А
П-71

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной
и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

Выпуск 8

А. В. ГЛАДКИЙ, А. Я. ДИКОВСКИЙ

ТЕОРИЯ ФОРМАЛЬНЫХ ГРАММАТИК
И
ЯЗЫКОВ

Москва — 1970

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

65106

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа
по экспериментальной и прикладной
лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

В ы п у с к 6

А.В. Гладкий, А.Я. Диковский

ТЕОРИЯ ФОРМАЛЬНЫХ ГРАММАТИК И ЯЗЫКОВ

М о с к в а - 1970

4
П-718
БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
ИН-ТА ЯЗЫКОЗНАНИЯ АН СССР
65106

Настоящий препринт содержит текст доклада на 2-ой Всесоюзной конференции по программированию (Новосибирск, февраль 1970 г.), исправленный и несколько дополненный по сравнению с вариантом, опубликованным в Трудах конференции.

ВВЕДЕНИЕ

Теория формальных грамматик и языков является центральным разделом математической лингвистики – специфической математической дисциплины, ориентированной на приложения в области изучения структуры естественных и искусственных языков. Эта теория возникла в 50-е годы в работах американского лингвиста Н. Хомского (хотя основы аппарата, на котором она базируется, были разработаны в математической логике значительно раньше). При этом Н. Хомский исходил исключительно из потребностей лингвистики, по традиции понимаемой как наука о строении естественных или, во всяком случае, универсальных языков; однако очень скоро обнаружилось, что методы его теории не в меньшей степени приложимы к искусственным специализированным языкам, и в частности – к языкам программирования; это, впрочем, довольно понятно, если вспомнить, что последние проще устроены и лучше поддаются формализации, чем естественные языки.

По характеру используемого математического аппарата теория формальных грамматик и языков близка к теории алгоритмов и особенно к теории автоматов, имея с последней много параллелей, точек соприкосновения и даже переплетений.

Цель настоящей публикации – дать краткий обзор некоторых важнейших направлений теории формальных грамматик и языков. Чтобы охватить в с е её разделы, потребовалось бы намного больше места и времени, чем мы в состоянии затратить. Выбранные нами направления принадлежат к тем, которые, с нашей точки зрения, являются в настоящее время наиболее актуальными. Именно, мы остановились на следующих вопросах: автоматные характеристики языков (раздел 2); сложность выводов в грамматиках, сложность распознавания языков различных классов (раздел 3); некоторые промежуточные классы грамматик и языков (раздел 4); надстройки,

управление выводом (раздел 5). Читатель, знакомый с предметом, сразу заметит отсутствие таких, например, "традиционных" разделов, как алгоритмические проблемы, категориальные грамматики, грамматики зависимостей. Это ни в коей мере не означает, разумеется, что мы считаем названные вопросы менее важными, чем включенные в обзор. Не затронуты и некоторые более новые области: окрестностные грамматики (см., например, [4, 9, I4, 24]), изучение порождаемых грамматиками множеств систем составляющих и деревьев подчинения (например, [3]) и ряд других. В частности, за пределами обзора остались работы, посвященные грамматикам, порождающим не цепочки, а деревья - хотя именно это направление авторы считают наиболее перспективным и важным. Это объясняется, с одной стороны, тем, что необходимый для изложения соответствующих понятий аппарат существенно отличен от "традиционного", и его рассмотрение резко увеличило бы объем, с другой - предварительным характером имеющихся пока что в этой области результатов.

Отметим также, что в нашем изложении нигде не затрагиваются вопросы интерпретации.

Сделаем ещё несколько замечаний по поводу некоторых особенностей изложения.

1. Все используемые понятия, включая основные понятия теории, определяются в тексте.

2. Мы стараемся избегать формулирования тривиальных утверждений ; однако в некоторых случаях, когда такие формулировки необходимы для полноты картины, этот принцип нарушается.

3. Доказательства не приводятся.

4. Мы не всюду стремимся к абсолютной строгости формулировок, но всякий раз, когда допускаются те или иные вольности, способ уточнения очевиден.

5. Всюду, где это возможно, формулировки утверждений сопровождаются ссылками на источники ; при этом, как

правило, предпочтение отдается более доступным. В частности, в ряде случаев мы ссылаемся на работы обобщающего характера - такие, как [8, 23, 30] - а не на те, в которых соответствующий результат получен впервые.

6. Помещенный в конце список литературы ни в малейшей степени не претендует на полноту.

7. При написании обзора мы всячески старались придерживаться принципов, изложенных в основополагающем труде [13]. Теперь, следуя одному из этих принципов, мы делаем отступ в одну строку и переходим к разделу I.

1. ОСНОВНЫЕ ПОНЯТИЯ

Мы будем рассматривать непустое конечное множество V, которое будем называть *словарем*, а его элементы - *элементарными символами* (или просто *символами*). Произвольную конечную последовательность ω элементов V будем называть *цепочкой* в словаре V; число членов этой последовательности назовем *длиной цепочки* (обозначение $|\omega|$). Например, ааббаа - цепочка в словаре $\{а,в,б,г\}$, $|аббаа|=5$. Будет рассматриваться, в частности, и *пустая цепочка* (обозначение Λ).

Множество всех цепочек в словаре V мы будем обозначать через V^*, множество всех непустых цепочек в V - через V^+. *Конкатенацией* цепочек φ и ψ (обозначение: $\varphi\psi$) называется операция, состоящая в записи ψ вслед за φ. Например, ввавбгб - конкатенация цепочек вв и вавбгб. В частности, для любой цепочки ω $\omega\Lambda=\Lambda\omega=\omega$. Операция конкатенации ассоциативна, но не коммутативна.

Произвольное множество $L \subseteq V^*$ мы будем называть *языком* в словаре V.

Теоретико-множественные операции над языками будут обозначаться обычным образом: $\cup$ (объединение), $\cap$

(пересечение), — (вычитание).

Нам понадобятся ещё две операции над языками: а) у м н о ж е н и е : $L_1 L_2 = \{\varphi\psi \mid \varphi \in L_1, \psi \in L_2\}$;
б) и т е р а ц и я: $L^* = \bigcup_{i=0}^{\infty} L^i$ где $L^i = \underbrace{L \dots L}_{i}$, $L^0 = \{\Lambda\}$.

(П о р о ж д а ю щ е й) г р а м м а т и к о й мы будем называть упорядоченную систему $\Gamma = (V, V_1, \mathcal{T}, R)$ где V и V_1 – непересекающиеся непустые конечные множества, $\mathcal{T}$ – элемент V_1, R – конечное множество цепочек вида $\varphi \to \psi$, где $\varphi, \psi \in (V \cup V_1)^*$ и $\to \notin V \cup V_1$. Множества V и V_1 называются соответственно о с н о в н ы м и и в с п о м о г а т е л ь н ы м с л о в а р я м и грамматики Γ, а их элементы соответственно о с н о в н ы м и и в с п о м о г а т е л ь н ы м и с и м в о л а м и Γ (вместо "основной" и "вспомогательной" часто говорят соответственно "терминальный" и "нетерминальный"), $\mathcal{T}$ – н а ч а л ь н ы м с и м в о л о м, R – с х е м о й и элементы R – п р а в и л а м и Γ.

