствии исследованные в работах Дехтяря, Диковского и Карлова[13,14], CDG получены из AB-грамматик при помощи добавления валентности у типов. Комбинáторные категориальные грамматики, изучавшиеся М. Стидманом[15], расширяют AB-грамматики добавлением новых правил преобразования строк типов. Конъюнктивные категориальные грамматики, введённые в статье С.Л. Кузнецова и А.С. Охотина[16], расширяют AB-грамматики за счёт добавления конъюнкций в знаменателях типов.

В теории формальных языков важную роль играют алгоритмические и алгебраические свойства различных классов грамматик. Классические алгебраические свойства — совпадения классов языков для различных формализмов, замкнутость относительно естественных операций. К основным алгоритмическим задачам относятся: проверка включения языков, проверка эквивалентности, задача синтаксического разбора и многие другие.

Известно множество работ, посвящённых различным алгоритмическим задачам для контекстно-свободных грамматик. Для задачи синтаксического анализа, то есть проверки принадлежности слова некоторому контекстно-свободному языку, И. Сакаи[17] доказана разрешимость за полиномиальное время (обычно алгоритм называют алгоритмом Кока — Янгера — Касами, по более поздним работам). Для класса контекстно-свободных грамматик, для которых у каждого слова существует одно «левое» древо разбора, А.Л. Семёновым[18] доказана теорема о разрешимости равенства автоматному языку. Однако многие задачи алгоритмиче-

[11]Dikovsky A. Dependencies as categories // In: Proceedings of the Workshop on Recent Advances in Dependency Grammar. – 2004. – P. 82–89.

[12]Dekhtyar M., Dikovsky A. Categorial dependency grammars // In: M. Moortgat, V. Prince (Eds.), Proceedings of the International Conference on Categorial Grammars, Montpellier, 2004. – P. 76–91.

[13]Dekhtyar M., Dikovsky A. Generalized categorial dependency grammars // Pillars of Computer Science: Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday. – Berlin, Heidelberg : Springer Berlin Heidelberg, 2008. – P. 230–255.

[14]Dekhtyar M., Dikovsky A., Karlov B. Categorial dependency grammars // Theoretical Computer Science. – 2015. – Vol. 579. – P. 33–63.

[15]Steedman M. The Syntactic Process. – MIT Press, 2001. – 348 pp.

[16]Kuznetsov S., Okhotin A. Conjunctive categorial grammars // Proceedings of the 15th Meeting on the Mathematics of Language. – ACL, 2017. – P. 140–151.

[17]Sakai I. Syntax in universal translation // In: Proceedings of the International Conference on Machine Translation and Applied Language Analysis. Her Majesty's Stationery Office, London, 1962. – P. 593–608.

[18]Семёнов А.Л. Алгоритмические проблемы для степенных рядов и контекстно-свободных грамматик // Доклады Академии наук СССР. – 1973. – Т. 212, № 1. – С. 50–52.

ски неразрешимы, что доказывается в работах Ш. Грейбах[19], А.В. Гладкого[20], Бар-Хиллела[21] и других авторов. В настоящее время эти результаты стали классическими[22,23].

Наряду с языками интерес представляют их гомоморфизмы. Одной из наиболее известных теорем о представлении посредством гомоморфизмов является теорема Хомского – Шюценберже[24], которая утверждает, что произвольный контекстно-свободный язык может быть получен как гомоморфный образ пересечения регулярного языка и языка правильных скобочных последовательностей, а также её различные усиленные варианты, например, полученный Охотиным[25] вариант для нестирающих гомоморфизмов.

Наряду с гомоморфными образами рассматривают и представления языков взятием прообраза при гомоморфизме. Классическим результатом для такого рода представлений является теорема Грейбах «о самом трудном языке»[26]. В работе Грейбах строится такой контекстно-свободный язык $L_0$, что для любого контекстно-свободного языка $L$ существует гомоморфизм $h$, для которого выполнено: $L = h^{-1}(L_0)$. Подобного рода представление имеет место не для всех грамматических формализмов. Грейбах доказала[27], что для класса языков, описываемых $LR(1)$-грамматиками, не существует подобного «самого трудного языка». В дальнейшем Л. Боассон и М. Нива доказали[28] аналогичный отрицательный результат для класса линейных контекстно-свободных языков.

В большинстве вариантов категориальных грамматик одному элементу алфавита $\Sigma$ может присваиваться несколько различных типов. С лингвистической точки зрения это свой-

[19]Greibach S. A note on undecidable properties of formal languages // Mathematical Systems Theory. – 1968. – Vol. 2, N. 1. – P. 1–6.

[20]Гладкий А.В. Алгоритмическая нераспознаваемость существенной неопределенности контекстно-свободных языков // Алгебра и логика. – 1965. – Т. 4, № 4. – С. 53–63.

