

Структура контекстно-свободных языков

С.Ю.Соловьев

МГУ имени М.В.Ломоносова, факультет ВМК, Москва, Россия

Поступила в редколлегию 24.03.2011

Аннотация—В работе показывается, что контекстно-свободный язык можно представить семейством π -сетей с помеченными ребрами. При этом оказывается, что достаточно представительные π -сети обладают рядом структурных свойств, позволяющим по конечному фрагменту сети построить правила контекстно-свободной грамматики.

1. ВВЕДЕНИЕ

На рисунке 1 представлена схема дорог между городами А и Б. Любой путь из А в Б всегда пролегает через поселок В, поэтому:

выбор конкретного пути [АБ] распадается на две подзадачи: $[АБ] \rightarrow [АВ]+[ВБ]$
причем путь [АВ] можно пройти двумя способами: $[АВ] \rightarrow 1 \mid 2$
и путь [ВБ] можно также пройти двумя способами: $[ВБ] \rightarrow 3 \mid 4$

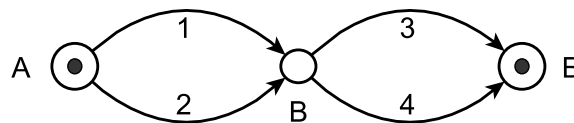


Рис. 1. Двухполюсная сеть как схема дорог

Полученная таким образом контекстно-свободная грамматика

$$\begin{aligned} [АБ] &\rightarrow [АВ]+[ВБ] \\ [АВ] &\rightarrow 1 \mid 2 \\ [ВБ] &\rightarrow 3 \mid 4 \end{aligned}$$

порождает все пути $\{1+3, 1+4, 2+3, 2+4\}$ из А в Б и служит альтернативным представлением исходной схемы дорог. Из примера следует, что между двухполюсными сетями и грамматиками существует определенная связь, остается лишь:

- ▷ уточнить классы сетей и грамматик;
- ▷ установить взаимно однозначный характер связи; и
- ▷ выявить "полезные" свойства сетевого представления.

2. ПОРОЖДАЮЩИЕ КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Контекстно-свободной (КС-) грамматикой [1] называется четверка $G = \langle N, \Sigma, P, S \rangle$, где N – алфавит нетерминальных символов (нетерминалов);

Σ – непересекающийся с N алфавит терминальных символов (терминалов);

P – конечное множество правил вывода вида $A \rightarrow \alpha$, где $A \in N$, α – цепочка символов из $N \cup \Sigma$;

S – выделенный символ из N , именуемый начальным символом.

Дополнительно будем полагать, что действуют следующие соглашения:

A, B, C, D, E – нетерминальные символы; S – начальный символ;

a, b, c, d – терминальные символы;

X, Y, Z – терминальные и нетерминальные символы;

α, β, γ – цепочки символов из $N \cup \Sigma$;

x, y, z – цепочки символов (предложения) из Σ ;

$||N||$ – количество символов в алфавите N ;

запись $A \rightarrow \alpha_1 | \dots | \alpha_n$, означает множество правил $\{ A \rightarrow \alpha_1, \dots, A \rightarrow \alpha_n \}$;

запись $\alpha \Rightarrow_G^+ \beta$ означает, что цепочка β непосредственно выводима из цепочки α в грамматике G , то есть для некоторого правила $A \rightarrow \gamma$ из P имеет место $\alpha = \alpha_1 A \alpha_2$, $\beta = \alpha_1 \gamma \alpha_2$;

$L(G) = \{ x \mid S \Rightarrow_G^+ x \}$ – КС-язык, порождаемый грамматикой G .

Посредством эквивалентных преобразований КС-грамматика может изменяться, сохраняя неизменным порождаемый язык. В дальнейшем будем рассматривать бесконечные КС-языки, порожденные КС-грамматиками, удовлетворяющими следующим пяти требованиям.

Требование 1. В КС-грамматике отсутствуют правила с пустой правой частью.

Требование 2. В КС-грамматике имеется правило $S \rightarrow \#$, причем терминальный символ $\#$ в других правилах не встречается.

Требование 3. В КС-грамматике отсутствуют бесполезные¹ правила вывода и бесполезные [1] символы.

Требование 4. В КС-грамматике отсутствуют цепные² правила вывода.

Требование 5. В КС-грамматике отсутствуют простые³ нетерминалы.

В [2] показано, что для исследования порождающих способностей КС-грамматик требования 1-5 не являются существенными. Покажем

что по известной КС-грамматике, удовлетворяющей требованиям 1-5, можно построить семейство π -сетей; ▷▷ разделы 3-7

что к перечисленным пяти требованиям можно, не нарушая общности, присоединить еще одно требование, гарантирующее обратимость перехода от грамматик к π -сетям; ▷▷ раздел 8

что структурные свойства достаточно представительной π -сети позволяют построить правила КС-грамматики, породившей эту сеть. ▷▷ раздел 9

3. ПРЕДСТАВЛЕНИЕ КС-ГРАММАТИК КС-ДИАГРАММАМИ

Пусть $G = \langle N, \Sigma, P, S \rangle$ – некоторая КС-грамматика, и множество правил для нетерминала A имеет вид:

$$A \rightarrow X_{11}X_{12}\dots X_{1n(1)} \mid X_{21}X_{22}\dots X_{2n(2)} \mid \dots \mid X_{k1}X_{k2}\dots X_{kn(k)}$$

¹ Бесполезным правилом вывода в грамматике G называется такое правило, которое нельзя применить ни в одном выводе предложений из $L(G)$.

² Цепным [1] называется правило вывода вида $A \rightarrow B$.

³ Простым называется нетерминал A , для которого в грамматике имеется ровно одно A -правило $A \rightarrow \alpha$.

Заметим, что из требований 2 и 5 вытекает, что $k > 1$. Следуя [3], назовем КС-диаграммой, изображенную на рисунке 2 двухполюсную сеть d_A в которой

– вершинами являются натуральные числа

$$\begin{aligned} v_1, v_{12}, \dots, v_{1n(1)}, v_n, \\ v_{22}, \dots, v_{2n(2)}, \dots, \\ v_{k2}, \dots, v_{kn(k)} \end{aligned}$$

– входным полюсом является вершина v_1 ;

– выходным полюсом является вершина v_n ;

– ребрами являются упорядоченные пары вершин

$$\begin{aligned} (v_1, v_{12})_{X_{11}}, (v_{12}, v_{13})_{X_{12}}, \dots, (v_{1n(1)}, v_n)_{X_{1n(1)}}, \\ (v_1, v_{22})_{X_{21}}, (v_{22}, v_{23})_{X_{22}}, \dots, (v_{2n(2)}, v_n)_{X_{2n(2)}}, \dots, \\ (v_1, v_{k2})_{X_{k1}}, (v_{k2}, v_{k3})_{X_{k2}}, \dots, (v_{kn(k)}, v_n)_{X_{kn(k)}}. \end{aligned}$$

В КС-диаграмме каждое ребро $(v_1, v_2)_X$ помечено терминальным или нетерминальным символом X . Ребра, помеченные терминальными символами, будем называть терминальными. Ребра, помеченные нетерминальными символами, будем называть нетерминальными. Терминальные ребра $(v_1, v_2)_X$ и $(v_1, v_2)_Y$ считаются разными, если $X \neq Y$. Исходя из требования 4, существование в КС-диаграммах кратных нетерминальных ребер невозможно.

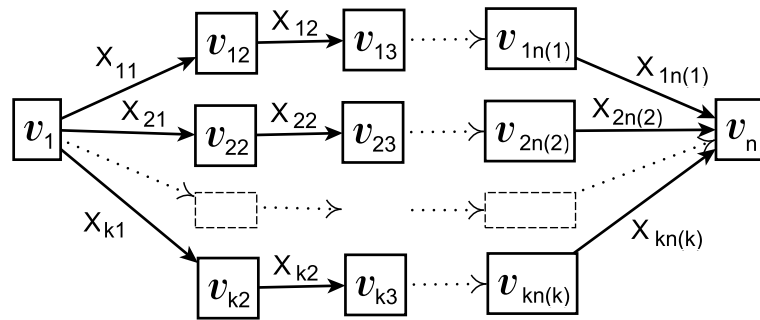


Рис. 2. КС-диаграмма d_A

Нетрудно видеть, что

- входной полюс есть вершина, в которую не входит ни одного ребра;
- выходной полюс есть вершина, из которой не исходит ни одного ребра;
- КС-диаграмма представляет собой параллельное соединение⁴ отдельных ребер и/или их последовательностей.

