ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Вишникина Максима Евгеньевича
«Категориальные грамматики с ограничением
на количество присваиваемых типов»,
представленную на соискание учёной степени
кандидата физико-математических наук
по специальности 1.1.5 «Математическая логика, алгебра,
теория чисел и дискретная математика»

В 1956 году Н. Хомский определил четыре класса порождающих грамматик для описания синтаксиса естественных и формальных языков. Одним из этих классов являются контекстно-свободные грамматики (КС-грамматики). Принципиально иной подход для описания языков основан на использовании синтаксических типов (или категорий), задающих для каждого слова в предложении его синтаксическую роль. Впервые подобный подход был описан К. Айдукевичем в 1935 году, а определение категориальных грамматик в современном виде было дано Й. Бар-Хиллелом в 1953 году. В честь Айдукевича и Бар-Хиллела такие грамматики называются AB-грамматиками. Синтаксические типы этих грамматик строятся из элементов изначально заданного множества примитивных типов с помощью бинарных связок левого деления ($\backslash$) и правого деления ($/$). С лингвистической точки зрения тип (A/B) означает, что слово имеет тип A, а справа от него стоит зависимое слово типа B. Вскоре было установлено, что класс языков, порождаемых AB-грамматиками, совпадает с классом КС-языков без пустого слова. Помимо AB-грамматик изучались и различные их варианты и обобщения: грамматики Ламбека (Й. Ламбек, 1958), разрывные грамматики Ламбека (Г. Моррилл, Х.-М. Меренсиано, 1996), комбинаторные категориальные грамматики (М. Стидман, 2001), категориальные грамматики зависимостей (М. И. Дехтярь, А. Я. Диковский, 2004) и пр.

Однако вместо того чтобы расширять возможности категориальных грамматик, можно изучать и более узкие подклассы AB-грамматик. Например, можно ограничиться базовыми категориальными грамматиками, в категориях которых допустимо деление только на примитивные типы. Хорошо известно, что такое ограничение не приводит к уменьшению класса порождаемых языков. В работах М. Канадзавы исследовались AB-грамматики с ограничением на число типов, присваиваемых каждому символу. Канадзава рассматривал k-значные грамматики (k-valued grammars), в которых каждому символу приписывается не более k типов, и их частный случай при $k = 1$ — жёсткие грамматики (rigid grammars). Канадзава показал, что классы языков, порождаемых k-значными грамматиками, образуют бесконечную иерархию, а также, что для любого фиксированного k класс k-значных языков является изучаемым в пределе в смысле Голда. Однако до сих пор были слабо исследованы алгебраические и алгоритмические свойства таких грамматик. Можно упомянуть результат А. Форе 2004 года о том, что проблема пустоты пересечения неразрешима для k-значных грамматик при любом $k \geqslant 1$, а также результат А. Н. Сафиуллина 2007 года о том, что любая грамматика Ламбека эквивалентна грамматике Ламбека, в которой каждому символу сопоставлен единственный тип.

Диссертация М. Е. Вишникина посвящена изучению свойств AB-грамматик с ограничением на количество типов, присваиваемых каждому символу. В ней изучаются выразительные возможности таких грамматик, свойства замкнутости классов порождаемых ими языков, а также алгоритмические проблемы для этих грамматик. Диссертация состоит из титульного листа, оглавления, введения, трёх глав основной

части, заключения и списка литературы.

Во введении обосновывается актуальность темы исследования и приводится обзор имеющих результатов.

В первой главе приведены основные определения: КС-грамматики и АВ-грамматики, k-значные грамматики, базовые категориальные грамматики (БКГ). Далее показано, как по произвольной БКГ построить эквивалентную ей КС-грамматику. В отличие от стандартного преобразования, конструкция М. Е. Вишникина даёт КС-грамматику, в которой правые части всех правил содержат ровно один терминал. Далее вводится обозначение $\mathcal{G}^k$ для класса всех k-значных БКГ и обозначение $\mathcal{L}^k$ для соответствующего класса языков. Основным результатом первой главы является теорема об иерархии: для любого $k \geqslant 1$ имеет место включение $\mathcal{L}^k \subsetneq \mathcal{L}^{k+1}$. Эта теорема аналогична результату Канадзавы, однако не является его прямым следствием, поскольку преобразование произвольной АВ-грамматики к БКГ изменяет число присваиваемых типов. Для доказательства этой теоремы М. Е. Вишникин рассматривает языки $L_n = \{\, a^i \mid i \leqslant n \,\}$ и показывает, что из $L_n \in \mathcal{L}^k$ следует, что $L_{n+1} \in \mathcal{L}^{k+1}$, а также что не существует класса $\mathcal{L}^k$, содержащего все L_n.