Если $\omega, \omega' \in (V \cup V_1)^*$, $\omega = \xi_1 \varphi \xi_2$, $\omega' = \xi_1 \psi \xi_2$ и $\varphi \to \psi \in R$, мы будем говорить, что ω' п о л у ч а е т с я и з ω п р и м е н е н и е м п р а в и л а $\varphi \to \psi$ к в ы д е л е н н о м у в х о ж д е н и ю φ. Если ω' получается из ω применением некоторого правила Γ, мы говорим, что ω' н е п о с р е д с т в е н н о в ы в о д и м а из ω в Γ, и пишем $\omega \vdash_\Gamma \omega'$. Транзитивное замыкание отношения $\vdash_\Gamma$ будет обозначаться $\vdash_\Gamma$ (или $\vdash$). $\varphi \vdash_\Gamma \psi$ читается "ψ в ы в о д и м а из φ в Γ". Таким образом, $\varphi \vdash_\Gamma \psi$ означает, что существует последовательность цепочек $\varphi_0, \varphi_1, \dots, \varphi_n$ такая, что для каждого $i, 0 \le i < n$, $\varphi_i \vdash_\Gamma \varphi_{i+1}$, $\varphi_0 = \varphi$ и $\varphi_n = \psi$.

Такая последовательность называется в ы в о д о м ψ из φ. Всякий вывод цепочки в словаре V из $\mathcal{T}$ называется п о л н ы м. Множество $L(\Gamma) = \{x \in V^* \mid \mathcal{T} \vdash_\Gamma x\}$ называется я з ы к о м, п о р о ж д а е м ы м г р а м м а т и к о й Γ.

Две грамматики называются эквивалентными, если они порождают один и тот же язык. (Иногда в этом смысле говорят о слабой эквивалентности в противоположность "сильной", под которой понимается, что грамматики не только порождают одни и те же цепочки, но и приписывают одинаковым цепочкам одинаковые описания структуры ; понятие сильной эквивалентности имеет смысл лишь для грамматик, естественным образом сопоставляющих цепочкам такие описания - см. замечание в конце настоящего раздела.)

В дальнейшем мы будем рассматривать только такие грамматики, у которых левые части всех правил содержат хотя бы по одному вхождению вспомогательного символа. Легко показать, что для всякой грамматики можно построить эквивалентную, удовлетворяющую этому условию.

Основным объектом изучения теории грамматик являются не произвольные грамматики, а грамматики некоторых специальных типов, наиболее важных как в теоретическом, так и в прикладном аспекте. Выделение этих типов производится по виду правил. Из них главными являются следующие:

1. Неукорачивающие грамматики грамматики, у которых для любого правила $\varphi \to \psi$ справедливо $|\varphi| \leq |\psi|$. (Сами правила такого вида также называются неукорачивающими).

2. Грамматики составляющих (или непосредственно составляющих), сокращенно НС-грамматики - грамматики с правилами вида $\eta_1 A \eta_2 \to \eta_1 \xi \eta_2$, где A - вспомогательный символ, ξ - произвольная непустая цепочка и η_1, η_2 - произвольные цепочки (сами правила такого вида называются НС - правилами).

3. Контекстно-свободные (или бесконтекстные) грамматики, сокращенно КС - грамматики - грамматики с правилами вида $A \to \omega$ (КС - правилами), где A - вспомогательный символ и ω - непустая цепочка.

4. Линейные грамматики – КС – грамматики, правые части правил которых содержат не более чем по одному вхождению вспомогательного символа.

5. Односторонние линейные грамматики: левосторонние – с правилами вида $A \to xB$ и $A \to x$ и правосторонние – с правилами вида $A \to Bx$ и $A \to x$; в обоих случаях A, B – вспомогательные символы, а x – непустая цепочка основных символов.

6. Автоматные грамматики (A – грамматики) – односторонние линейные грамматики, у которых в каждом правиле цепочка x состоит из одного символа a.

Из определений непосредственно ясно, что в этой последовательности каждый следующий класс содержится в предыдущем.

Если обозначить класс языков, порождаемых произвольными грамматиками, через $\mathcal{L}_0$, классы языков, порождаемых грамматиками типов 1–4, соответственно через $\mathcal{L}_1, \ldots, \mathcal{L}_4$, а классы языков, порождаемых левосторонними/правосторонними грамматиками типов 5,6 – через $\mathcal{L}_5^л, \mathcal{L}_6^л / \mathcal{L}_6^п, \mathcal{L}_6^п$, то оказывается [23, 8], что между ними имеют место следующие соотношения:

$$\mathcal{L}_0 \supset \mathcal{L}_1 = \mathcal{L}_2 \supset \mathcal{L}_3 \supset \mathcal{L}_4 \supset \mathcal{L}_5^п = \mathcal{L}_5^л = \mathcal{L}_6^п = \mathcal{L}_6^л .$$

Таким образом, восьми специальным классам грамматик (не считая $\mathcal{L}_0$) отвечают всего четыре класса языков. Языки этих классов называются соответственно НС –, КС –, линейными и A – языками (A – языки называют ещё регулярными множествами). Мы будем придерживаться следующих очевидных обозначений для указанных классов языков: $\mathcal{L}$(НС), $\mathcal{L}$(КС), $\mathcal{L}$(Л), $\mathcal{L}(A)$. Кроме того, будем обозначать через $\mathcal{L}(\Gamma)$ класс языков, порождаемых произвольными грамматиками. Приведем примеры языков, принадлежащих этим классам.

$\{a^k b^l / k, l = 1, 2, \ldots\} \in \mathcal{L}(A)$;

$\{a^n b^n / n = 1, 2, \ldots\}, \{x \hat{x} / x \in \{a, b\}^+\} \in \mathcal{L}(\Pi) \setminus \mathcal{L}(A)$

($\hat{x}$ – зеркальный образ x);

$\{a^n b^n / n = 1,2,\ldots\}^* \in \mathcal{L}(КС) - \mathcal{L}(Л)$;
$\{a^n b^n c^n / n = 1,2,\ldots\}, \{xx / x \in \{a,b\}^+\} \in \mathcal{L}(НС) - \mathcal{L}(КС)$.

Замечание. Каждый способ вывода ("история порождения") цепочки $x \in L(\Gamma)$ в грамматике Γ может рассматриваться как некоторое описание синтаксической структуры этой цепочки. В случае, когда Γ - НС -грамматика, такое описание по существу равносильно широко употребительному в лингвистике представлению структуры предложения с помощью системы составляющих ; именно этим обстоятельством обусловливается особая роль НС -грамматик в задачах моделирования естественных - а также и искусственных - языков.

2. АВТОМАТНЫЕ ХАРАКТЕРИСТИКИ ОСНОВНЫХ КЛАССОВ ГРАММАТИК

Все автоматы, которыми мы будем пользоваться, суть частные случаи (н е д е т е р м и н и р о в а н н о й) м а ш и н ы Т ь ю р и н г а с о в х о д н о й л е н т о й, которая может быть описана следующим образом. Машина имеет две ленты - в х о д н у ю и р а б о ч у ю. Каждой ленте сопоставлен алфавит (словарь) с выделенным в нем п у с т ы м с и м в о л о м и головка, также называемые соответственно в х о д н ы м и и в ы х о д н ы м и. Каждая лента ограничена слева, и крайняя левая ячейка всегда занята особым символом, г р а н и ч н ы м м а р к е р о м, не входящим в соответствующий алфавит. Кроме того, машина имеет конечное число (в н у т р е н н и х) с о с т о я н и й, среди которых выделены н а ч а л ь н о е и з а к л ю ч и т е л ь н о е. Элементарный акт машины представляет собой либо чтение символа на входной ленте с одновременным сдвигом (входной) головки на одну ячейку вправо и переходом машины в другое состояние, либо одно из следующих действий на рабочей ленте: а) замена обозреваемого символа другим с одновременным сдвигом головки вправо и переходом машины в другое состояние ; б) то же, но со сдвигом влево ; в) то же,

но без сдвига. Когда машина находится в заключительном состоянии, никакой элементарный акт невозможен.