[21]Bar-Hillel Y., Perles M., Shamir E. On formal properties of simple phrase structure grammars // In: Language and Information: Selected Essays on their Theory and Application, 1964. – P. 116–150.

[22]Гладкий А.В. Формальные грамматики и языки. – М.: Изд-во «Наука» Главная редакция физико-математической литературы, 1973. – 368 с.

[23]Kozen D.C. Automata and Computability. – Springer Science & Business Media, 1997.

[24]Chomsky N., Schützenberger M. The algebraic theory of context-free languages // Computer Programming and Formal Systems. – 1963. – Vol. 35. – P. 118–161.

[25]Okhotin A. Non-erasing variants of the Chomsky–Schützenberger theorem // In: International Conference on Developments in Language Theory. – Berlin, Heidelberg : Springer Berlin Heidelberg, 2012. – P. 121–129.

[26]Greibach S. The hardest context-free language // SIAM Journal on Computing. – 1973. – Vol. 2, N. 4. – P. 304–310.

[27]Greibach S. Jump PDA's and hierarchies of deterministic context-free languages // SIAM Journal on Computing. – 1974. – Vol. 3, N. 2. – P. 111–127.

[28]Boasson L., Nivat M. Le cylindre des langages linéaires // Mathematical Systems Theory. – 1977. – Vol. 11. – P. 147–155.

ство соответствует явлению омонимии: например, слово «течь» может быть как глаголом, так и существительным в различных контекстах.

Одной из предпосылок данной диссертации является работа А.Н. Сафиуллина[29], где доказывается, что для грамматик Ламбека ограничение на один присваиваемый тип одному символу не сужает класс порождаемых языков. Таким образом, упомянутый выше результат Пентуса переносится на случай однозначного присвоения типов. Основная идея в работе Сафиуллина заключается в том, чтобы имитировать множественность назначений типов одному символу при помощи одного типа. Позднее данный результат и общая конструкция были использованы в работах Кузнецова[30,31], в которых доказывается результат о сложности для исчисления Ламбека, пополненного дополнительной унарной операцией — звёздочкой Клини.

Представляется естественным вопрос — как условие однозначности присвоения типов повлияет на класс AB-грамматик. В отличие от грамматик Ламбека, подобное ограничение в данном случае оказывается очень существенным. Даже язык $a^+$ всех непустых слов из одной буквы такие грамматики задать не способны.

Отметим, что AB-грамматики с ограничением на количество присваиваемых типов изучались ранее. Так, в работах В. Бушковского[32] и М. Канадзавы[33] рассмотрены “rigid” и “$k$-valued” грамматики, в которых одному символу присвоено не более чем один или не более чем $k$ типов соответственно. В работе Канадзавы используются идеи, заложенные в работах Бушковского и Пенна[34,35], и доказывается, что данные классы грамматик обучаемы в смысле Голда.

Данное свойство, формализующее понятие обучения языку по его конечным фрагментам, было введено в работе Э.М. Голда[36]. Класс языков обучаем по Голду, если для лю-

[29]Сафиуллин А. Н. Выводимость допустимых правил с простыми посылками в исчислении Ламбека // Вестник Московского университета. Серия 1: Математика. Механика. – 2007. – № 4. – С. 72–76.

[30]Kuznetsov S. The Lambek calculus with iteration: two variants // International Workshop on Logic, Language, Information, and Computation. – Berlin, Heidelberg : Springer Berlin Heidelberg, 2017. – P. 182–198.

[31]Kuznetsov S. Complexity of the infinitary Lambek calculus with Kleene star // The Review of Symbolic Logic. – 2021. – Vol. 14, N. 4. – P. 946–972.

[32]Buszkowski W. Solvable problems for classical categorial grammars // Bulletin of the Polish Academy of Sciences: Mathematics. – 1987. – Vol. 35. – P. 373–382.

[33]Kanazawa M. Indentification in the Limit of Categorial Grammars // Journal of Logic, Language and Information. – 1996. – Vol. 5, N. 2. – P. 115–155.

[34]Buszkowski W. Discovery procedures for categorial grammars // In: J. van Benthem and E. Klein (Eds.), Categories, Polymorphism and Unification – Universiteit van Amsterdam, 1987. – С. 35–64.

[35]Buszkowski W., Penn G. Categorial grammars determined from linguistic data by unification // Studia Logica. – 1990. – Vol. 49. – P. 431–454.