Рассмотрим КС-грамматику G , удовлетворяющую требованиям 1-5. Для каждого нетерминала этой грамматики можно построить КС-диаграмму. Набор КС-диаграмм для всех нетерминалов можно считать альтернативной формой описания правил вывода. Например, набор КС-диаграмм на рисунке 3 содержит всю информацию о грамматике

$$\begin{aligned} G_E : \quad S &\rightarrow cAc \mid \# \\ A &\rightarrow aB \mid b \mid Ab \\ B &\rightarrow bB \mid b \mid AcB \end{aligned}$$

⁴ Параллельное соединение – здесь – результат раздельного склеивания входных и выходных полюсов двухполюсных сетей.

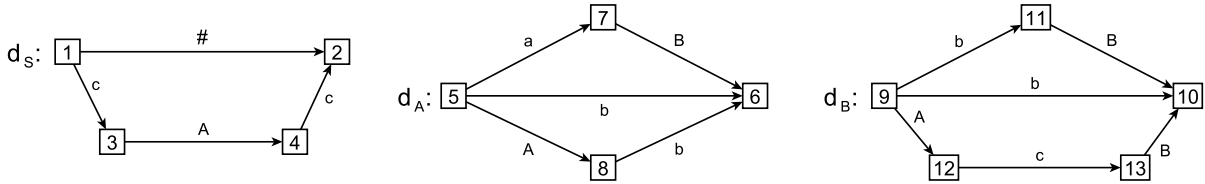


Рис. 3. Набор КС-диаграмм

В связи с грамматиками будем рассматривать КС-диаграммы с уникальными вершинами, не встречающимися в других КС-диаграммах. Свойство уникальности вершин позволяет по заданной вершине v однозначно найти в наборе КС-диаграмм диаграмму d_A , содержащую v .

4. ОПЕРАЦИЯ ПОДСТАНОВКИ ДВУХПОЛЮСНЫХ СЕТЕЙ

Если КС-диаграмма или иная сеть имеет нетерминальное ребро, скажем A , то операция подстановки вместо этого ребра КС-диаграммы d_A буквально "напрашивается". В общем случае

- исходными данными операции подстановки являются:
 - двухполюсная сеть T ;
 - ребро $(w', w'')_X$ сети T ; и
 - двухполюсная сеть t :
 - с входным полюсом v_1 ;
 - с выходным полюсом v_n ;
 - с внутренними вершинами v_{ij} ;
- результатом операции подстановки является двухполюсная сеть.

Заметим, что предлагаемое определение подстановки есть версия "операции подстановки в двухполюсную сеть вместо ее ребра некоторой другой двухполюсной сети" [4]. Особенность настоящей версии состоит в том, что

- сети T и t могут иметь общие вершины и ребра, но
- результат операции определяется однозначно.

Предлагаемое определение операции подстановки основано на специальном представлении вершин сети. А именно, будем полагать, что вершины сети имеют вид

$$i_1.j_1|i_2.j_2| \dots |i_l.j_l|m, \quad \text{где } i_1, j_1, \dots, i_l, j_l, m - \text{натуральные числа, } l \geq 0.$$

Часть $i_1.j_1|i_2.j_2| \dots |i_l.j_l|$ называется префиксом вершины и l - длина префикса.

Число m называется прототипом вершины.

Вершины могут, например, выглядеть так: 347, 3.4|7, 3.4|7.8|7.9|10.

Пара вершин, соединенных ребром $(v', v'')_X$, где $v' = i_1.j_1|i_2.j_2| \dots |i_l.j_l|m$

и $v'' = p_1.q_1|p_2.q_2| \dots |p_r.q_r|u$,

однозначно порождает префикс $pref(v', v'')$ длины $\max(l, r) + 1$:

если $l < r$, то $pref(v', v'') = p_1.q_1|p_2.q_2| \dots |p_r.q_r|m.u$;

если $l \geq r$, то $pref(v', v'') = i_1.j_1|i_2.j_2| \dots |i_l.j_l|m.u$.

Например:

v'	v''	$pref(v', v'')$
3	4	3.4
3.4 7	4	3.4 7.4
3.4 7	3.4 7.4 12	3.4 7.4 7.12

Подстановка сети t вместо некоторого ребра $(w', w'')_X$ предполагает предварительную модификацию t , основанную на функции g переименования вершин:

$$g(v_1) = w', \quad g(v_n) = w'', \quad g(v_{ij}) = \text{pref}(w', w'')v_{ij}.$$

Функция переименования вершин естественным образом порождает функцию переименования ребер

$$g : (v', v'')_X \rightarrow (g(v'), g(v''))_X.$$

Результатом модификации сети t является двухполюсная сеть $\text{mdf}(t, w', w'')$

- вершины которой есть переименованные с помощью функции g вершины сети t , причем
 - входным полюсом является вершина w' ;
 - выходным полюсом является вершина w'' ;
- ребра которой есть переименованные с помощью функции g ребра сети t .

На рисунке 4 приводится пример модификации сети t для подстановки вместо ребра $(w', w'')_B$, где $w' = 7.2|3$ и $w'' = 7.2|4.6|8$, $\text{pref}(w', w'') = 7.2|4.6|3.8|$.

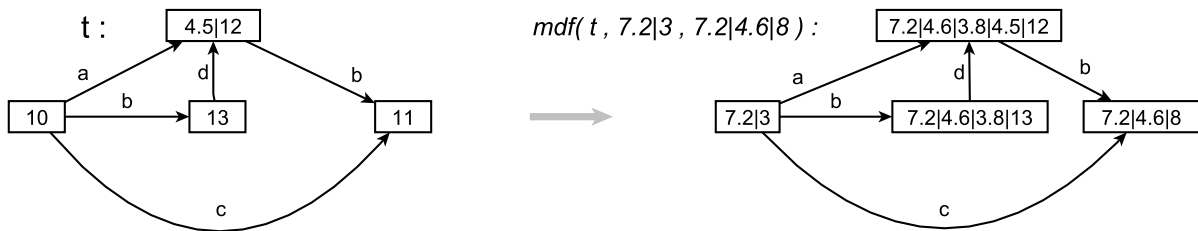


Рис. 4. Пример модификации сети t

Операция подстановки в сеть T вместо ребра $(w', w'')_X$ сети t состоит из трех этапов:

Этап 1. Модифицировать сеть t . $t' = \text{mdf}(t, w', w'')$.

Этап 2. Множество вершин результирующей сети вычислить как объединение вершин сети T и вершин сети t' .

Этап 3. Множество ребер результирующей сети вычислить как объединение ребер сети T за вычетом ребра $(w', w'')_X$ и ребер сети t' .

На рисунке 5 приводится пример операции подстановки вместо ребра $(w', w'')_B$, где $w' = 7.2|3$ и $w'' = 7.2|4.6|8$.

Обобщением операции подстановки является операция одновременной подстановки в сеть T вместо заданного подмножества ребер заданного множества сетей. В операции одновременной подстановки этап 1 выполняется для каждой заданной пары

$$\text{ребро} \leftarrow \text{сеть для подстановки}.$$

На этапах 2 и 3 выполняются операции объединения множеств для всех вершин и ребер ранее модифицированных сетей.