Вторая глава посвящена изучению свойств замкнутости класса $\mathcal{L}^1$, а также представлению КС-языков посредством гомоморфизмов. Доказывается, что класс $\mathcal{L}^1$ не замкнут относительно конкатенации, усечённой итерации (плюс Клини), объединения, пересечения и дополнения, однако замкнут относительно обращения. Отметим, что на самом деле доказаны даже более сильные результаты. Так, показано, что существуют два языка из класса $\mathcal{L}^1$, пересечение которых вообще не является контекстно-свободным. Для дополнения установлено, что дополнение языка из класса $\mathcal{L}^1$ никогда не принадлежит $\mathcal{L}^1$. Далее вводится понятие сводимости посредством гомоморфизмов, аналогичное понятию m-сводимости или полиномиальной сводимости: язык $L_1 \subseteq \Sigma^*$ сводится посредством гомоморфизма к языку $L_2 \subseteq \Omega^*$, если существует гомоморфизм $h\colon \Sigma^* \to \Omega^*$ такой, что $w \in L_1$ тогда и только тогда, когда $h(w) \in L_2$ для любого $w \in \Sigma^*$ (обозначается $L_1 \leqslant L_2$). Основной результат состоит в том, что для любого КС-языка L без пустого слова существует АВ-грамматика G с однозначным присвоением типов (не обязательно БКГ), такая что L сводится к $L(G)$ посредством гомоморфизма. Для доказательства этого результата М. Е. Вишникин рассматривает КС-грамматику в нормальной форме Грейбах, порождающую язык L, строит по ней эквивалентную БКГ, а далее заменяет каждый примитивный тип p на тип (p/p). Результат такой замены в типе A обозначается через $\phi(A)$, а через $\psi(A)$ обозначается тип $\ell\backslash(\phi(A)/r)$. Далее доказывается, что по любым типам $A_1, \ldots, A_n$, входящим в БКГ, можно построить строку типов $w_{A_1,\ldots,A_n}$, которая редуцируется к типам $\psi(A_1), \ldots, \psi(A_n)$. Эта строка моделирует выбор одного из n типов, а замена p на (p/p) позволяет обойти ограничение на сохранение интерпретации типов в свободной группе при любой последовательности редукций. В качестве следствия получается усиление известной теоремы Грейбах о самом трудном КС-языке: существует язык L_0, заданный АВ-грамматикой с однозначным присвоением типов, такой, что произвольный контекстно-свободный язык L сводится посредством гомоморфизма к L_0. Далее доказывается усиленный вариант этого результата для класса $\mathcal{G}^2$: для произвольного контекстно-свободного языка L существует АВ-грамматика $G \in \mathcal{G}^2$ такая, что $L \leqslant L(G)$ и $L(G) \leqslant L$. Усиление состоит в том, что в случае $\mathcal{G}^2$ можно ограничиться только базовыми грамматиками, а сводимость имеет место в обе стороны.

Третья глава посвящена исследованию алгоритмических проблем для БКГ. Хорошо известно, что многие естественные алгоритмические проблемы неразрешимы в

классе всех КС-грамматик. В третьей главе исследуются некоторые проблемы для грамматик из классов $\mathcal{G}^1$ и $\mathcal{G}^2$. Используя полученные выше результаты о представлении языков с помощью гомоморфизмов, М. Е. Вишникин доказывает, что что в классе $\mathcal{G}^2$ неразрешимы проблемы эквивалентности двух грамматик и регулярности порождаемого языка. Далее установлено, что не существует алгоритма, который по произвольной КС-грамматике G определял бы, принадлежит ли язык $L(G)$ классу $\mathcal{L}^1$. Доказывается, что для грамматик из класса $\mathcal{G}^1$ неразрешимы проблемы бесконечности пересечения двух языков и пустоты пересечения. Отметим, что неразрешимость второй проблемы была установлена в 2004 году А. Форе, однако конструкция М. Е. Вишникина позволяет доказать неразрешимость уже для алфавита мощности 11, в то время как в доказательстве Форе необходимы 63 символа. Доказательства для $\mathcal{G}^1$ основаны на сведении к рассматриваемым проблемам известной неразрешимой проблемы соответствий Поста. Они аналогичны доказательству неразрешимости проблем тотальности, бесконечности пересечения и пустоты пересечения для произвольных КС-грамматик, но более сложны из-за ограничений на количество типов. Для общих АВ-грамматик с однозначным присвоением типов доказывается неразрешимость проблемы включения.