бой бесконечной последовательности слов $w_1, w_2, \ldots$, начиная с некоторого момента $t_0$, существует не более одного языка из данного класса, порождающего все слова $w_1, \ldots, w_{t_0}$. Таким образом, любой язык из данного класса можно «угадать», зная только его достаточно большой конечный фрагмент. Класс языков, порождаемых AB-грамматиками (т.е. класс контекстно-свободных языков без пустого слова), не обучаем по Голду. С другой стороны, в выше упомянутой работе Канадзавы доказана обучаемость класса языков, задаваемых AB-грамматиками с ограничением $k$ на число присваиваемых типов для любого фиксированного $k$. Более того, процесс обучения можно сделать эффективным: существует алгоритм, принимающий на вход последовательность слов $w_1, w_2, \ldots$ и пытающийся «угадать» грамматику. Начиная с некоторого момента $t_0$, ответы, выдаваемые данным алгоритмом, стабилизируются на грамматике, правильно задающей данный язык.

Одним из часто встречающихся вариантов сужения AB-грамматик является рассмотрение подклассов, полученных ограничением на типы, которые могут быть использованы при присвоении. Упомянутое выше соответствие между AB-грамматиками и контекстно-свободными выражается не только в классе порождаемых языков, но и в эффективном преобразовании данных грамматик друг в друга. А именно, в преобразовании произвольной контекстно-свободной грамматики без пустого слова в форму AB-грамматики, за счёт формы Грейбах[37] возможно добиться отсутствия деления на сложные типы. Более того, любой контекстно-свободный язык без пустого слова может быть порождён AB-грамматикой с типами трёх видов ($p, q, r$ — произвольные примитивные типы): $p$, $(p/q)$, $((p/q)/r)$. Таким образом, естественно рассматривать AB-грамматики с делениями только на примитивные типы, такие грамматики в данной работе носят название «базовые категориальные грамматики». Подобного рода ограничения встречаются во многих формализмах, основанных на AB-грамматиках. Категориальные грамматики зависимостей и конъюнктивные категориальные грамматики обладают именно таким ограничением. Отметим, что базовыми категориальными грамматиками с однозначным присвоением типов посредством естественного гомоморфизма можно представить любой контекстно-свободный язык без пустого слова.

Хотя в работе Канадзавы исследуются AB-грамматики с однозначным присвоением типов, основной акцент делается на понятии обучаемости грамматике по Голду. При этом оставались открытыми вопросы, насколько данный подкласс AB-грамматик ограничен, а также

[36]Gold E.M. Language identification in the limit // Information and Control. – 1967. – Vol. 10, N. 5. – P. 447–474.

[37]Greibach S. A new normal-form theorem for context-free phrase structure grammars // Journal of the ACM. – 1965. – Vol. 12, N. 1. – P. 42–52.

насколько сложны связанные с ним алгоритмические задачи. Вопрос об алгоритмической разрешимости задачи проверки пустоты пересечения языков, заданных данными грамматиками, решен А. Форе[38]. Другие алгоритмические свойства данного класса, однако, оставались неизученными. Интуитивно, условие однозначности убирает весь возможный «недетерминизм» в грамматике, единственное, что остаётся, — это возможность различным образом редуцировать одну выбранную строку типов. Следовательно, распознавание и другие задачи могли бы решаться проще, чем для контекстно-свободных языков.

В то время как ограничение на количество присвоенных типов одному символу для AB-грамматик было исследовано в упомянутой выше работе Канадзавы (в частности, получена теорема об иерархии по числу присвоений), для базовых категориальных грамматик данный результат ранее не был известен. Для данных грамматик мы ставим вопросы, аналогичные полученным для более широких классов.

### Цели и задачи исследования

В работе изучаются классы категориальных грамматик, полученные из класса AB-грамматик при помощи следующих ограничений: ограничение множества возможных присваиваемых типов и ограничение на количество типов, присваиваемых одному символу в грамматике. Основной задачей работы является изучение свойств, в частности алгоритмических, для указанных классов грамматик.

### Научная новизна и положения, выносимые на защиту

Все основные результаты диссертации являются новыми, получены автором самостоятельно и состоят в следующем.

- Доказательство строгости иерархии классов языков, порождаемых базовыми категориальными грамматиками с присвоением не более $k$ типов одному символу ($\mathcal{L}^k$).
- Доказательство незамкнутости класса языков, порождаемых базовыми категориальными грамматиками с однозначным присвоением типов ($\mathcal{G}^1$), относительно операций объединения, пересечения, конкатенации, итерации Клини.
- Построение сведения посредством гомоморфизмов произвольного контекстно-свободного языка к языку, порождённому базовой категориальной грамматикой

[38]Foret A. The emptiness of intersection problem for languages of k-valued categorial grammars (classical and Lambek) is undecidable // Electronic Notes in Theoretical Computer Science. – 2004. – Vol. 53. – P. 81–93.

с двузначным присвоением типов (из класса $\mathcal{G}^2$). Как следствие, доказана алгоритмическая неразрешимость некоторых классических задач для категориальных грамматик из данного класса.