Пусть машина находится в начальном состоянии, на её входной ленте записана непосредственно вслед за граничным маркером цепочка, не содержащая пустого символа, рабочая лента пуста (т.е. все её ячейки, кроме граничной, заняты пустыми символами) и обе головки обозревают граничные маркеры. В этом случае мы скажем, что машина находится в н а ч а л ь н о й с и т у а ц и и. Аналогично, машина находится в з а к л ю ч и т е л ь н о й с и т у а ц и и, если её состояние - заключительное, на входной ленте обозревается первая слева пустая ячейка, а на рабочей ленте картина та же, что в начальной ситуации.

Машина д о п у с к а е т цепочку x, если, начав работу в начальной ситуации с цепочкой x на входной ленте, она может при "удачном" ходе вычисления (в общем случае машина недетерминированная !) оказаться в заключительной ситуации.

Я з ы к о м, д о п у с к а е м ы м м а ш и н о й M, называется множество допускаемых ею цепочек (обозначение: $L(M)$).

Класс языков, допускаемых всевозможными машинами с входной лентой, совпадает с классом $\mathcal{L}(\Gamma)$. Отсюда следует, что $\mathcal{L}(\Gamma)$ есть не что иное, как класс всех рекурсивно перечислимых языков.

Теперь мы будем налагать на рассматриваемые машины некоторые все более сильные ограничения и сравнивать получаемые при этом классы языков с классами, введенными в разделе 1.

1) Машина Тьюринга с входной лентой называется л и н е й н о - о г р а н и ч е н н о й, если существует такое число C, что в процессе "допускания" цепочки x машина не может использовать более чем $C|x|$ рабочих ячеек.

Класс языков, допускаемых такими машинами, совпадает с $\mathcal{L}(\text{НС})$ [37].

2) Машина Тьюринга с входной лентой называется а в-

томатом с магазинной памятью (МП - автоматом), если при каждом её элементарном акте на рабочей ленте либо α) в обозреваемой ячейке непустой символ заменяется пустым, после чего головка сдвигается влево, либо β) головка сдвигается вправо из непустой ячейки (не меняя её содержимого) в пустую, либо γ) пустой символ заменяется непустым без сдвига головки.

Ясно, что акты типов β) и γ) могут происходить только попарно и в указанном порядке. Поэтому мы будем объединять их в один элементарный акт, при котором β') головка сдвигается вправо, после чего в обозреваемой теперь ячейке пустой символ заменяется непустым. Таким образом, мы имеем два двойственных типа элементарных актов α) и β') - стирание и запись. Ясно, что в каждый момент в некотором начальном отрезке рабочей ленты записана какая-то непустая цепочка ξ, остальная часть ленты пуста, а рабочая головка обозревает последний символ цепочки ξ.

Класс языков, допускаемых МП- автоматами, совпадает с $\mathcal{L}$(КС) [23, 30].

3) Назовем МП - автомат одноходовым, если при допускании любой цепочки направление движения его рабочей головки меняется не более одного раза (таким образом, "успешное" вычисление разбивается на два этапа: на первом рабочие символы могут только записываться, на втором - только стираться.)

Класс языков, допускаемых одноходовыми МП- автоматами, совпадает с $\mathcal{L}(\Lambda)$.

4) Мы снабдили машину входной лентой исключительно из соображений удобства. Однако затем, определяя различные типы машин, мы налагали ограничения не на входную, а на рабочую ленту. (Это можно сравнить со следующей воображаемой ситуацией из другой области: к некоторой работающей организации ради удобства присоединяется административное подразделение, которое со временем становится фактически главным, тогда как функции работающей части сокращаются.) Что-

бы довести дело до логического конца, остается наложить на рабочую ленту наиболее сильное из всех мыслимых ограничений - потребовать, чтобы на ней ничего не происходило, или, что то же самое, чтобы её не было совсем. В этом случае мы получим к о н е ч н ы й а в т о м а т (который, очевидно, может рассматриваться как частный случай одноходового МП- автомата).

Класс языков, допускаемых конечными автоматами, совпадает с $\mathcal{L}(A)$ [30]. Существует также другая характеристика этого класса: он совпадает с классом так называемых регулярных событий в смысле Клини [22]. Имеются и другие характеристики этого класса [8].

Назовем машину Тьюринга со входной лентой д е т е р м и н и р о в а н н о й, если она удовлетворяет следующим двум условиям: (i) в каждом состоянии может выполняться элементарный акт только одного из типов а), б), в) (см. определение в начале настоящего раздела); (i i) каждый элементарный акт однозначно определяется состоянием машины и символом, обозреваемым функционирующей при данном акте головкой.

Таким образом, процесс работы детерминированной машины полностью предопределяется входной цепочкой.

Класс языков, допускаемых всевозможными детерминированными машинами Тьюринга со входной лентой (линейно-ограниченными машинами Тьюринга со входной лентой, МП - автоматами, одноходовыми МП- автоматами, конечными автоматами) мы будем обозначать через $\mathcal{T}_{\mathcal{D}}$(М) ($\mathcal{T}_{\mathcal{D}}$(ЛОМ), $\mathcal{T}_{\mathcal{D}}$(МПА), $\mathcal{T}_{\mathcal{D}}$(ОМПА), $\mathcal{T}_{\mathcal{D}}$(КА)).

Имеют место следующие соотношения [8, 39, 34, 10]:

$$\mathcal{T}_{\mathcal{D}}(М) = \mathcal{L}(Г);$$

$$\mathcal{L}(КС) \subsetneq \mathcal{T}_{\mathcal{D}}(ЛОМ) \subseteq \mathcal{L}(НС);$$

$$\mathcal{T}_{\mathcal{D}}(КА) = \mathcal{L}(А) \subsetneq \mathcal{T}_{\mathcal{D}}(ОМПА) \subsetneq$$

$$\subsetneq \mathcal{T}_{\mathcal{D}}(МПА) \cap \mathcal{L}(Л) \overset{\mathcal{T}_{\mathcal{D}}(МПА)}{\subsetneq} \mathcal{T}_{\mathcal{N}}(МПА) \cup \mathcal{L}(Л) \subsetneq \mathcal{L}(КС)$$

$$\subsetneq \mathcal{L}(Л).$$

примеры: $\{a^n b^n c^n / n = 1,2,\dots\} \in \mathcal{T}_{\mathcal{D}}(\text{ЛОМ}) - \mathcal{L}(\text{КС})$;

$\{xc\hat{x} / x \in \{a,b\}^+\} \in \mathcal{T}_{\mathcal{D}}(\text{ОМПА}) - \mathcal{L}(\text{А})$;

$\{a^n b^n e / n = 1,2,\dots\} \cup \{a^k b^l c^p d^p / k,p,l = 1,2,\dots\} \in \mathcal{T}_{\mathcal{D}}(\text{МПА}) \cap \mathcal{L}(\text{Л}) - \mathcal{T}_{\mathcal{D}}(\text{ОМПА})$;

$\{a^n b^n c^p d^p / n,p = 1,2\dots\} \in \mathcal{T}_{\mathcal{D}}(\text{МПА}) - \mathcal{L}(\text{Л})$;

$\{x\hat{x} / x \in \{a,b\}^+\} \in \mathcal{L}(\text{Л}) - \mathcal{T}_{\mathcal{D}}(\text{МПА})$;

$\{xx\hat{x}y\hat{y} / x,y \in \{a,b\}^+\} \in \mathcal{L}(\text{КС}) - [\mathcal{T}_{\mathcal{D}}(\text{МПА}) \cup \mathcal{L}(\text{Л})]$.