Описанную операцию подстановки можно считать корректной только в том случае, когда внутренние вершины модифицированной сети t не пересекаются с вершинами сети T . В дальнейшем изложении будут использоваться только особые сети, которые

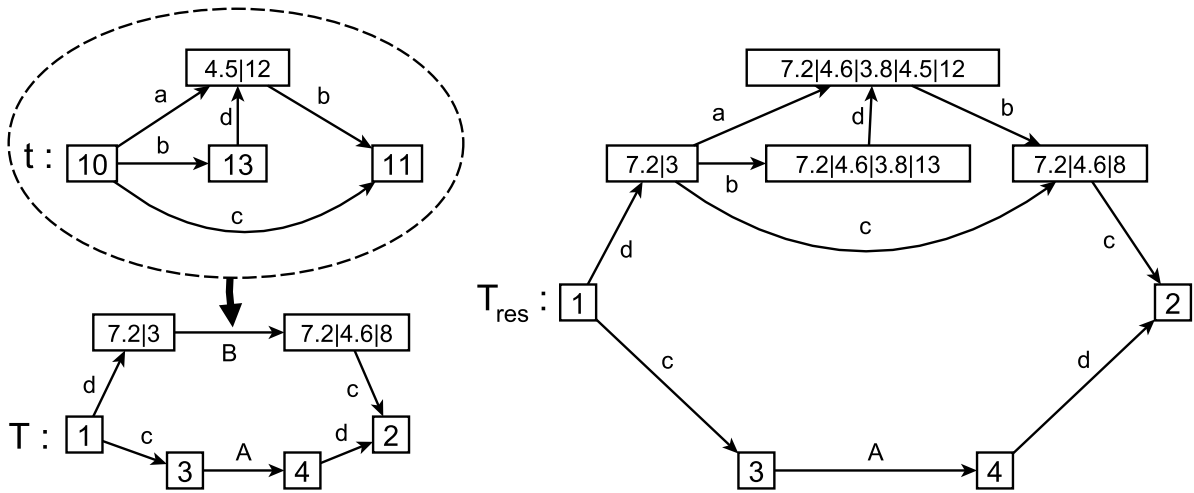


Рис. 5. Пример операции подстановки

либо являются КС-диаграммами грамматики, удовлетворяющей требованиям 1-5;
 либо являются результатом подстановки одной особой сети в другую особую сеть вместо нетерминального ребра.

Корректность операции подстановки для класса особых сетей следует из требования 4 для КС-грамматик и из свойства уникальности вершин набора КС-диаграмм.

Набор КС-диаграмм для каждого нетерминала A естественным образом порождает последовательности сетей

$$T_{0/A}, T_{1/A}, T_{2/A}, \dots,$$

где $T_{0/A} = d_A$, а $T_{i+1/A}$ получается из $T_{i/A}$ путем одновременной подстановки вместо нетерминальных ребер соответствующих КС-диаграмм. На рисунке 6 приводятся сети $T_{0/S}, T_{1/S}, T_{2/S}$ для грамматики G_E .

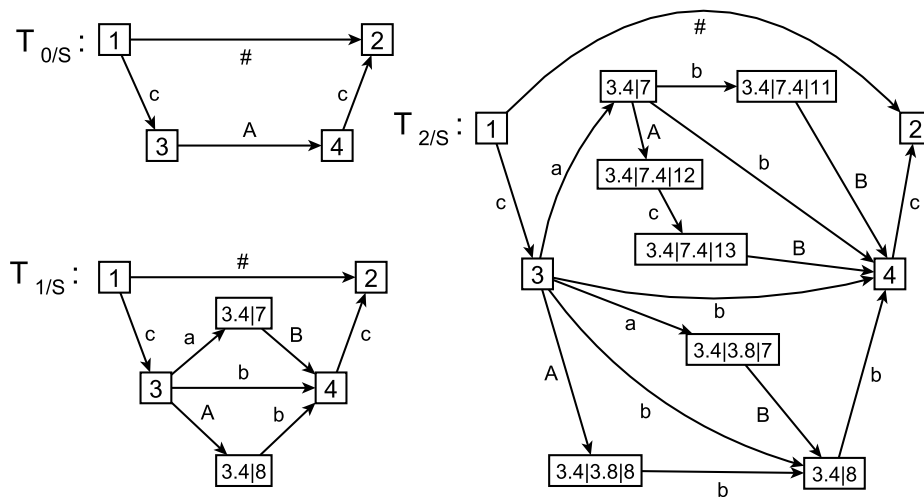


Рис. 6. Пример сетей $T_{0/S}, T_{1/S}, T_{2/S}$

В общем случае сеть $T_{i+1/A}$ получается из сети $T_{k/A}$ одновременной подстановкой вместо нетерминальных ребер соответствующих сетей $T_{i-k/B}$.

Замечание 4.1 ($i = n - 1, k = 0$)

Сеть $T_{n/A}$ состоит

из вершин КС-диаграммы d_A ,

из терминальных ребер d_A , а также

из вершин и ребер сетей $mdf(T_{n-1/A'}, v', v'')$, построенных для каждого нетерминального ребра $(v', v'')_{A'}$ из d_A .

Других вершин и ребер в сети $T_{n/A}$ нет.

Замечание 4.2 Каждый путь, связывающий полюса

– представляет собой последовательность помеченных ребер; и

– порождает цепочку меток, составленную в порядке перечисления ребер.

Например, в сети $T_{2/S}$ путь, изображенный на рисунке 7, порождает цепочку $caAcBc$.

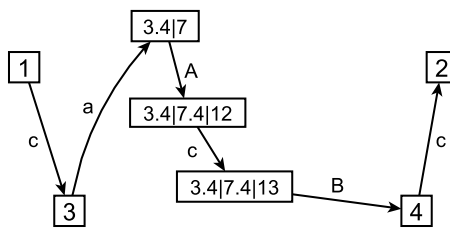


Рис. 7. Пример пути

Множество всех путей, связывающих полюса сети T , порождает множество цепочек $L(T)$. Так, для сети $T_{2/S}$, изображенной на рисунке 6,

$$L(T_{2/S}) = \{ \#, cabBc, cabc, caAcBc, cbc, caBbc, cbbc, cAbbc \}$$

Очевидно, что $L(T_{i/A})$ составляют сентенциальные формы, выводимые из A :

$$\text{Если } \alpha \in L(T_{i/A}), \text{ то } A \Rightarrow_G^* \alpha.$$

5. СПЕЦИАЛЬНОЕ РАСЩЕПЛЕНИЕ ДВУХПОЛЮСНЫХ СЕТЕЙ

По терминологии, принятой в [4], сети $T_{0/A}, T_{1/A}, T_{2/A}, \dots$ являются связными двухполюсными π -сетями из помеченных двухкомпонентных наборов. По определению π -сеть есть двухполюсная сеть, полученная в результате конечного числа подстановок-вместо-ребер сетей, изображенных на рисунке 8.

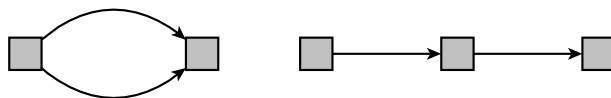


Рис. 8. Γ_2^p и Γ_2^s - примитивы для построения π -сетей

Π -сети подразделяются на s-сети и p-сети:

- s-сеть⁵ есть последовательное соединение отдельных ребер и/или p-сетей;
- p-сеть⁶ есть параллельное соединение отдельных ребер и/или s-сетей.

В дальнейшем ребра и p-сети, из которого путем последовательного соединения получается s-сеть, будем называть компонентами s-сети. Аналогично, ребра и s-сети, из которых путем параллельного соединения получается p-сеть, будем называть компонентами p-сети. На рисунке 9 представлены p-сеть и ее три компонента: одно ребро и две s-сети; в свою очередь для обеих s-сетей представлены их компоненты: две p-сети для первой и два ребра плюс p-сеть для второй.