- Доказательство неразрешимости задачи проверки пустоты пересечения двух языков, порождаемых базовыми категориальными грамматиками с однозначным присвоением типов, над фиксированным алфавитом из 11 букв.

- Доказательство алгоритмической неразрешимости задачи проверки включения языков, заданных AB-грамматиками с однозначным присвоением типов (поставлена в статье Форе 2004 г.).

**Теоретическая и практическая значимость**

Работа имеет теоретический характер. Разработанные методы могут быть использованы в дальнейших исследованиях. В частности, предполагается, что кодирование множественности выбора типов применимо при рассмотрении других формализмов. Некоторые результаты могут иметь практическое применение при автоматизированном анализе конструкций языков программирования.

**Методы исследования**

В работе используются методы математической логики, теории формальных грамматик и теории алгоритмов.

**Степень достоверности и апробация результатов**

Достоверность результатов обеспечена строгими математическими доказательствами. Они опубликованы в рецензируемых научных журналах из списка ВАК, которые входят в базы данных Scopus и Web of Science. Список работ приведён в конце диссертации.

Результаты диссертации докладывались на следующих научных семинарах и конференциях.

- Международная конференция «Мальцевские чтения» (Новосибирск, ИМ им. С. Л. Соболева СО РАН и НГУ, с 11.11.2024 по 15.11.2024).

- Семинар «Теория доказательств» (Logic Online Seminar) отдела математической логики МИАН (Москва, 01.03.2021, 28.10.2024).

- Семинар «Категориальные грамматики» (Москва, МГУ, 12.05.2024).
- Объединённый семинар «Логические проблемы информатики» и «Модальная и алгебраическая логика» (Москва, МГУ, 05.12.2019).
- Семинар «Теоретические основы информатики» (Тверь, ТвГУ, 24.12.2024).
- Однодневный семинар по математической логике (Москва, НИУ ВШЭ, 29.06.2023).
- Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2021» (Москва, МГУ, с 12.04.2021 по 23.04.2021).
- Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2025» (Москва, МГУ, с 11.04.2025 по 25.04.2025).

**Структура и объём диссертационной работы**

Диссертация состоит из введения, трёх глав, заключения и библиографии. Объём работы составляет 92 страницы. Библиография включает 49 наименований.

## Содержание работы

Во введении даётся общая характеристика работы, формулируются постановки задач, цели работы и обосновывается её актуальность. Кроме того, приводится обзор результатов, полученных ранее другими авторами по теме диссертации и смежным направлениям исследований, а также кратко излагается содержание глав диссертации.

Первая глава вводная. Основной результат данной главы служит мотивировкой для рассмотрения условия ограниченного присвоения типов в базовых категориальных грамматиках. Сперва вводятся AB-грамматики и базовые категориальные грамматики. Затем доказывается равенство класса языков, заданных вторым формализмом, и классического класса контекстно-свободных языков без пустого слова. В конце рассматриваются классы грамматик и, соответственно, заданных ими языков, полученных из данного формализма наложением ограничения на количество типов, присваиваемых одному символу. Перечислим основные положения данной главы.

В параграфе 1.1 вводится понятие контекстно-свободной грамматики и AB-грамматики.

**Определение 1.1.3** (Бар-Хиллел и др., 1960)**.** Пусть дано конечное множество $\mathrm{Pr} = \{p, q, r, \dots\}$ *примитивных типов*. Определим множество *типов* Tp индуктивно:

- если $p$ — примитивный тип, то $p$ — тип;
- если $A$ и $B$ — типы, то $(B\backslash A)$ и $(A/B)$ — типы.

При записи строк типов типы разделяются точкой с запятой (;), а внешние скобки опускаются.

*AB-грамматика* есть упорядоченная четвёрка $(\Sigma, \mathrm{Pr}, S, \triangleright)$, где $\Sigma$ — конечный алфавит, Pr — конечное множество примитивных типов, $S \in \mathrm{Tp}$ ($S$ называется *результирующим типом*) и $\triangleright$ — произвольное конечное бинарное отношение $\triangleright \subseteq \Sigma \times \mathrm{Tp}$, *присваивающее* каждой букве алфавита некоторый набор типов.

Строки типов могут быть *преобразованы* при помощи следующих *правил редукции*, каждое из которых сокращает два рядом стоящих типа в один.

$$\frac{A/B \qquad B}{A} \qquad\qquad \frac{B \qquad B\backslash A}{A}$$

Соответственно, будем говорить, что строка типов $\alpha$ *редуцируется за один шаг* к строке типов $\beta$ ($\alpha \Rightarrow \beta$), если для некоторых типов $A$ и $B$, а также для некоторых строк типов $\eta$ и $\theta$ выполнено, что: $\alpha = \eta; (A/B); B; \theta$ или $\alpha = \eta; B; (B\backslash A); \theta$, а строка $\beta$ равна $\eta; A; \theta$. Будем говорить, что строка типов $\alpha$ *редуцируется* к строке типов $\beta$ ($\alpha \Rightarrow^* \beta$), если строку $\beta$ можно получить из $\alpha$ при помощи нуля или нескольких применений указанных выше правил.