Вопрос о том, совпадают ли классы $\mathcal{T}_{\mathcal{D}}$ (ЛОМ) и $\mathcal{L}$(НС), остается открытым.

3. СЛОЖНОСТЬ ВЫВОДА И РАСПОЗНАВАНИЯ ЯЗЫКОВ

Сопоставим произвольной грамматике $\Gamma = (V, V_1, \mathcal{Y}, R)$ функцию $f_\Gamma(\mathcal{D}, i)$, определенную на всевозможных парах $\langle \mathcal{D}, i \rangle$, где $\mathcal{D} = (\mathcal{Y}_0, \mathcal{Y}_1, \dots, \mathcal{Y}_n)$ – произвольный вывод в Γ и $i = 0,1,\dots,n$, и принимающую в качестве значений натуральные числа. Для произвольной цепочки $x \in L(\Gamma)$ и произвольного вывода $\mathcal{D} = (\mathcal{Y}_0, \mathcal{Y}_1, \dots, \mathcal{Y}_n)$ цепочки x из $\mathcal{Y}$ положим $\tilde{f}_\Gamma(\mathcal{D}, x) = \max\limits_{0 \le i \le n} f_\Gamma(\mathcal{D}, i)$. Далее, для любой цепочки $x \in L(\Gamma)$ положим $\tilde{F}_\Gamma(x) = \min \tilde{f}_\Gamma(\mathcal{D}, x)$, где минимум берется по всем выводам цепочки x из $\mathcal{Y}$ в Γ. Наконец, положим $F_\Gamma(n) = \max\limits_{x \in L(\Gamma), |x| \le n} \tilde{F}_\Gamma(x)$; $F_\Gamma(n)$ не определена для тех n, для которых не существует таких $x \in L(\Gamma)$, что $|x| \le n$.

Если выбрать функцию $f_\Gamma(\mathcal{D}, i)$ так, чтобы она в каком-то смысле характеризовала "сложность" цепочки $\mathcal{Y}_i$ по отношению к её "месту" (номеру) в выводе $\mathcal{D}$, то соответствующая функция F_Γ может рассматриваться как некоторая

характеристика сложности грамматики Γ, точнее – работы этой грамматики (т.е. выводов в ней). Такие функции F_Γ мы будем называть сигнализирующими функциями грамматики Γ.

Мы будем рассматривать следующие типы сигнализирующих функций.

1. Временная сигнализирующая (обозначение: $T_\Gamma(n)$), соответствующая случаю $f_\Gamma(\mathfrak{D}, i) = i$.

2. Емкостная сигнализирующая $S_\Gamma(n)$, получаемая при $f_\Gamma(\mathfrak{D}, i) = |\varphi_i|$.

3. Сигнализирующая левой глубины $Y'_\Gamma(n)$, возникающая, когда $f_\Gamma(\mathfrak{D}, i)$ – увеличенное на единицу расстояние от самого левого вхождения в цепочку φ_i вспомогательного символа до её правого конца; точнее: если $\varphi_i = x A \eta$, где $x \in V^*$ и $A \in V_1$, то $f_\Gamma(\mathfrak{D}, i) = |\eta| + 1$; если же $\varphi_i \in V^*$, то $f_\Gamma(\mathfrak{D}, i) = 0$.

4. Сигнализирующая правой глубины $Y''_\Gamma(n)$, определяемая симметрично $Y'_\Gamma(n)$.

5. Сигнализирующая разброса $P_\Gamma(n)$: $f_\Gamma(\mathfrak{D}, i)$ есть длина наименьшей подцепочки φ_i, содержащей все вхождения вспомогательных символов.

6. Сигнализирующая активной емкости $I_\Gamma(n)$, которая получается, если $f_\Gamma(\mathfrak{D}, i) = |\varphi_i|_1$ ($|\varphi_i|_1$ – число вхождений вспомогательных символов в φ_i).

Тривиальным образом получаются следующие соотношения:

$$T_\Gamma(n) \leqslant K^{S_\Gamma(n)+1};$$

где K – общее число основных и вспомогательных символов грамматики Γ;

$$I_\Gamma(n) \leqslant P_\Gamma(n) \leqslant \begin{matrix} Y'_\Gamma(n) \\ Y''_\Gamma(n) \end{matrix} \leqslant S_\Gamma(n) \leqslant (d+1) \cdot T_\Gamma(n),$$

где $d = \max\limits_{\varphi \to \psi \in R} \{|\psi| - |\varphi|\}$.

Введенные здесь понятия позволяют изучать связь между "порождающей силой" грамматик и сложностью их работы, характеризуемой с помощью сигнализирующих функций. Отметим наибо-

лее интересные из относящихся сюда результатов. Предварительно введем следующие обозначения.

Обозначим через $\mathcal{L}^{p}_{c\cdot n}(\Gamma)$, соответственно $\mathcal{L}^{p}_{c\cdot n}$ (НС), $\mathcal{L}^{p}_{c\cdot n}$ (КС), класс языков, порождаемых произвольными грамматиками (НС – грамматиками, КС – грамматиками) с сигнализирующей разброса, не превосходящей некоторой линейной функции. Далее, через $\mathcal{L}^{p}_{n}(\Gamma)$, $\mathcal{L}^{p}_{n}$ (НС), ... будем обозначать классы языков, порождаемых произвольными грамматиками, НС – грамматиками и т.д. с сигнализирующей разброса $P_r(n) = n \cdot \mathcal{L}^{p}_{c}(\Gamma)$, $\mathcal{L}^{p}_{c}$ (НС), ... будет означать класс языков, порождаемых произвольными грамматиками с сигнализирующей разброса, не превосходящей некоторой константы; наконец, $\mathcal{L}^{p}_{1}(\Gamma)$, $\mathcal{L}^{p}_{1}$ (НС), ... – класс языков, порождаемых грамматиками соответствующих классов с сигнализирующей разброса $P_r(n) = 1$.

Совершенно аналогично можно определить классы $\mathcal{L}^{V^{n}}_{c\cdot n}(\Gamma)$, $\mathcal{L}^{V^{n}}_{c\cdot n}$ (НС), ..., $\mathcal{L}^{V^{n}}_{n}(\Gamma)$, ..., $\mathcal{L}^{V^{n}}_{1}(\Gamma)$, ..., $\mathcal{L}^{V^{n}}_{c\cdot n}(\Gamma)$, ..., $\mathcal{L}^{T}_{c\cdot n}(\Gamma)$, ..., $\mathcal{L}^{S}_{c\cdot n}(\Gamma)$, ..., $\mathcal{L}_{c\cdot n}(\Gamma)$, ...

Имеют место следующие соотношения:

1. а) $\mathcal{L}(\text{НС}) \supseteq \begin{matrix} \mathcal{L}^{T}_{c\cdot n}(\Gamma) \\ \cup \\ \mathcal{L}^{T}_{c\cdot n}(\text{НС}) \end{matrix} \supseteq \mathcal{L}(\text{КС}) = \mathcal{L}^{T}_{c\cdot n}(\text{КС})$.

[Примеры, обеспечивающие строгие включения:
$\{xcx/x \in \{a,b\}^{+}\} \in \mathcal{L}^{p}(\text{НС}) - \mathcal{L}^{T}_{c\cdot n}(\text{НС})$;
$\{ba^{2^{n}}b/n = 1,2,\ldots\} \in \mathcal{L}^{T}_{c\cdot n}(\text{НС}) - \mathcal{L}(\text{КС})$ [6].]