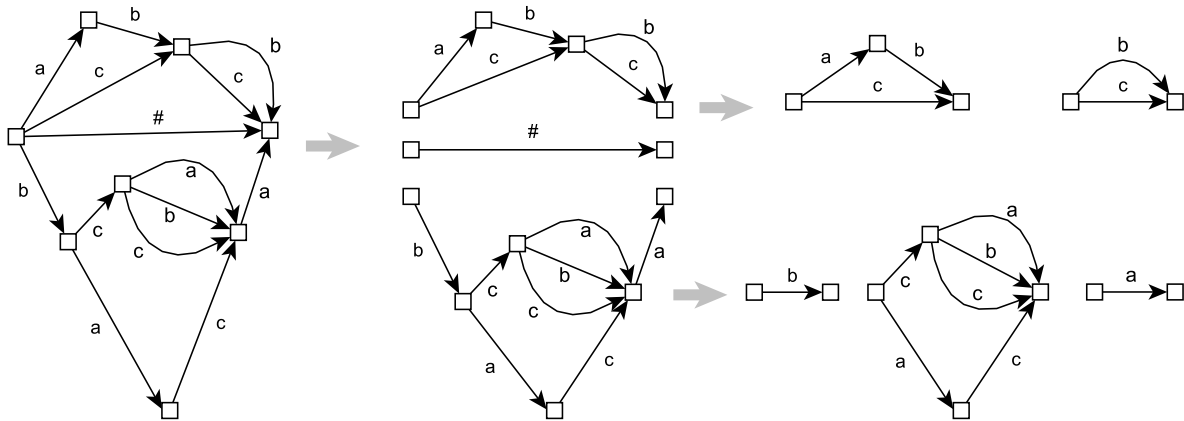


Рис. 9. Компоненты одной p-сети и двух s-сетей

Операция специального расщепления применяется к заданной p-сети и состоит в выделении из нее подсетей и их обработке. Процесс начинается с занесения в очередь заданной сети. Очередная p-сеть \mathcal{T} , подлежащая расщеплению, выбирается из очереди и подвергается обработке. При этом возможны два случая.

В простом случае \mathcal{T} представляет собой прототип КС-диаграммы – суть – параллельное соединение отдельных ребер и/или их последовательностей. В этом случае результатом специального расщепления является сама сеть \mathcal{T} .

В общем случае в \mathcal{T} однозначно выделяются несколько p-сетей. Эти выделенные p-сети представляют собой отличные от ребер компоненты-от-компонент сети \mathcal{T} . В этом случае операция специального расщепления

- ▷ производит прототип КС-диаграммы, полученный из \mathcal{T} заменой каждой выделенной p-сети ребром с ее именем; и
- ▷ заносит в очередь набор поименованных выделенных p-сетей.

Процесс расщепления продолжается до исчерпания очереди⁷.

Пример специального расщепления приведен на рисунке 10. В этом примере из заданной сети $S_{\text{сеть}}$ выделяются сети $A_{\text{сеть}}$, $B_{\text{сеть}}$, $C_{\text{сеть}}$ и $D_{\text{сеть}}$. Под общий случай операции специального расщепления подпадают сети $S_{\text{сеть}}$ и $A_{\text{сеть}}$; под простой случай – $B_{\text{сеть}}$, $C_{\text{сеть}}$ и $D_{\text{сеть}}$.

Все выделенные p-сети специального расщепления образуют, так называемый, p-спектр заданной сети. Иначе говоря, элементами p-спектра являются только те сети, которые "прошли"

⁵ По терминологии, принятой в [4], s-сеть есть s-разложимая π -сеть.

⁶ По терминологии, принятой в [4], p-сеть есть p-разложимая π -сеть.

⁷ Очередь как структура данных поддерживает перебор подсетей в ширину.

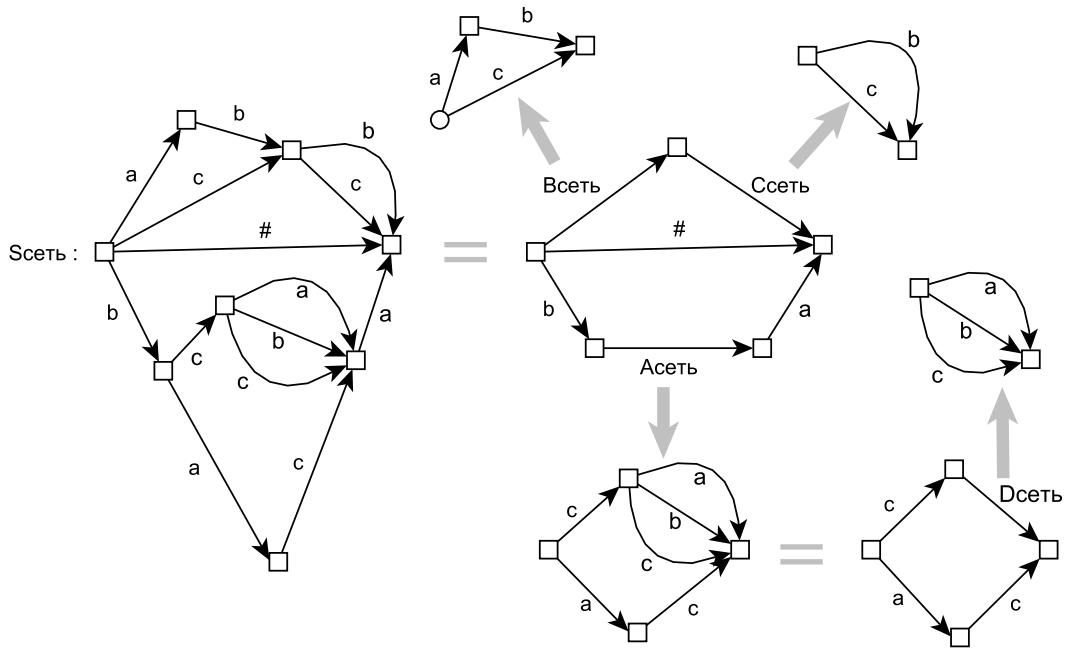


Рис. 10. Специальное расщепление π -сети $S_{\text{сеть}}$ на р-сети

через очередь; прототипы КС-диаграмм произведенные в "общем случае" элементами р-спектра не являются. Элементы р-спектра (см., например, рисунок 11) связаны иерархическими связями, в соответствии с которыми каждому элементу однозначно приписывается уровень вложенности.

Замечание 5.1. Элемент р-спектра t как и другие подсети сети T содержит вершины и ребра из T . Других вершин и ребер в t нет.

Замечание 5.2. Пусть v' и v'' - полюса некоторого элемента р-спектра сети $T_{n/A}$, построенной из набора КС-диаграмм с уникальными вершинами. Тогда по известным v' и v'' можно однозначно установить нетерминальную метку ребра (v', v'') на месте которого располагается элемент р-спектра:

- ▷ из вершин v' и v'' выбрать вершину v с префиксом максимальной длины;
- ▷ в наборе КС-диаграмм по прототипу вершины v определить d_B .

Ответ: элементу р-спектра с вершинами v' и v'' соответствует нетерминал B .

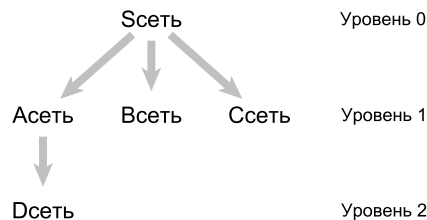


Рис.11. Структура р-спектра сети $S_{\text{сеть}}$

Замечание 5.3. Операция специального расщепления неявно предполагает наличие функции именования выявленных r -сетей. В рассмотренном примере предполагалось, что каждая вновь выделенная r -сеть получает уникальное имя: $A_{\text{сеть}}$, $B_{\text{сеть}}$ и т.д. Иногда одинаковые имена могут получать, вообще говоря, разные элементы r -спектра. В этом случае сети с уже встречавшимися именами в очередь не заносятся.

Замечание 5.4. Побочным продуктом операции специального расщепления является КС-грамматика, в которой

- имена элементов r -спектра используются в качестве нетерминальных символов, а
- прототипы КС-диаграмм d_E трансформируются в правила вывода $\{ E \rightarrow \alpha \mid \alpha \in L(d_E) \}$.