Слово $a_1 \dots a_n$ *порождается* AB-грамматикой $G = (\Sigma, \mathrm{Pr}, S, \triangleright)$, если существуют такие типы $B_1, \dots, B_n$, что для каждого $i$ выполняется $a_i \triangleright B_i$, и строка $B_1; \dots; B_n$ может быть редуцирована к строке из единственного типа $S$.

Язык $L(G)$, *задаваемый* AB-грамматикой $G$, определяется как множество всех слов $a_1 \dots a_n$ в алфавите $\Sigma$, порождаемых грамматикой $G$.

Известно, что AB-грамматики задают все контекстно-свободные языки без пустого слова и только их. Затем даётся определение «$k$-значной» AB-грамматики, которое ранее было введено в работе Канадзавы, а также теорема, утверждающая, что «$k$-значные» грамматики при разных значениях параметра $k$ задают строго вложенные классы языков.

**Определение 1.1.5** (Канадзава, 1996)**.** AB-грамматика $(\Sigma, \mathrm{Pr}, S, \triangleright)$ называется *«$k$-значной»* AB-грамматикой, если для любого символа $a \in \Sigma$ выполнено, что $|\{A \mid a \triangleright A\}| \leq k$.

**Теорема 1.1.6** (Канадзава, 1996)**.** Для произвольного натурального числа $k$ существует такой язык $L$, что $L$ задаётся «$(k+1)$-значной» AB-грамматикой, но не задаётся никакой «$k$-значной» AB-грамматикой.

В параграфе 1.2 вводится понятие базовых категориальных грамматик, подкласса AB-грамматик, в котором запрещены типы вида $(A/B)$ и $(B\backslash A)$, в которых $B$ — не примитивный тип.

В параграфе 1.3 определяется способ преобразования произвольной базовой категориальной грамматики в контекстно-свободную грамматику особенного вида.

В параграфе 1.4 вводится класс базовых категориальных грамматик с присвоением не более $k$ типов.

**Определение 1.4.1.** Базовая категориальная грамматика $(\Sigma, \mathrm{Pr}, S, \triangleright)$ принадлежит классу $\mathcal{G}^k$, если для любого символа $a \in \Sigma$ выполнено, что $|\{A \mid a \triangleright A\}| \leq k$. Обозначим через $\mathcal{L}^k$ класс языков, задаваемых грамматиками из класса $\mathcal{G}^k$.

Дальше изучается соотношение классов $\mathcal{L}^k$ и $\mathcal{L}^{k+1}$ при произвольном $k$. Очевидно, что данные классы вложены, но появляется вопрос, а могло ли иметь место совпадение классов языков при некотором $k$. При конкретных небольших значениях $k$ это оказывается достаточно легко проверить на примерах языков, лежащих в классе $\mathcal{L}^2$, но не лежащих в классе $\mathcal{L}^1$, а также лежащих в $\mathcal{L}^3$, но не в $\mathcal{L}^2$. В итоге параграф завершается доказательством теоремы об иерархии, аналог которой для случая AB-грамматик доказан в работе Канадзавы (1996).

**Теорема 1.4.7.** Для произвольного натурального числа $k$ существует такой язык $L$, что $L$ задаётся грамматикой из класса $\mathcal{G}^{k+1}$, но не задаётся никакой грамматикой из класса $\mathcal{G}^k$.

Вторая глава содержит основные технические результаты данной работы. Сперва доказывается существование естественных инвариантов для сокращения произвольной строки типов. Далее доказывается, что класс $\mathcal{G}^1$ не замкнут относительно естественных операций на языках. В третьем параграфе доказываются теоремы о сведении посредством гомоморфизмов контекстно-свободных языков к языкам, заданным AB-грамматиками с однозначным присвоением типов. В последнем параграфе доказывается усиленный вариант аналогичной теоремы для языков, заданных грамматиками из класса $\mathcal{G}^2$.

В параграфе 2.1 вводится понятие счётчика по примитивному типу $p$, интерпретации в свободной группе, а также числителя и знаменателя произвольной строки типов. Доказываются основные технические свойства, требуемые в дальнейших построениях.

В параграфе 2.2 доказывается, что класс языков $\mathcal{L}^1$ не замкнут относительно большинства операций на языках.

**Теорема 2.2.2.** Для произвольного языка $L$ из класса $\mathcal{L}^1$ язык $\Sigma^+ - L$ не принадлежит классу $\mathcal{L}^1$.