Неизвестно, являются ли включения $\mathcal{L}(\text{НС}) \supseteq \mathcal{L}^{V^{n}}_{c\cdot n}(\Gamma)$ и $\mathcal{L}^{T}_{c\cdot n}(\Gamma) \supseteq \mathcal{L}^{T}_{n}(\text{НС})$ собственными.

б) $\mathcal{L}^{T}_{n}(\text{КС}) = \mathcal{L}(\text{КС})$.

Последнее равенство непосредственно вытекает хотя бы из так называемой теоремы о стандартной форме [35]: по всякой КС-грамматике Γ может быть построена эквивалентная ей КС-грамматика, каждое правило которой имеет либо вид $A \to aBC$, либо вид $A \to aB$, либо вид $A \to a$, где A, B, C – вспомогательные символы, a – основной символ.

2. а) $\mathcal{L}^{s}_{с.п}(Г) = \mathcal{L}^{s}_{с.п}(НС) = \mathcal{L}^{s}_{п}(Г) = \mathcal{L}^{s}_{п}(НС) = \mathcal{L}(НС)$ [5];

б) $\mathcal{L}^{s}_{с.п}(КС) = \mathcal{L}^{s}_{п}(КС) = \mathcal{L}(КС)$ тривиальным образом.

3 а) $\mathcal{L}(Г) \supsetneq \mathcal{L}^{У\Pi}_{с.п}(Г) = \mathcal{L}^{У\Pi}_{п}(Г) \supseteq \mathcal{L}(НС)$;

б) $\mathcal{L}^{У\Pi}_{с}(Г) = \mathcal{L}^{У\Pi}_{с}(НС) = \mathcal{L}^{У\Pi}_{с}(КС) = \mathcal{L}^{У\Pi}_{1}(Г) =$

$= \mathcal{L}^{У\Pi}_{1}(НС) = \mathcal{L}^{У\Pi}_{1}(КС) = \mathcal{L}(А)$

Авторам неизвестно, является ли включение $\mathcal{L}^{У\Pi}_{п}(Г) \supseteq \mathcal{L}(НС)$ собственным.

4 Для $У^{\Pi}$ – аналогично.

5. а) $\mathcal{L}^{Р}_{с.п}(Г) = \mathcal{L}^{Р}_{п}(Г) = \mathcal{L}^{Р}_{с}(Г) = \mathcal{L}^{Р}_{1}(Г) = \mathcal{L}(Г)$;

б) $\mathcal{L}^{Р}_{с}(НС) = \mathcal{L}^{Р}_{с}(КС) = \mathcal{L}^{Р}_{1}(НС) = \mathcal{L}^{Р}_{1}(КС) = \mathcal{L}(Л)$.;

6. а) $\mathcal{L}^{\bar{1}}_{с.п}(Г) = \mathcal{L}^{\bar{1}}_{п}(Г) = \mathcal{L}^{\bar{1}}_{с}(Г) = \mathcal{L}^{\bar{1}}_{1}(Г) = \mathcal{L}(Г)$;

б) $\mathcal{L}(КС) \supsetneq \mathcal{L}^{\bar{1}}_{с}(НС) = \mathcal{L}^{\bar{1}}_{с}(КС) \supsetneq \mathcal{L}^{\bar{1}}_{1}(НС) = \mathcal{L}^{\bar{1}}_{1}(КС) = \mathcal{L}(Л)$ [12]

[Примеры: множество всех бинарных скобочных последовательностей принадлежит $\mathcal{L}(КС) - \mathcal{L}^{\bar{1}}(КС)$;

$\{a^n b^n c^p d^p \mid n, p = 1, 2, \ldots\} \in \mathcal{L}^{\bar{1}}_{с}(КС) - \mathcal{L}(Л)$.]

Замечание. Мы не рассматриваем соотношений, относящихся к сигнализирующим для языков классов $\mathcal{L}(Л)$ и $\mathcal{L}(КА)$, ввиду их полной очевидности.

Перейдем к рассмотрению вопроса о сложности распознавания языков различными детерминированными автоматами (машинами). Мы говорим, что детерминированная машина Тьюринга р а с п о з н а е т язык L, если она допускает L и каждое её вычисление заканчивается.

Для этого прежде всего нужно определить для машин сигнализирующие функции аналогично тому, как это было сделано для грамматик.

Пусть M – детерминированная машина Тьюринга со входной лентой и $x \in L(M)$. Обозначим через $\tilde{T}_M(x)$ число элементарных актов, выполняемых машиной M при "допускании" x, т.е. при переходе от начальной ситуации со входной цепочкой x к заключительной. Положим $T_M(n) =$

$= \max_{x \in L(M);\, |x| \le n} \tilde{T}_M(x)$; $T_M(n)$ не определена для тех n, для которых не существует таких $x \in L(M)$, что $|x| \le n$. Функцию $T_M(n)$ назовем временной сигнализирующей машины M.

Если в приведенной только что формулировке заменить число элементарных актов числом ячеек рабочей ленты, используемых машиной M при распознавании цепочки x, мы получим определение емкостной сигнализирующей машины M.

Совершенно аналогично могут быть определены временные сигнализирующие для всех других типов машин, которые будут рассматриваться дальше – с неограниченной подвижностью входной головки, многоголовочных и т.п. Что касается емкостной сигнализирующей, то для всех типов машин она определяется так же, как выше – с учетом только рабочей ленты.

Из довольно многочисленных известных соотношений, относящихся к оценкам сигнализирующих функций машин, распознающих языки различных классов, мы приведем лишь некоторые.

В дальнейшем будут рассматриваться только детерминированные машины; слово "детерминированная" будет опускаться.

I. Рассмотрим сначала распознавание на машинах Тьюринга с входной лентой, как они были определены выше. В этом классе машин

а) всякий язык, распознаваемый с емкостной сигнализирующей, которая мажорируется константой, является А-языком; не существует языков, распознаваемых с емкостными сигнализирующими, по порядку бо́льшими константы, но меньшими $\log n$; язык $\{a^n b^n / n = 1, 2, \ldots\}$ распознается с емкостной сигнализирующей $\log n$ · [38,42];

б) существует бесконечная последовательность функций $l_1(n) \cdot l_2(n) \ldots$, заключенных по порядку между $\log n$

65106

БИБЛИОТЕКА ИНСТИТУТА РУССКОГО ЯЗЫКА АН СССР

и n и таких, что для каждого $j = 1,2,\ldots$: (i) $f_{j+1}(n)$ по порядку больше $f_j(n)$; (ii) существует КС-язык, распознаваемый некоторой машиной с емкостной сигнализирующей $f_j(n)$ и не распознаваемый никакой машиной с емкостной сигнализирующей, по порядку меньшей $f_j(n)$ [38];

в) язык $\{x c \hat{x} / x \in \{a,b\}^+\}$ распознается машиной с емкостной сигнализирующей n и не распознается никакой машиной с емкостной сигнализирующей, по порядку меньшей n [38].