Для примера на рисунке 10 упомянутая КС-грамматика выглядит так:

$$\begin{array}{lcl}
 S_{\text{сеть}} & \rightarrow & B_{\text{сеть}}C_{\text{сеть}} \mid \# \mid bA_{\text{сеть}}a \\
 A_{\text{сеть}} & \rightarrow & cD_{\text{сеть}} \mid ac \\
 B_{\text{сеть}} & \rightarrow & ab \mid c \\
 C_{\text{сеть}} & \rightarrow & b \mid c \\
 D_{\text{сеть}} & \rightarrow & a \mid b \mid c
 \end{array}$$

6. ПРЕОБРАЗОВАНИЯ ДВУХПОЛЮСНЫХ СЕТЕЙ

Для π -сетей с помеченными ребрами определим ряд стандартных преобразований. Каждое преобразование

- имеет условное наименование $\phi \mapsto \psi$, в которое могут включаться параметры;
- применяется к заданной π -сети T ;
- порождает π -сеть с помеченными ребрами, которая обозначается $[T]_{\phi \mapsto \psi}$.

Будем различать четыре типа преобразований:

- отсечение подсетей $dp(k) \mapsto \{U\}$;
- переименование нетерминальных ребер $N \mapsto N'$;
- обезличение нетерминальных ребер $N \mapsto \{U\}$;
- удаление нетерминальных ребер $N \mapsto \emptyset$.

Преобразование $dp(k) \mapsto \{U\}$ состоит в том, что в π -сети T каждый элемент r -спектра глубины k заменяется ребром с меткой U . Результатом преобразования является сеть $[T]_{dp(k) \mapsto \{U\}}$. На рисунке 12 представлены π -сети, полученные из сети $S_{\text{сеть}}$ в результате преобразований $dp(1) \mapsto \{U\}$ и $dp(2) \mapsto \{U\}$.

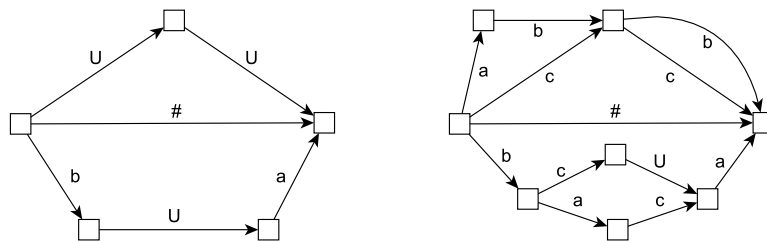


Рис.12. Преобразования $dp(1) \mapsto \{U\}$ и $dp(2) \mapsto \{U\}$ для сети $S_{\text{сеть}}$

Преобразование $N \mapsto N'$ исходит из того, что существует функция переименования нетерминалов $h: N \rightarrow N'$, где N' - множество новых нетерминальных символов. Сеть $[T]_{N \mapsto N'}$ получается из сети T посредством замены нетерминальных меток ребер A на метки $h(A)$.

Преобразование $N \mapsto \{U\}$ есть вариант преобразования $N \mapsto N'$ на случай $N' = \{U\}$. Результатом операции является сеть $[T]_{N \mapsto \{U\}}$.

Преобразование $N \mapsto \emptyset$ состоит в том, что из сети T удаляются ребра, помеченные нетерминальными символами, и ребра, через которые не проходят пути, связывающие полюса.

Результатом преобразования является сеть $[T]_{N \mapsto \emptyset}$.

Если $T = T_{n/A}$ для некоторой КС-грамматики, то сеть $[T_{n/A}]_{N \mapsto \emptyset}$ будем обозначать $t_{n/A}$.

На рисунке 13 представлены сети $t_{0/S}$, $t_{1/S}$, $t_{2/S}$ для грамматики G_E из раздела 3:

$$L(t_{0/S}) = \{\#\}, \quad L(t_{1/S}) = \{\#, abc\}, \quad L(t_{2/S}) = \{\#, abc, cabc, cbbc\}.$$

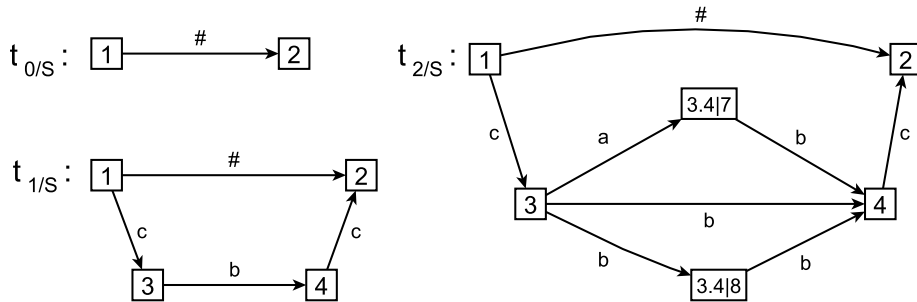


Рис. 13. Пример сетей $t_{0/S}$, $t_{1/S}$, $t_{2/S}$

Замечание 6.1 $L(t_{i/A}) \subseteq L(T_{i/A})$. Подсеть $t_{i/A}$, порождает исключительно предложения из терминальных символов.

Замечание 6.2 Сеть $t_{i+1/A}$ содержит в качестве собственной подсети сеть $t_{i/A}$ и, возможно, некоторое количество новых ребер, помеченных терминальными символами:

$$L(t_{0/A}) \subseteq L(t_{1/A}) \subseteq L(t_{2/A}) \subseteq \dots \quad L(t_{i/S}) \subset L(G).$$

Замечание 6.3 Теоретически можно определить бесконечную π -сеть t_A , состоящую из вершин и ребер сетей $t_{n/A}$. Тогда сеть t_S можно рассматривать как структуру всего КС-языка, а не его отдельного фрагмента. Кроме того, р-спектр⁸ любой бесконечной сети t_A оказывается конечным множеством (с точностью до изоморфизма). В настоящем изложении бесконечные сети не исследуются, что приводит к необходимости зонирования⁹ конечных π -сетей.

Замечание 6.4. Сети $T_{0/A}$, $T_{1/A}$, $T_{2/A}$, ... являются р-сетями, это следует из требований 2 и 5. Относительно сетей $t_{0/S}$, $t_{1/S}$, $t_{2/S}$, ... утверждение о их строении выглядит сложнее.

Во-первых, существует целое $n = n_S$, зависящее от грамматики G , такое, что

$$\begin{aligned} t_{0/S}, \dots, t_{n-1/S} & \text{ являются s-сетями,} \\ t_{n/S}, t_{n+1/S}, \dots & \text{ являются р-сетями.} \end{aligned}$$

Во-вторых, для каждого нетерминала A , $A \neq S$, существует целое $j = j_A$, зависящее от грамматики G , такое, что

$$\begin{aligned} t_{0/A}, \dots, t_{j-1/A} & \text{ не имеют ребер вообще,} \\ t_{j/A}, t_{j+1/A}, \dots & \text{ являются s- или р-сетями.} \end{aligned}$$

⁸ При корректном переопределении р-спектра.

⁹ См. замечания 6.4 и 6.5, а также утверждение 3.

Таким образом, каждая грамматика G , удовлетворяющая требованиям 1-5, позволяет вычислить характеристику

$$join(G) = \max\{j_A \mid A \in N, A \neq S\},$$

относительно которой справедливо

если $A \neq S$, то для $j = join(G)$ все сети $t_{j/A}$ являются связными;
 если $A \in N$, то для $j > join(G)$ все сети $t_{j/A}$ являются р-сетями,
 причем количество компонент в $t_{j/A}$ совпадает с d_A .

Для грамматики G_A из раздела 8 $j_A = j_B = j_E = 0, j_C = 1, j_D = 2, join(G_A) = 2$.

Для грамматики G_E из раздела 2 $join(G_E) = 0$.

Конец замечания 6.4.

Целочисленная характеристика $j = join(G)$ связывает семейства сетей $t_{n/A}$ и $T_{n/A}$:

- сеть $t_{n+j/A}$ содержит все вершины сети $T_{n/A}$;
- в сети $t_{n+j+1/A}$, на местах нетерминальных ребер из $T_{n/A}$ размещаются элементы р-спектра, то есть

$$[T_{n/A}]_{N \rightarrow \{U\}} = [t_{n+j+1/A}]_{dp(n+1) \rightarrow \{U\}}.$$

Замечание 6.5. Зафиксируем целое число n , $n > \|N\| + join(G)$. В сети $t_{n/S}$ каждый р-элемент однозначно соответствует определенному нетерминалу¹⁰:

- р-элемент 0-го уровня - сеть $t_{n/S}$ - соответствует нетерминалу S ;
- р-элементы 1-го уровня соответствуют нетерминалам из правых частей S -правил и т.д.