**Теорема 2.2.3.** Если язык $L$ содержит хотя бы одно непустое слово, то язык $L^+$ не принадлежит классу $\mathcal{L}^1$.

**Теоремы 2.2.4–6.** Класс языков $\mathcal{L}^1$ не замкнут относительно конкатенации, объединения и пересечения.

Параграф 2.3 посвящён AB-грамматикам с однозначным присвоением типов. Основной целью данного параграфа является доказательство теоремы о том, что произвольный контекстно-свободный язык без пустого слова можно задать как гомоморфный прообраз некоторого языка, заданного AB-грамматикой с однозначным присвоением типов.

Параграф начинается с рассмотрения основной идеи доказательства финального результата. Идея заключается в имитации выбора из нескольких типов для одной буквы за счёт различных вариантов редукций некоторой строки типов. Для этого сперва для любых двух типов $A$ и $B$ нужно построить строку типов, которая редуцируется как к $A$, так и к $B$.

Однако данная цель не реализуема напрямую, что является следствием инвариантности относительно интерпретации в свободной группе. Таким образом, чтобы обойти данные инварианты, можно сделать преобразование, которое произвольные два типа преобразует в типы, интерпретации которых в свободной группе равны. Необходимое преобразование устроено следующим образом.

**Определение 2.3.1.** Определим функцию $\varphi$ как отображение из множества типов во множество типов по индукции:

- если $C$ — примитивный тип $p$, то $\varphi(C) = p/p$;
- если $C$ имеет вид $A/B$, то $\varphi(C) = \varphi(A)/\varphi(B)$;
- если $C$ имеет вид $B\backslash A$, то $\varphi(C) = \varphi(B)\backslash\varphi(A)$.

Для произвольных типов $A$ и $B$ существует строка типов, которая редуцируется как к $\varphi(A)$, так и к $\varphi(B)$. Однако от этих строк нам требуется больше, чем просто возможность редукции к нескольким разным типам. Нам необходимо, чтобы составленная из конкатенации такого рода «блоков» строка, в случае редукции к некоторому типу $C$, всегда редуцировалась «правильным» образом. Чтобы каждый из «блоков» сперва редуцировался к единственному типу, а затем полученные типы редуцировались между собой. Иначе говоря, данные «блоки» не должны «перемешиваться» между собой. Для этого вместо $\varphi$ используется модифицированное преобразование $\psi$.

**Определение 2.3.7.** Для произвольного типа $A$ положим значение $\psi(A)$ равным $\ell\backslash(\varphi(A)/r)$.

Уже для этого преобразования доказывается следующая лемма.

**Лемма 2.3.8.** Пусть $A$ и $B$ — различные типы вида $p$, $p/q$ или $(p/q)/s$. Тогда существует строка типов $u_{A,B}$, такая что:

I. $u_{A,B}$ редуцируется к $\psi(A)$ и к $\psi(B)$;

II. никакой собственный суффикс $u_{A,B}$ не редуцируется к $\psi(A)$, и никакой собственный префикс $u_{A,B}$ не редуцируется к $\psi(B)$.

Далее при помощи строки $u_{A,B}$ строится финальная строка $w_{A_1,\ldots,A_n}$, которая обобщает выбор между двумя типами на множественный случай:

$$
\begin{aligned}
&w_{A_1,\ldots,A_n} = x_1; u_{A_1,A_2}; x_2; u_{A_2,A_3}; x_3; \ldots; x_{n-1}; u_{A_{n-1},A_n}; x_n, \text{ где}\\
&x_1 = \psi(A_1)/e_1; e_1/\psi(A_2),\\
&x_i = \psi(A_{i-1})\backslash e_{2i-2}; (e_{2i-2}\backslash\psi(A_i))/e_{2i-1}; e_{2i-1}/\psi(A_{i+1}),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{для всех } i \in \{2,\ldots,n-1\},\\
&x_n = \psi(A_{n-1})\backslash e_{2(n-1)}; e_{2(n-1)}\backslash\psi(A_n).
\end{aligned}
$$

Таким образом, для построенной строки выполнена следующая лемма.

**Лемма 2.3.11.** Пусть $A_1, \ldots, A_n$, где $n \geqslant 1$ — произвольные попарно различные типы форм $p$, $p/q$ или $(p/q)/s$. Тогда строка типов $w_{A_1,\ldots,A_n}$ редуцируется к каждому из типов $\psi(A_1)$, $\ldots$, $\psi(A_n)$.