В заключении подводятся итоги исследования и формулируются некоторые вопросы для дальнейшего изучения.

Общий объём работы составляет 92 страницы, список литературы содержит 49 наименований.

Существенных недостатков в диссертации нет. В качестве небольшого замечания можно указать, что в параграфе 2.2 не рассматриваются такие операции над языками, как пересечение с регулярными языками, неукорачивающий гомоморфизм и обращение гомоморфизма. Эти операции также весьма естественны, в частности, они входят в определение абстрактного семейства языков. Незамкнутость класса $\mathcal{L}^1$ относительно этих операций устанавливается непосредственно, и было бы хорошо упомянуть этот результат.

Также диссертация содержит некоторые неточности и опечатки.

1) На стр. 23 в определении 1.1.2 написано, что отношение $\to^*$ является рефлексивно-транзитивным замыканием отношения R. На самом деле это рефлексивно-транзитивное замыкание отношения, порождённого R на множестве всех слов.

2) На стр. 26 в примере 2 написан тип $q\backslash(p/p)/p$ вместо $q\backslash((p/p)/p)$.

3) На стр. 28 в доказательстве теоремы 1.3.3 рассматриваются нетерминальные символы $h_{i,j}$ для $i \in \{1, \ldots, k\}$, $j \in \{1, \ldots, m\}$, однако далее используется нетерминал $h_{0,0}$. В определении должно быть написано $i \in \{0, \ldots, k\}$, $j \in \{0, \ldots, m\}$

4) На стр. 30 в строке 4 снизу написан лишний символ C.

5) На стр. 35 в строке 4 написано $L(G)$ вместо $L(G')$.

6) На стр. 40 в шаге индукции не указано, что $num(B\backslash C) = num(C/B)$, а значит, доказательство справедливо как для типа $B\backslash C$, так и для C/B.

7) На стр. 43 в доказательстве теоремы 2.2.5 вторая грамматика порождает не язык $L_2 = \{ab\}$, а язык $\{ab^n \mid n \geqslant 0\}$. Язык L_2 порождается грамматикой $a \triangleright p$, $b \triangleright (p\backslash r)$.

8) На стр. 55 в строке 4 написано $x_{t,B}$ вместо $x_{A,t}$.

9) На стр. 55 примитивный тип r используется как один из сторожевых типов. Однако до этого r использовался в качестве результирующего типа.

10) На стр. 58 в строках 2 и 3 написано x вместо $z_{A,B}$.

11) На стр. 59 в строке между деревьями вывода написано u вместо $u_{A,B}$, а в третьем дереве — v вместо u.

12) На стр. 60 в строке 4 написано r вместо ℓ, в строке 6 — $\phi(B)$ вместо $\phi(A)$, а в выключных формулах — o_1 и o_2 вместо o_3 и o_4 соответственно.

13) На стр. 62 в строке 3 после деревьев написано $u_{A_n,A_{n+1}}$ вместо u_{A_{n-1},A_n}, а в последнем типе этой же строки написано $\psi(A_{n-1})$ вместо $\psi(A_n)$.

14) На стр. 62 в строке 4 второго абзаца после деревьев написано, что по индукционному предположению $w_{A_1,\ldots,A_n}$ редуцируется к $\psi(A_n); x_{n+1}$. На самом деле $w_{A_1,\ldots,A_n}$ редуцируется к $\psi(A_n)$, а значит, вся строка $w_{A_1,\ldots,A_{n+1}}$ редуцируется к $\psi(A_n); x_{n+1}$ и далее к $\psi(A_{n+1})$.

15) На стр. 65 в третьей выключной формуле в начале обеих строк пропущено ℓ.

16) На стр. 68 в конце второго абзаца случай типа r невозможен не по утверждению 2.3.15, а по утверждению 2.3.14.