Перебирая р-элементы в порядке возрастания их уровней и отмечая впервые встретившиеся нетерминалы, на некотором уровне - обозначим его $main(G)$ - неизбежно обнаружится, что новые нетерминалы более не встречаются. Это следует из конечности множества нетерминалов и требования 3.

Для грамматики G_A из раздела 8 $main(G_A) = 4, \|N\| + join(G_A) = 8$.

Для грамматики G_E из раздела 2 $main(G_E) = 2$.

7. ИЗОМОРФИЗМ ДВУХПОЛЮСНЫХ СЕТЕЙ

При сравнении π -сетей естественно сохранять разметку ребер. В частности, это верно для отношения изоморфизма.

Взаимно однозначное соответствие между вершинами и ребрами π -сетей Π_1 и Π_2 называется изоморфизмом (отношением изоморфизма), если соответствующие ребра имеют одинаковые метки и соединяют соответствующие вершины.

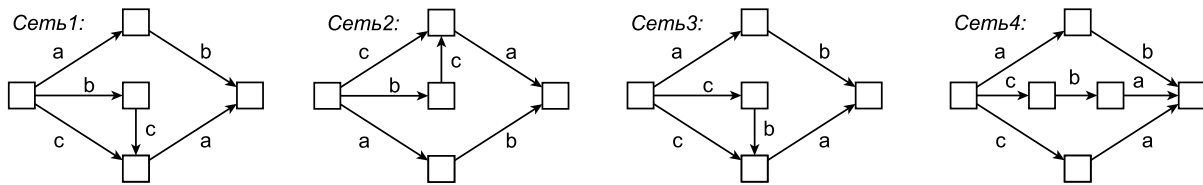
Замечание 7.1 Фактически изоморфизм представляет собой множество пар двух сортов: *вершина* \leftrightarrow *вершина* и *ребро* \leftrightarrow *ребро*. Вершины и ребра, входящие в одну пару, именуются соответствующими.

Π -сети Π_1 и Π_2 называются изоморфными, если для них существует хотя бы одно отношение изоморфизма. Факт наличия изоморфизма будем записывать так $\Pi_1 \approx \Pi_2$.

На рисунке 14 представлены четыре сети, для которых справедливо:

- o *Сеть1* \approx *Сеть2*;
- o *Сеть1* $\not\approx$ *Сеть3* из-за невозможности сохранить разметку ребер;
- o *Сеть3* $\not\approx$ *Сеть4* из-за топологических различий, хотя $L(\text{Сеть3}) = L(\text{Сеть4})$.

¹⁰ См. также замечание 5.2.

Рис. 14. Примеры π -сетей

Описанная в разделе 4 операция модификации особой сети t порождает изоморфную ей особую сеть $mdf(t, w', w'')$, причем функция переименования вершин и ребер g задает отношение изоморфизма.

Для наших целей интерес представляет изоморфизм-с-оговорками, при котором заданные сети Π_1 и Π_2 , вообще говоря, неизоморфны, но при некотором фиксированном преобразовании $\phi \mapsto \psi$ из них все-таки получается пара изоморфных сетей. Для отношений такого рода будем использовать запись

$$[\Pi_1 \approx \Pi_2]_{\phi \mapsto \psi}$$

При исследовании пары сетей на изоморфизм приходится считаться с тем, что отношений изоморфизма может быть несколько. При этом некоторые изоморфизмы оказываются "полезными", а другие необходимо "нейтрализовать".

8. СТРУКТУРНАЯ ИЗБЫТОЧНОСТЬ

Каждый нетерминал грамматики порождает семейство π -сетей. Рассмотрим вопрос о различимости нетерминалов по структурным свойствам этих сетей. Вынужденно начнем издалека.

Будем рассматривать разбиения множества нетерминалов КС-грамматик. Разбиение нетерминалов K в КС-грамматике $G = \langle N, \Sigma, P, S \rangle$ образуют такие непересекающиеся подмножества (классы) нетерминалов, объединение которых есть N . $K = \{K_C \mid C \in N'\}$. Здесь в качестве индексов используются символы некоторого алфавита N' .

Среди возможных разбиений выделим, так называемое, тривиальное разбиение, каждый класс которого состоит ровно из одного нетерминала. Соответственно в нетривиальном разбиении хотя бы один класс содержит два и более нетерминалов.

Каждое разбиение K порождает

- (а) функцию переименования нетерминалов $h_K : N \rightarrow N'$ такую, что $h_K(A) = C$, если $A \in K_C$;
- (б) преобразование H_K цепочек из $(N \cup \Sigma)^*$ и правил вывода из P :
 - если $\alpha = y_0 A_1 y_1 \dots A_m y_m$, то $H_K(\alpha)$ есть $y_0 A'_1 y_1 \dots A'_m y_m$, где $A'_i = h_K(A_i)$, $i = 1, \dots, m$;
 - если p имеет вид $A \rightarrow \alpha$, то $H_K(p)$ есть $h_K(A) \rightarrow H_K(\alpha)$;
- (в) КС-грамматику $G_K = \langle N', \Sigma, H_K(P), h_K(S) \rangle$.

В качестве примера рассмотрим "абстрактную" КС-грамматику G_A и разбиение нетерминалов $K : \{S\}_{S'}, \{A, B\}_{A'}, \{C, D, E\}_{C'}$.

$$G_A : \begin{array}{l} S \rightarrow bC \mid aAD \mid \# \\ A \rightarrow bC \mid aAD \mid b \\ B \rightarrow cE \mid aAE \mid aA \mid c \\ C \rightarrow aB \mid aBC \\ D \rightarrow BC \mid AaD \mid dD \\ E \rightarrow Aa \mid AaE \mid dE \mid d \end{array}$$

В этом случае: $H_K(bC) = bC'$, $H_K(aAD) = aA'C'$ и т.д.
 $H_K(S \rightarrow bC)$ есть $S' \rightarrow bC'$,
 $H_K(A \rightarrow aAD)$ есть $A' \rightarrow aA'C'$,
 $H_K(B \rightarrow aAE)$ есть $A' \rightarrow aA'C'$ и т.д.

$$(G_A)_K : \begin{array}{l} S' \rightarrow bC' \mid aA'C' \mid \# \\ A' \rightarrow bC' \mid aA'C' \mid aA' \mid cC' \mid c \mid b \\ A' \rightarrow dC' \mid aA'C' \mid aA' \mid A'C' \mid A'aC' \mid A'a \mid d \end{array}$$

Пусть на множестве нетерминалов КС-грамматики $G = \langle N, \Sigma, P, S \rangle$ зафиксировано разбиение K , $h_K: N \rightarrow N'$.

Будем обозначать $A \cong B$,

если $H_K(P_A) = H_K(P_B)$, где P_A и P_B – множества A - и B -правил из P .

Будем называть нетривиальное разбиение K согласованным,

если $A \cong B$ для любой пары нетерминалов A и B из одного класса разбиения.

$$\text{Для КС-грамматики } G_1 : \begin{array}{l} S \rightarrow \# \mid A + B \\ A \rightarrow a \mid aA \\ B \rightarrow a \mid aB \mid aC \\ C \rightarrow a \mid aC \end{array}$$

разбиение нетерминалов $K: \{S\}_{S'}, \{A\}_{A'}, \{B, C\}_{B'}$

является согласованным:

$$P_B = \{ B \rightarrow a, B \rightarrow aB, B \rightarrow aC \}, \quad H_K(P_B) = \{ B' \rightarrow a, B' \rightarrow aB' \}.$$

$$P_C = \{ C \rightarrow a, C \rightarrow aC \}, \quad H_K(P_C) = \{ B' \rightarrow a, B' \rightarrow aB' \} = H_K(P_B).$$

$$\text{В полученной грамматике } (G_1)_K : \begin{array}{l} S' \rightarrow \# \mid A' + B' \\ A' \rightarrow a \mid aA' \\ B' \rightarrow a \mid aB' \end{array}$$

можно еще раз выделить согласованное разбиение нетерминалов и окончательно преобразовать грамматику G_1 к виду:

$$\begin{array}{l} S_r \rightarrow \# \mid A_r + A_r \\ A_r \rightarrow a \mid aA_r \end{array}$$

Замечание 8.1 Задача обнаружения в заданной грамматике согласованного разбиения алгоритмически разрешима и сводится к перебору и проверке всех разбиений конечного множества нетерминалов.