После этого фиксируется контекстно-свободная грамматика в нормальной форме Грейбах. Для каждого символа $a$ рассматриваются все правила вида $X \to a$, $X \to aY$ и $X \to aYZ$, где данный символ стоит справа на первом месте. Нетерминальные символы далее будут использованы как примитивные типы. Затем $a$ сопоставляется строка типов, которая имитирует выбор одного из этих правил. Итого, определяется гомоморфизм $h$, отображающий каждую букву $a$ в соответствующую строку типов. Технические условия, которым удовлетворяла $u_{A,B}$, устроены так, чтобы была выполнена следующая ключевая лемма.

**Лемма 2.3.13.** Если $h(y) = h(a_1 \ldots a_n)$ редуцируется к типу $X/X = \varphi(X)$, то существует строка типов формы $\varphi(C_1); \ldots; \varphi(C_n)$, которая редуцируется к $X/X$, где каждый тип $C_i$ выбран из множества:

$$\{\, X \mid X \to a_i \in R \,\} \cup \{\, X/Y \mid X \to a_iY \in R \,\} \cup \{\, (X/Y)/Z \mid X \to a_iZY \in R \,\}.$$

С использованием данного свойства для гомоморфизма $h$ доказывается ключевая лемма.

**Лемма 2.3.12.** Для произвольного слова $y \in \Sigma^+$ и произвольного нетерминального символа $X \in N$, слово $y$ выводимо в грамматике $G$ из нетерминального символа $X$ тогда и только тогда, когда его образ $h(y)$ редуцируется к типу $\varphi(X)$.

В итоге, используя понятие сводимости языков, доказывается основная теорема параграфа.

**Определение 2.3.17.** Язык $L_1 \in \Sigma^*$ *сводится посредством гомоморфизма* к языку $L_2 \in \Omega^*$ ($L_1 \leq L_2$), если существует гомоморфизм $g\colon \Sigma \to \Omega^*$ такой, что слово $w$ принадлежит $L_1$ тогда и только тогда, когда $h(w)$ принадлежит $L_2$.

**Теорема 2.3.18.** Для произвольного контекстно-свободного языка $L \subseteq \Sigma^+$ существует AB-грамматика с однозначным присвоением типов $G$, такая, что $L \leq L(G)$.

Таким образом, устанавливается, что AB-грамматики с однозначным присвоением типов достаточно сильны, чтобы закодировать произвольный контекстно-свободный язык.

В параграфе 2.4 показывается, что произвольный контекстно-свободный язык эквивалентен по сводимости языку из класса $\mathcal{L}^2$.

**Теорема 2.4.2.** Для произвольного контекстно-свободного языка $L \subseteq \Sigma^+$ существует $G$ — грамматика из класса $\mathcal{G}^2$ такая, что $L \leq L(G)$ и $L(G) \leq L$.

Третья глава посвящена алгоритмическим свойствам ранее введенных формализмов. Сперва рассматриваются грамматики из класса $\mathcal{G}^2$, затем из класса $\mathcal{G}^1$ и в конце AB-грамматики с однозначным присвоением типов.

В параграфе 3.1 показывается, как при помощи теоремы 2.4.2 доказать, что классические задачи для класса грамматик $\mathcal{G}^2$ алгоритмически неразрешимы.

**Теорема 3.1.1.** Следующие задачи неразрешимы для грамматик из класса $\mathcal{G}^2$:

- проверить по двум грамматикам из класса $\mathcal{G}^2$, равны ли заданные ими языки;
- по произвольной грамматике $G$ из класса $\mathcal{G}^2$ проверить, является ли язык, заданный грамматикой $G$, регулярным.

В параграфе 3.2 доказываются следующие теоремы для класса $\mathcal{G}^1$.

**Теорема 3.2.6.** Задача определения для произвольной контекстно-свободной грамматики $G$

того, задаётся ли язык $L(G)$ некоторой грамматикой из класса $\mathcal{G}^1$, алгоритмически неразрешима.

**Теорема 3.2.7.** Задача определения для произвольных грамматик $G_1$ и $G_2$ из класса $\mathcal{G}^1$ того, является ли пересечение языков $L(G_1) \cap L(G_2)$ бесконечным языком, алгоритмически неразрешима.

**Теорема 3.2.9.** Задача определения для произвольных грамматик $G_1$ и $G_2$ из класса $\mathcal{G}^1$, где количество символов в алфавитах обеих грамматик равно 11, того, является ли пересечение языков $L(G_1) \cap L(G_2)$ пустым, алгоритмически неразрешима.

Доказательство неразрешимости основано на сведении известной неразрешимой задачи. В данном случае в качестве такой задачи выступает проверка существования решения для произвольной системы соответствий Поста.

Теорема 3.2.9 усиливает ранее известный результат из работы Форе (2004), где доказана аналогичная теорема для AB-грамматик, при этом требуется алфавит из существенно большего количества символов.

В параграфе 3.3, при помощи теоремы из параграфа 2.3, доказывается усиление классической теоремы Грейбах о самом трудном языке.