2. Если лишить машину входной ленты (т.е. перейти к обычной одноленточной машине Тьюринга), то

а) всякий КС-язык можно распознать с временной сигнализирующей, по порядку не превосходящей n^4 [11];

б) всякий линейный язык можно распознать с временной сигнализирующей, по порядку не превосходящей n^2 [36]; язык $\{x c \hat{x} / x \in \{a,b\}^+\}$ нельзя распознать с временной сигнализирующей, по порядку меньшей n^2 [2];

в) существует бесконечная последовательность функций $f_1(n), f_2(n), \ldots$, заключенных по порядку между $n \log n$ и n^2 и таких, что для каждого $j = 1,2,\ldots$: (i) $f_{j+1}(n)$ по порядку больше $f_j(n)$; существует КС-язык, распознаваемый некоторой машиной с временной сигнализирующей $f_j(n)$ и не распознаваемый никакой машиной с временной сигнализирующей, по порядку меньшей $f_j(n)$;

г) язык $\{a^n b^n / n = 1,2,\ldots\}$ распознается с временной сигнализирующей $n \log n$;

д) всякий язык, распознаваемый с временной сигнализирующей, по порядку меньшей n, является А-языком; не существует языков, распознаваемых с временными сигнализирующими, заключенными по порядку между n и $n \log n$ [20].

Если снабдить одноленточную машину тремя головками, то

е) всякий КС-язык можно распознать с временной сигнализирующей, по порядку не превосходящей n^3 [43].

3. Ещё один класс образуют машины Тьюринга с одной

входной и одной рабочей лентой, которые на рабочей ленте функционируют так же, как машины первого из рассматривавшихся нами типов, в то время как входная головка может передвигаться по своей ленте без ограничений, не меняя, однако, её содержимого. Для этих машин:

а) всякий КС-язык можно распознавать с емкостной сигнализирующей, по порядку не превосходящей $(\log n)^2$ [38];

б) всякий язык, распознаваемый с емкостной сигнализирующей, по порядку меньшей $\log \log n$, является А-языком; НС-язык $L=\{cb_1^{\alpha}cb_2^{\alpha}\dots cb_{2i-1}^{\alpha}cb_{2i}^{\alpha} \mid i=1,2,\dots\}$, где b_i^{α} - двоичная запись числа i, в которой перед каждым разрядом стоит двоичная запись номера этого разряда в алфавите $\{\bar{0},\bar{1}\}$ /например, $b_9^{\alpha}=\bar{1}\bar{0}\bar{0}\bar{1}\bar{1}\bar{1}\bar{0}\bar{1}\bar{0}\bar{0}\bar{1}\bar{0}$/, и C - символ, отличный от $0, 1, \bar{0}$ и $\bar{1}$, распознается с емкостной сигнализирующей $\log \log n$ [38, 42]; / L - не КС-язык, но дополнение к L - КС-язык/.

в) существует бесконечная последовательность функций $l_1(n)$, $l_2(n)$, ..., заключенных по порядку между $\log \log n$ и $\log n$ и таких, что для каждого $j=1,2,\dots$: (i) $l_{j+1}(n)$ по порядку больше $l_j(n)$; (ii) существует КС-язык, распознаваемый некоторой машиной с емкостной сигнализирующей $l_j(n)$ и не распознаваемый никакой машиной с емкостной сигнализирующей, по порядку меньшей $l_j(n)$ [38].

Мы не будем останавливаться на результатах, относящихся к другим сигнализирующим, а также к другим концепциям машин и автоматов. Некоторые из этих результатов можно найти в [21].

4. НЕКОТОРЫЕ ПРОМЕЖУТОЧНЫЕ КЛАССЫ ГРАММАТИК И ЯЗЫКОВ

В ряде прикладных задач приходится сталкиваться с ситуацией, когда КС- грамматики оказываются недостаточно удобными, и в то же время использовать произвольные НС-

грамматики заведомо не нужно. Поэтому возникает вопрос об исследовании различных промежуточных классов НС- грамматик - в первую очередь с целью выяснения объема соответствующих классов языков.

Сформулируем некоторые результаты этого типа.

1. Класс языков, порождаемых л е в о - к о н т е к с т н ы м и НС - г р а м м а т и к а м и - т.е. НС- грамматиками, все правила которых имеют вид $\varphi A \to \varphi \omega$ (A - вспомогательный символ) - шире класса КС- языков; именно, он содержит языки $\{a^n b^n c^j \mid n = 1, 2, \ldots; 1 \le j \le n\}$ [35] и $\{a^n b^n a^n \mid n = 1, 2, \ldots\}$ [16].

Замечание. В недавно вышедшей заметке [34] анонсировано утверждение о совпадении этого класса с классом всех НС- языков.

2. Класс языков, порождаемых такими лево-контекстными НС- грамматиками, все правила которых имеют вид $\varphi A \to \varphi \omega' a$ (A - вспомогательный символ, a - основной символ) совпадает с классом КС - языков [17].

Аналогично могут быть определены п р а в о - к о н т е к с т н ы е НС - г р а м м а т и к и, для которых, разумеется, имеют место такие же факты.

3. Класс языков, порождаемых НС- грамматиками, все правила которых имеют вид $xAy \to x\omega y$, где x и y - цепочки в основном словаре, совпадает с классом КС- языков [26].

4. Класс языков, порождаемых НС- грамматиками, все правила которых имеют вид $xA\psi \to x\omega\psi$, где x цепочка в основном словаре и $|x| \ge |\psi|$, совпадает с классом КС- языков [26].

Отметим ещё следующий результат несколько иного харак

дает КС-язык [31]. (Отсюда в силу теоремы о стандартной форме (см. раздел 3) следует, что класс языков, порождаемых такими грамматиками, совпадает с классом КС- языков).

Кроме ограничений на вид правил грамматики, возможны также ограничения на способ их применения. Строго говоря, подобные ограничения приводят к исчислениям, не являющимся порождающими грамматиками в смысле определения, сформулированного в разделе I; однако они достаточно близки к грамматикам, и мы упомянем здесь два результата, относящихся к таким исчислениям.

1. Назовем вывод $(\omega_0, \omega_1, \dots, \omega_n)$ в произвольной грамматике р е г у л я р н ы м, если для каждого $i = 0, \dots, n-1$ цепочки ω_i и ω_{i+1} представляются в виде $\omega_i = \omega_i' \varphi$, $\omega_{i+1} = \omega_i' \psi$, где $\varphi \to \psi$ - правило грамматики (иначе - правила применяются только на правых концах цепочек). Оказывается, что для любой грамматики множество цепочек в её основном словаре, выводимых из её начального символа с помощью регулярных выводов, является A - языком [27].

2. Пусть Γ - НС -грамматика, каждое правило которой имеет вид $x A \psi \to x \omega \psi$, где x - цепочка в основном словаре. Назовем вывод в такой грамматике л е в ы м, если на каждом его шаге заменяется самое левое вхождение вспомогательного символа в соответствующую цепочку.

Множество цепочек основных символов, выводимых в грамматике указанного типа из её начального символа с помощью левых выводов, является КС- языком [15, 39, 7].

5. НАДСТРОЙКИ И УПРАВЛЕНИЕ ВЫВОДОМ

Пусть $\Gamma = (V, V_1, \mathfrak{I}, R)$ - некоторая грамматика. Сопоставим каждому правилу из R символ τ (разным правилам

отвечают разные символы); множество таких символов обозначим R. Каждому выводу в грамматике Γ сопоставим его характеристику – цепочку $r_1 \dots r_k$, где r_i $(1 \le i \le k)$ – символ, соответствующий правилу, применяемому на i-ом шаге вывода. Множество характеристик всевозможных полных выводов в грамматике Γ назовем надстройкой этой грамматики.

Классы языков, являющихся надстройкой для НС-($КС$-, A-) грамматик будут обозначаться $\mathcal{P}(НС)$ ($\mathcal{P}(КС)$, $\mathcal{P}(A)$).