Утверждение 1. Если K - согласованное разбиение множества нетерминалов КС-грамматики G , то $L(G) = L(G_K)$.

Доказательство. Зафиксируем некоторое согласованное разбиение K произвольной КС-грамматики $G = \langle N, \Sigma, P, S \rangle$. Для сокращения записи введем обозначения: $h = h_K$, $H = H_K$, $G' = \langle N', \Sigma, P', C' \rangle = G_K = \langle N', \Sigma, H(P), h(S) \rangle$.

Во-первых, покажем $L(G) \subseteq L(G_K)$. Рассмотрим произвольное предложение x из $L(G)$ и некоторый левый вывод x в грамматике G .

$$\alpha_0, \alpha_1, \dots, \alpha_{k-1}, \alpha_k : \quad \alpha_0 \Rightarrow_G \alpha_1 \Rightarrow_G \dots \Rightarrow_G \alpha_{k-1} \Rightarrow_G \alpha_k = x.$$

Докажем по индукции, что последовательность цепочек $H(\alpha_0), \dots, H(\alpha_k)$ есть левый вывод предложения x в грамматике G' .

Базис индукции. По определению левого вывода $\alpha_0 = S$ и $H(\alpha_0) = C'$, то есть $H(\alpha_0)$ - цепочка, выводимая в грамматике G_K .

Индуктивный переход. Предположим, что цепочка $H(\alpha_i)$ выводима в грамматике G' для некоторого $i < k$. По определению левого вывода для грамматики G цепочка α_{i+1} получена из α_i посредством некоторого правила вывода $A_1 \rightarrow x_0 B_1 x_1 \dots B_n x_n$ из P :

$$\alpha_i = z_0 A_1 \beta_i \Rightarrow_G z_0 x_0 B_1 x_1 \dots B_n x_n \beta_i = \alpha_{i+1}.$$

По определению грамматики G' имеем

$$h(A_1) \rightarrow x_0 h(B_1) x_1 \dots h(B_n) x_n \in P'.$$

Следовательно,

$$H(\alpha_i) = z_0 h(A_1) H(\beta_i) \Rightarrow_{G'} z_0 x_0 h(B_1) x_1 \dots h(B_n) x_n H(\beta_i) = H(\alpha_{i+1}).$$

То есть цепочка $H(\alpha_{i+1})$ также выводима в G' , а значит и $H(\alpha_k)$ выводима в G' . Учитывая, что $H(\alpha_k) = \alpha_k = x$, окончательно получаем: $x \in L(G')$ и $L(G) \subseteq L(G')$.

Во-вторых, покажем $L(G) \supseteq L(G_K)$. Рассмотрим произвольное предложение y из $L(G')$ и некоторый левый вывод $\beta_0, \beta_1, \dots, \beta_k$ этого предложения в грамматике G' .

$$C' = \beta_0 \Rightarrow_{G'} \beta_1 \Rightarrow_{G'} \dots \Rightarrow_{G'} \beta_{k-1} \Rightarrow_{G'} \beta_k = y$$

Докажем по индукции факт существования левого вывода $\alpha_0, \alpha_1, \dots, \alpha_k$ в грамматике G , удовлетворяющего условиям $H(\alpha_i) = \beta_i, i = 1, \dots, k$.

Базис индукции. Положим $\alpha_0 = S$, при этом оказывается, что $H(\alpha_0) = H(S) = h(S) = C' = \beta_0$.

Индуктивный переход. Предположим, что для некоторого $i, i < k$, удалось построить вывод

$$\alpha_0 \Rightarrow_{G'} \alpha_1 \Rightarrow_{G'} \dots \Rightarrow_{G'} \alpha_i, \text{ причем } H(\alpha_i) = \beta_i.$$

Равенство $H(\alpha_i) = \beta_i$ означает, что

$$\begin{aligned} \alpha_i &= z_0 A_1 z_1 A_2 z_2 \dots A_m z_m, \\ \beta_i &= z_0 C_1 z_1 C_2 z_2 \dots C_m z_m, \text{ и } h(A_1) = C_1, \dots, h(A_m) = C_m. \end{aligned}$$

Как следует из определения левого вывода, в грамматике G' существует правило

$$C_1 \rightarrow x_0 C'_1 x_1 \dots C'_n x_n \text{ такое, что } \beta_{i+1} = z_0 x_0 C'_1 x_1 \dots C'_n x_n z_1 C_2 z_2 \dots C_m z_m.$$

Для этого правила, согласно определению множества P' , в грамматике G существует, по крайней мере, одно "родительское" правило

$$B_1 \rightarrow x_0 B'_1 x_1 \dots B'_n x_n \text{ такое, что } h(B_1) = C_1, h(B'_1) = C'_1, \dots, h(B'_n) = C'_n.$$

Итак, $h(A_1) = C_1$ по индуктивному предположению,

и $h(B_1) = C_1$ из свойств левого вывода в грамматике G' (для некоторого B_1).

Поэтому $A_1 \in K_{C_1}, B_1 \in K_{C_1}$ и

либо (случай 1) A_1 и B_1 совпадают,

либо (случай 2) A_1 и B_1 различны, но $A_1 \cong B_1$ в силу согласованности разбиения K .

В первом случае "родительское" правило имеет вид

$$A_1 \rightarrow x_0 B'_1 x_1 \dots B'_n x_n \text{ и с помощью этого правила } \alpha_i \Rightarrow_G \alpha_{i+1},$$

где α_{i+1} есть цепочка $z_0 x_0 B'_1 x_1 \dots B'_n x_n z_1 A_2 z_2 \dots A_m z_m$,

$$h(B'_1) = C'_1, \dots, h(B'_n) = C'_n,$$

$$h(A_2) = C_2, \dots, h(A_m) = C_m.$$

Во втором случае в грамматике G существует также правило

$$A_1 \rightarrow x_0 A'_1 x_1 \dots A'_n x_n \quad \text{такое, что} \quad h(A'_1) = h(B'_1) = C'_1, \quad \dots, \quad h(A'_n) = h(B'_n) = C'_n,$$

и с помощью этого правила $\alpha_i \Rightarrow_G \alpha_{i+1}$,

где α_{i+1} есть цепочка $z_0 x_0 A'_1 x_1 \dots A'_n x_n z_1 A_2 z_2 \dots A_m z_m$,

$$h(A'_1) = C'_1, \quad \dots, \quad h(A'_n) = C'_n,$$

$$h(A_2) = C_2, \quad \dots, \quad h(A_m) = C_m.$$

В обоих случаях $H(\alpha_{i+1}) = \beta_{i+1}$ и $\alpha_0 \Rightarrow_G^+ \alpha_{i+1}$. Сравнивая цепочки α_i и β_i , а также, учитывая, что для $i = k$ имеет место равенство $m = 0$, получаем: $\alpha_k = \beta_k = y$. Итак, $y \in L(G)$ и $L(G_K) \subseteq L(G)$.

Утверждение 1 доказано.

Замечание 8.2 Если в КС-грамматике $\langle N, \Sigma, P, S \rangle$ для пары нетерминалов A и B имеет место $A \cong B$, то грамматики $\langle N, \Sigma, P, A \rangle$ и $\langle N, \Sigma, P, B \rangle$ структурно эквивалентны [5]. Обратное утверждение неверно.