17) На стр. 69–70 в доказательстве для шага индукции перепутаны m и n. В самом начале написано, что $y = a_1 \ldots a_m$, но последовательность типов имеет вид $\phi(C_1) \ldots \phi(C_n)$. А в строке перед деревом она уже записана как $\phi(C_1); \ldots; \phi(C_m)$.

18) На стр. 70 перед определением 2.3.17 написано, что будет введено понятие сильной сводимости. Однако далее определяется сводимости посредством гомоморфизма, а термин «сильная сводимость» больше нигде не встречается.

19) На стр. 70 в определении 2.3.17 написано $L_1 \in \Sigma^*$ и $L_2 \in \Omega^*$ вместо $L_1 \subseteq \Sigma^*$ и $L_2 \subseteq \Omega^*$ соответственно, а гомоморфизм обозначается сначала h, а потом g.

20) На стр. 77 в самом начале доказательства предложения 3.2.5 известным результатом является эквивалентность не первого и второго, а первого и третьего утверждений.

21) На стр. 81 в строках 3 и 6 после первой выключной формулы написано $\alpha_2(x_1)$ и $\alpha_2(x_i)$ вместо $(\alpha_2(x_1))^R$ и $(\alpha_2(x_i))^R$ соответственно.

В работе имеются и другие опечатки. Например, в некоторых местах вместо длинного тире «—» используется среднее «–»; пустое множество обозначается то $\emptyset$ (стр. 22 и 84), то $\varnothing$ (стр. 39); гомоморфизм определяется то как отображение вида $h\colon \Sigma \to \Delta^*$ (стр. 19, 70), то как отображение вида $h\colon \Sigma^* \to \Delta^*$ (стр. 78); некоторые предложения «разорваны» на части (стр. 40, после выключной формулы; стр. 54, абзац 4 снизу; стр. 58, в самом низу; стр. 63, первый абзац; стр. 71, абзац после примера 9; стр. 84 конец доказательства). Имеется и небольшое количество орфографических и пунктуационных ошибок (на стр. 46 в строке 3 перед леммой 2.3.2 написано «в» вместо «к»; на стр. 68 в предпоследней строке доказательства написано «редуцированны»

вместо «редуцированы»; на стр. 70 пропущена запятая перед словом «используя»; на стр. 75 в строке 6 пропущены запятые вокруг причастного оборота «порожденных грамматиками из класса $\mathcal{G}^2$» и т. п.).

Все перечисленные замечания носят технический характер и не снижают научной ценности проведённых исследований и корректности полученных результатов.

Таким образом, диссертация Вишникина Максима Евгеньевича представляет собой завершённое научное исследование, посвящённое решению актуальных проблем математической логики. Полученные в ней результаты являются весомым вкладом в области математической лингвистики и её приложений и могут быть использованы в дальнейших исследованиях категориальных грамматик. Все доказываемые в диссертационной работе результаты являются новыми и принадлежат автору. Все утверждения снабжены строгими математическими доказательствами, поэтому их справедливость не вызывает сомнений. Автор опубликовал шесть работ, в том числе три — в ведущих рецензируемых изданиях, рекомендованных ВАК для публикации результатов диссертаций и индексируемых в международных системах цитирования. Текст оформлен качественно и аккуратно. Автореферат правильно отражает содержание диссертации.

Диссертация Вишникина Максима Евгеньевича «Категориальные грамматики с ограничением на количество присваиваемых типов» удовлетворяет требованиям п. 9 «Положения о порядке присуждения учёных степеней», утверждённого постановлением Правительства РФ от 24.09.2013 №842, а её автор заслуживает присуждения учёной степени кандидата физико-математических наук по специальности 1.1.5 — «Математическая логика, алгебра, теория чисел и дискретная математика».

Официальный оппонент:
кандидат физико-математических наук (специальность 01.01.06 «Математическая логика, алгебра и теория чисел»), доцент,
доцент кафедры информатики
факультета прикладной математики и кибернетики
ФГБОУ ВО «Тверской государственный университет»
170100, г. Тверь, ул. Желябова, 33

6 августа 2025 года Карлов

Карлов Борис Николаевич
+7 (903) 806-03-99, bnkarlov@gmail.com

Подпись руки Карлова Б. Н.
Удостоверяю Сухих Н. Е. —
проректор по образовательной деятельности