институт русского языка им. в. в. виноградова

Приведем некоторые результаты, устанавливающие связь между типом грамматики и типом её надстройки.

1. Надстройка любой грамматики является НС-языком [18].

2. $\mathcal{P}(НС) \supsetneq \mathcal{P}(КС) \supsetneq \mathcal{P}(A)$

3. Надстройка любой A-грамматики является A-языком [18].

(Соответствующие примеры: $\{a^n b c^n / n \ge 1\} \in \mathcal{P}(НС) - \mathcal{P}(КС)$; класс $\mathcal{P}(КС) - \mathcal{P}(A)$ содержит язык $\{x / x \in \{a, b\}^+; |x|_a + 1 = |x|_b$; для всякого начального отрезка y цепочки $x\ |y|_a \ge |y|_b\}$, где $|z|_a$ ($|z|_b$) – число вхождений символа a (b) в цепочку z. Первый пример языка, принадлежащего классу $\mathcal{P}(КС) - \mathcal{P}(A)$, построен в [19а]).

www.ruslang.ru

4. Будем рассматривать грамматики, у которых левые части всех правил являются непустыми цепочками вспомогательных символов. Среди таких грамматик автоматные надстройки имеют те и только те, которые имеют равномерно ограниченную активную емкость [28, 12]. (Мы говорим, что грамматика имеет равномерно ограниченную активную емкость, если существует такая константа K, что ни в одном полном выводе грамматики ни одна промежуточная цепочка не содержит более K вхождений вспомогательных

символов. Для всякой грамматики с емкостью, равномерно ограниченной числом K, функция $I_\Gamma(n)$ будет, очевидно, ограничена тем же числом. Обратное неверно; более того, существует КС - язык $L \in \mathcal{L}_c^{I}(Kc)$, не порождаемый никакой грамматикой рассматриваемого в этом пункте вида C равномерно ограниченной активной емкостью. Можно, например, взять $L = \{a^n b^{n^2} / n = 1,2,\ldots\}^+$ [12].)

5. Существуют языки, порождаемые грамматиками с КС - надстройками, но не порождаемые никакими грамматиками введенного в предыдущем пункте типа с A - надстройками (примером может служить язык L предыдущего пункта) [12].

Остается открытым вопрос о существовании языков, не порождаемых никакими грамматиками с КС- надстройками.

Несколько иной подход к изучению структуры множеств выводов в грамматиках предложен в [40].

В прикладных задачах, как правило, приходится иметь дело не с произвольными выводами в грамматиках, а с выводами, удовлетворяющими определенным условиям. Поэтому представляет интерес изучение различных ограничений на выводы и получающихся при этих ограничениях языков.

1. Одним из естественных способов "управления выводом" в грамматике $\Gamma = (V, V_1, \mathcal{Y}, R)$ является задание множества характеристик "допустимых выводов" с помощью некоторой грамматики Γ' с основным словарем $\bar{R}$. Таким образом, мы приходим к понятию обобщенной грамматики [18]. Именно, о б о б щ е н н а я г р а м м а т и к а - это пара грамматик (Γ, Γ') такая, что язык $L(\Gamma')$ является подмножеством множества всех характеристик выводов в Γ. Я з ы к о м, п о р о ж д а е м ы м о б о б щ е н н о й г р а м м а т и к о й (Γ, Γ') называется множество $\{x / x \in L(\Gamma);$ существует вывод x из начального символа Γ, характеристика которого принадлежит $L(\Gamma')\}$:

Мы будем пользоваться обозначениями $\mathcal{L}(\Gamma,\Gamma)$, $\mathcal{L}(\Gamma,НС),\ldots,\mathcal{L}(A,A)$, имеющими следующий смысл: "класс языков, порождаемых обобщенными грамматиками (Γ,Γ'), где Γ - произвольная грамматика и Γ' - произвольная грамматика, соответственно Γ - произвольная и Γ' — НС- грамматика,..., Γ-A - грамматика и Γ'-A - грамматика."

Имеют место следующие соотношения [18, 19]:

$\mathcal{L}(НС,НС)=\mathcal{L}(КС,НС)=\mathcal{L}(\Gamma)$;

$\mathcal{L}(НС,A)=\mathcal{L}(A,НС)=\mathcal{L}(НС)$;

$\mathcal{L}(A,КС)=\mathcal{L}(КС)$;

$\mathcal{L}(A,A)=\mathcal{L}(A)$.

Замечания. 1) Любой класс $\mathcal{L}(\alpha,\beta)$ где α или β есть Γ, тривиальным образом совпадает с $\mathcal{L}(\Gamma)$. 2) О языках классов $\mathcal{L}(НС, КС)$ и $\mathcal{L}(КС, КС)$ известно только, что они являются рекурсивными множествами достаточно простой природы (в смысле сложности их распознавания). 3) Класс $\mathcal{L}(КС, A)$ будет рассмотрен в следующем пункте.

2. Одним из наиболее простых и естественных ограничений на вывод является объединение правил в группы так, чтобы правила из одной группы применялись вместе. В наиболее интересном случае, когда правила являются контекстно-свободными, это ограничение приводит к понятию матричной грамматики (впервые введенному в [25]). М а т р и ч н а я г р а м м а т и к а отличается от обычной КС-грамматики только тем, что её правила собраны в конечные последовательности - м а т р и ц ы, причем одно и то же правило может принадлежать нескольким разным матрицам и встречаться в одной матрице несколько раз. (В частности, возможны матрицы, состоящие из одного правила). Н е п о с р е д с т в е н н а я в ы в о д и м о с т ь в матричной грамматике может быть определена двумя различными способами: а) цепочка ω' непосредственно выводима из цепоч-

ки ω, если ω' получается из ω последовательным применением всех правил некоторой матрицы в том порядке, в каком они в ней записаны; б) ω' непосредственно выводима из ω, если ω' получается из ω одновременным применением всех правил некоторой матрицы (причем, если какое-то правило входит в матрицу K раз, оно должно применяться к K различным вхождениям своей левой части). Сформулированные определения непосредственной выводимости естественным образом приводят к двум разным определениям языка, порождаемого матричной грамматикой. Оказывается, однако, что эти определения равносильны. Класс языков, порождаемых матричными грамматиками, обозначим через $\mathcal{L}(M)$.

Установлено, что:

а) $\mathcal{L}(M) = \mathcal{L}(KC, A)$ [1, 18, 19].

б) $\mathcal{L}(KC) \subsetneq \mathcal{L}(M)$.

(Примеры: $\{a^n b^n c^n / n = 1,2,\dots\}$, $\{xcx^R / x \in \{a,b\}^+\} \in \mathcal{L}(M) - \mathcal{L}(KC)$).

в) $\mathcal{L}(M) \subsetneq \mathcal{L}(HC)$ [41].

3. Ещё один очень важный для приложений способ ограничения возможностей вывода состоит в упорядочении множества правил. Именно, пусть на множестве правил грамматики введено некоторое отношение частичного порядка, и правило разрешается применять к цепочке лишь в том случае, когда к ней не применимо ни одно из правил, следующих за данным. Полученный так объект называется грамматикой с порядком [29]. Язык, порождаемый грамматикой с порядком, определяется очевидным образом.

Можно показать, что класс языков, порождаемых КС- грамматиками с порядком, содержит $\mathcal{L}(KC)$ и строго содержится в $\mathcal{L}(HC)$.