**Следствие 3.3.1.** Существуют алфавит $\Omega$ и язык $L_0$, заданный AB-грамматикой с однозначным присвоением типов, такие, что для произвольного контекстно-свободного языка $L \subseteq \Sigma^+$ верно $L \leq L_0$.

А также решается следующий вопрос, поставленный в вышеупомянутой статье Форе.

**Теорема 3.3.2.** Задача определения для произвольных AB-грамматик с однозначным присвоением типов $G_1$ и $G_2$ того, что $L(G_1) \subseteq L(G_2)$, алгоритмически неразрешима.

## Заключение

В первой главе работы введены базовые категориальные грамматики с условием присвоения не более $k$ типов одному символу. Доказана теорема об иерархии по параметру $k$ для классов задаваемых языков.

Во второй главе доказана незамкнутость класса $\mathcal{G}^1$ относительно большинства естественных операций на языках. Затем рассмотрены AB-грамматики с однозначным присвоением типов. Доказана возможность сведения любого контекстно-свободного языка к языку, заданному данным формализмом. В частности, разработана техника кодирования множе-

ственного присвоения типов за счёт единственной строки. Однако, техника нетривиальна, и пока неизвестно, существует ли более простой способ. Более того, имеются основания полагать, что, в силу крайне ограниченных возможностей, подобные конструкции будут настолько же технически сложными. Доказано, что произвольный контекстно-свободный язык эквивалентен по сводимости некоторому языку из класса $\mathcal{L}^2$.

В третьей главе рассматриваются алгоритмические свойства грамматических формализмов, введённых в первой главе. Доказано, что для грамматик из класса $\mathcal{G}^2$ неразрешимы классические алгоритмические задачи. Дальше доказано, что задача проверки пустоты пересечения языков, заданных грамматиками из класса $\mathcal{G}^1$, неразрешима. Более того, последнее верно при фиксированном алфавите из 11 символов. В третьем параграфе рассматриваются AB-грамматики с однозначным присвоением типов. Используя теорему из второй главы, доказывается, что задача проверки вложения языков, заданных грамматиками из данного класса, неразрешима.

Стоит отметить, что, хотя и доказан ряд теорем для класса $\mathcal{G}^1$, до сих пор остаются открытыми многие вопросы, касающиеся данного формализма. Например, неизвестно, разрешима ли задача проверки включения языков, порожденных данными грамматиками. Предполагается, что техника, разработанная в данной диссертации, поможет решить эту задачу.

Методы, разработанные во второй главе, применимы в большинстве грамматических формализмов, основанных на AB-грамматиках. Следовательно, последующие работы могут использовать полученный подход для исследований. Так, результаты о сложности в более богатых исчислениях могут быть доказаны напрямую, используя разработанную конструкцию кодирования.

## Благодарности

Автор выражает глубокую благодарность своему научному руководителю — доктору физико-математических наук Степану Львовичу Кузнецову за выбор темы, постановку задач и постоянную поддержку. Особенную признательность автор адресует доктору физико-математических наук, профессору Мати Рейновичу Пентусу за внимательное прочтение рукописей и ценнейшие замечания. Кроме того, автор приносит благодарность Ph.D., профессору Александру Сергеевичу Охотину за полезные обсуждения и плодотворную совместную деятельность, а также всем сотрудникам отдела математической логики за создание тёплой и дружелюбной атмосферы.

## Публикации автора по теме диссертации


1. Вишникин М.Е. Базовые категориальные грамматики с однозначным присвоением типов // Вестник Московского университета. Серия 1: Математика. Механика. – 2022. – № 2. – С. 64–67.

2. Вишникин М.Е. Базовые категориальные грамматики с ограничением на количество присваиваемых категорий // Сибирские электронные математические известия. – 2025. – Т. 22, № 1. – С. 343–353.

3. Вишникин М.Е. Алгоритмические свойства базовых категориальных грамматик с однозначным присвоением категорий // Доклады Российской академии наук. Математика, информатика, процессы управления. – 2025. – Т. 523. – С. 21–25.


*Тезисы конференций:*


4. Вишникин М.Е. Базовые категориальные грамматики с однозначным присвоением типов // тез. докл. в сборнике Материалы Международного молодежного научного форума «Ломоносов-2021», М.: МАКС Пресс, 2021.

5. Вишникин М.Е., Охотин А.С. Категориальные грамматики с однозначным присвоением категорий // тез. докл. в эл. сборнике Международной конференции «Мальцевские чтения», Новосибирск, 2024.

6. Вишникин М.Е. Базовые категориальные грамматики с ограничением присваиваемые категории // тез. докл. в сборнике Материалы Международного молодежного научного форума «Ломоносов-2025», М.: МАКС Пресс, 2025.