4. Если каждому правилу τ некоторой грамматики сопоставить А-язык $\ell(\tau)$ и разрешить применять это правило только к цепочкам, принадлежащим $\ell(\tau)$, получится исчисление, называемое грамматикой с авто-

матными ограничениями [29]. Класс языков, порождаемых НС- грамматиками с автоматными ограничениями, совпадает с $\mathcal{L}(HC)$ [29]. Можно показать, что класс языков, порождаемых КС- грамматиками с автоматными ограничениями, также совпадает с $\mathcal{L}(\text{НС})$.

ЛИТЕРАТУРА

1. М.В.Арапов, В.Б.Борщев, Ю.А.Шрейдер, Язык, грамматика, семантика, в сб. "Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации", т.2, М., 1967, 5-15.

2. Я.М.Барздинь, Сложность распознавания симметрии на машинах Тьюринга, "Пробл. кибернетики", 15, 1965, 245-248.

3. М.И.Белецкий, О различных типах эквивалентности доминационных грамматик, журн. "Кибернетика", № 4, 1968, 91-95.

4. В.Б.Борщев, Окрестностные грамматики, "Научно-техническая информация", сер. 2, № 11, 1967, 39-41.

5. А.В.Гладкий, Грамматики с линейной памятью, Алгебра и логика. Сем., т.2, № 5, 1963, 43-55.

6. А.В.Гладкий, О сложности вывода в грамматиках непосредственно составляющих, Алгебра и логика. Сем., т.3, № 5-6, 1964, 29-44.

7. А.В.Гладкий, Прямое доказательство теоремы Мэтьюза, Алгебра и логика. Сем., т.4, № 4, 1965, 65-70.

8. А.В.Гладкий, Лекции по математической лингвистике для студентов НГУ, Новосибирск, 1966.

9. А.В.Гладкий, Именные окрестностные грамматики и НС-грамматики, "Научно-техническая информация", сер. 2, № 5, 1969, 22-24.

10. А.Я.Диковский, О соотношении между классом всех контекстно-свободных языков и классом детерминированных контекстно-свободных языков, Алгебра и логика. Сем., т.8, № I, 1969, 44-61.

11. А.Я.Диковский, Замечание о распознавании КС-языков на машинах Тьюринга, Всесоюзная конференция по проблемам теоретической кибернетики. Тезисы докладов, Новосибирск, 1969, II.

12. А.Я.Диковский, Языки с ограниченной активной емкостью, в печати.

13. А.П.Ершов, Инструкция по оформлению докладов, представленных на 2-ю Всесоюзную конференцию по программированию, Новосибирск, 1969.

14. М.В.Ломковская, Об именных окрестностных и бесконтекстных языках, "Научно-техническая информация", сер.2, № 6, 1969, 21-24.

15. Дж.Мэтьюз, Разрывность и асимметрия в грамматиках непосредственных составляющих, сб. "Математическая лингвистика" М., "Мир", 1964, 150-159.

16. Л.Г.Самойленко, Об одном классе грамматик непосредственно составляющих, журн. "Кибернетика", № 2, 1968, 102-103.

17. Л.Г.Самойленко, Некоторые вопросы теории грамматик непосредственных составляющих. Автореферат, Киев, 1968.

18. Э.Д.Стоцкий, О некоторых ограничениях на способ вывода в грамматиках непосредственно составляющих, "Научно-техническая информация", сер.2, № 7, 1967, 35-38.

19. Э.Д.Стоцкий, Об одном обобщении порождающих грамматик, Автореферат, Москва, 1969.

19 а. Э.Д.Стоцкий, Об ограничениях на вывод в грамматиках непосредственных составляющих, Ученые записки Тартуского Гос. Ун-та, 1969, вып. 226, 107-111.

20. Б.А.Трахтенброт, Тьюринговы вычисления с логарифмическим замедлением, Алгебра и логика. Сем., 1964, т.3, 4, 33-48.

21. Б.А.Трахтенброт, Сложность алгоритмов и вычислений, Новосибирск, 1967.

22. Б.А.Трахтенброт, Я.М.Барздинь, Конечные автоматы (поведение и синтез), М., "Наука", 1970.

23. Н.Хомский, Формальные свойства грамматик, Кибернетический сборник, вып.2 (новая серия), М., "Мир", 1966, 121-230.

24. Ю.А.Шрейдер, Окрестностная модель языка, Tartu Ülikooli Toimetised, Уч.зап.Тартуского ун-та, 1969, вып. 226, 129-142.

25. S.ABRAHAM, Some questions of phrase structure grammars I, Computational linguistics, IV, Budapest, 1965, 61-70.

26. R.V.BOOK, On the effect of terminal context on context sensitive grammars, Math. Linguistics and Automatic Translation, Harvard Univ. Computation Lab. Rept. NSF-18, 1967, pp. IV-I - IV-30.

27. J.BÜCHI, Regular canonical systems, Arch. Math. Logik Grundlagenforsch., Band 6, 1964, 91-111.

28. J.FRIANT, Langages ultralinéaires et superlinéaires. Nouvelles caractérisations, Rapport MA-102, CETADOL, Univ. de Montréal (1968).

29. I.FRIŠ, Grammars with Partial Ordering of the Rules, Information and Control, v.12, 1968, 415-425.

30. GINSBURG, S., Mathematical Theory of Context-Free Languages, McGraw-Hill, Inc., New York, 1966.
(Русский перевод: С.ГИНЗБУРГ, Математическая теория контекстно-свободных языков. М., "Мир", 1970).

31. S.GINSBURG and S.A.GREIBACH, Mappings which preserve Context-sensitive languages, Information and Control, v.9, 1966, 563-582.

32. S.GINSBURG, S.A.GREIBACH, Deterministic context-free languages, Information and Control, v.9, 1966, 620-648.

33. S.A.GREIBACH, A simple proof of the standard form theorem for context-free grammars, Math. Linguistics and Automatic Translation, Harvard Univ. Computation Lab. Rept. NSF-18, 1967, pp. 11-1 - 11-4.

34. L.H.HAINERS, Representation theorems for context sensitive languages, Notices of the AMS, vol. 16, N 3 (1969), p.527.

35. I.HAVEL, A note on One-Sided Context-Sensitive Grammars, Kybernetika, číslo 3, ročník 5, 1969, 186-189.

36. T.KASAMI, A note on computing time for recognition of languages generated by linear grammars, Information and Control, v.10, 1967, 209-214.

37. S.Y.KURODA, Classes of languages and linear-bounded automata, Information and Control, v.7, 1964, 207-223.

38. P.M.LEWIS, J.HARTMANIS, R.E.STEARNS, Memory bounds for recognition of context-free and context-sensitive languages, IEEE Conference Record on Switching Cirquit Theory and Logical Design, 1965, 191-202.

39. G.H.MATTEWS, A note on asymmetry of phrase structure grammars, Information and Control, v.7, 1964, 360-365.

40. G.ROSENBERG, Languages de derivations, Bulletin de l'Académie Polonaise des Sciences, Série de Sc. Math., Astr. et Phys., v.15, 1967, 753-758.

41. D.I.ROSENKRANTZ, Programmed grammars and classes of formal languages, Journ. of the Assoc. for Comput. Mach., v.16, 1969, 107-131.

42. R.E.STEARNS, J.HARTMANIS, P.M.LEWIS II, Hierarchies of Memory Limited Computations, IEE Conference Record on Switching Cirquit Theory and Logical Design, 1965, 179-190.

43. D.F.YOUNGER, Recognition and parsing of context-free languages in time n^3, Information and Control, v.10, 1967, 189-208.

Т-10447 от 13.08.70 2 п.л. 100 экз. зак.735

Тип. МГПИИЯ