Следствие 1. Если K - согласованное разбиение множества нетерминалов КС-грамматики $G = \langle N, \Sigma, P, S \rangle$, то для любого нетерминала A , $A \in N$,

$$L(G^A) = L(G_K^A), \quad \text{где } L(G^A) = \langle N, \Sigma, P, A \rangle.$$

Следствие 2. Из утверждения 1 следует, что устранение согласованного разбиения в грамматике G , состоящее в переходе к G_K , является эквивалентным преобразованием КС-грамматик. Причем устранение одного такого разбиения уменьшает количество нетерминалов в преобразованной грамматике. Таким образом, к КС-грамматикам можно формально предъявить дополнительное

Требование 6. В КС-грамматике отсутствуют согласованные разбиения нетерминалов.

Как показано в [2] свойств, сформулированных в следствии 2 и в замечаниях 8.1 и 8.2, достаточно для доказательства факта существования грамматик, одновременно удовлетворяющих требованиям 1-5 и 6.

9. ЧТО ЖЕ ИЗ ЭТОГО СЛЕДУЕТ?

Отсутствие в КС-грамматике согласованных разбиений позволяет установить ряд структурных свойств π -сетей, ассоциированных с этой грамматикой.

Замечание 9.1 Если грамматика удовлетворяет требованиям 1-5 и 6, то по известной сети $\mathcal{T} = \text{mdf}(d_A, w', w'')$ диаграмма d_A (а равно - нетерминал A) вычисляется однозначно:

либо в \mathcal{T} имеются внутренние вершины, и тогда по прототипу одной из них в наборе КС-диаграмм однозначно находится d_A ;

либо d_A есть параллельное соединение некоторых терминальных ребер, которое из-за отсутствия согласованных разбиений является единственным.

Утверждение 2.

Пусть (а) G - КС-грамматика, удовлетворяющая требованиям 1-5 и 6,

(б) A и B - пара различных нетерминалов из G таких, что

$$t_{n/A} \approx t_{n/B} \quad \text{для любого } n > \text{join}(G), \quad (1)$$

тогда в грамматике G существует согласованное разбиение нетерминалов.

Доказательство. Для некоторого $n > (||N|| - 1) * (||N|| - 2)/2 + \text{join}(G)$ зафиксируем отношение изоморфизма f для сетей

$$\Pi_A = t_{n/A} \quad \text{и} \quad \Pi_B = t_{n/B}.$$

По заданному f построим множество F , элементами которого являются некоторые взаимно однозначные функции вида $f_{AB}: d_A \rightarrow d_B$.

О множестве F известно, что первоначально $F = \emptyset$, а при занесении f_{AB} в F также в F заносится функция $f_{BA} = f_{AB}^{-1}$ и выполняется транзитивное замыкание:

$$\text{если } f_{AB} \in F \text{ и } f_{BC} \in F, \text{ то } f_{AC} \in F, \quad \text{где } f_{AC}(x) \stackrel{\text{def}}{=} f_{BC}(f_{AB}(x)).$$

Итерация No.1 Из строения сетей Π_A и Π_B – замечания 4.1 и 6.4 – вытекает существование взаимно однозначного соответствия $f_{AB}: d_A \rightarrow d_B$, при котором

- соответствие вершин d_A вершинам d_B порождается изоморфизмом f ;
- соответствие терминальных ребер d_A терминальным ребрам d_B также порождается изоморфизмом f ;
- каждому нетерминальному ребру $(v', v'')_{A'}$ из d_A соответствует нетерминальное ребро $(w', w'')_{B'}$ из d_B такое, что
 - $w' = f_{AB}(v')$,
 - $w'' = f_{AB}(v'')$,
 - $\Pi_{A'} = \text{mdf}(t_{n-1/A'}, v', v'')$
 - и $\Pi_{B'} = \text{mdf}(t_{n-1/B'}, w', w'')$ – пара сетей, связанных отношением f .

Функция f_{AB} заносится в F , а пары $\Pi_{A'}$ и $\Pi_{B'}$, выявленные при ее построении, заносится в очередь для обработки. *Конец итерации No.1.*

Итерация No.2 Если в очереди нашлась пара сетей $\Pi_{A'}$ и $\Pi_{B'}$ таких, что $A' \neq B'$ и $f_{A'B'} \notin F$, то по аналогии с парой Π_A и Π_B существует взаимно однозначное соответствие $f'_{A'B'}: d'_{A'} \rightarrow d'_{B'}$,

где $d'_{A'} = \text{mdf}(d_{A'}, v', v'')$ и $d'_{B'} = \text{mdf}(d_{B'}, w', w'')$, при котором

- соответствие вершин $d'_{A'}$ вершинам $d'_{B'}$ порождается изоморфизмом f ;
- соответствие терминальных ребер $d'_{A'}$ терминальным ребрам $d'_{B'}$ также порождается изоморфизмом f ;
- каждому нетерминальному ребру $(v'_2, v''_2)_{A''}$ из $d'_{A'}$ соответствует нетерминальное ребро $(w'_2, w''_2)_{B''}$ из $d'_{B'}$ такое, что
 - $w'_2 = f'_{A'B'}(v'_2)$,
 - $w''_2 = f'_{A'B'}(v''_2)$,
 - $\Pi_{A''} = \text{mdf}(t_{n-2/A''}, v'_2, v''_2)$
 - и $\Pi_{B''} = \text{mdf}(t_{n-2/B''}, w'_2, w''_2)$ – пара сетей, связанных отношением f .

Из замечания 9.1 следует, что по известным $d'_{A'}$ и $d'_{B'}$ КС-диаграммы $d_{A'}$ и $d_{B'}$ восстанавливаются однозначно, и это позволяет однозначно перейти от $f'_{A'B'}: d'_{A'} \rightarrow d'_{B'}$ к $f_{A'B'}: d_{A'} \rightarrow d_{B'}$. Функция $f_{A'B'}$ заносится в F , а пары $\Pi_{A''}$ и $\Pi_{B''}$, выявленные при ее построении, заносится в очередь для обработки. *Конец итерации No.2.*

Процесс продолжается до тех пор пока в очереди обнаруживаются пары сетей Π_{A^*} и Π_{B^*} такие, что $A^* \neq B^*$ и $f_{A^*B^*} \notin F$. В самом неблагоприятном случае для построения множества F потребуются $(\|N\| - 1) * (\|N\| - 2)/2$ итераций, каждая из которых выполняется на новом уровне р-спектра сетей Π_A и Π_B .

Множество F порождает разбиение нетерминалов $K(F)$: два нетерминала A и B относятся к одному классу K_C разбиения $K(F)$, если $f_{AB} \in F$. По построению $K(F)$ является согласованным разбиением.

Утверждение 2 доказано.

Утверждение 3.

Для любой КС-грамматики, удовлетворяющей требованиям 1-5 и 6, существует целое число $chis(G)$, $chis(G) > join(G)$ такое, что для любой пары нетерминалов C и D $t_{n/C} \not\approx t_{n/D}$ при $n > chis(G)$.

Доказательство. Предположим обратное. Тогда для некоторой пары нетерминалов A и B имеет место (1), и в силу утверждения 2 в грамматике G существует согласованное разбиение нетерминалов, что противоречит условиям утверждения.

Из доказанного утверждения 3 вытекает главное

Утверждение 4.

Пусть (а) $m = main(G)$ и $i \geq chis(G)$ для КС-грамматики G , удовлетворяющей требованиям 1-5 и 6,

(б) Q - множество p -элементов глубины $0, 1, \dots, m$ из $T_{m+i/B}$, тогда два p -элемента t' и t'' из Q относятся к одному и тому же нетерминалу тогда и только тогда, когда

$$[t' \approx t'']_{dp(i) \rightarrow \{U\}}. \quad (2)$$

Проще говоря, для выявления одноименных нетерминалов по известной сети $T_{m+i/B}$ достаточно ограничиться изоморфизмом p -элементов, пренебрегая конкретным представлением вершин.

10. ЗАКЛЮЧЕНИЕ

Из приведенных рассуждений следует, что по известной π -сети \mathcal{T} , которую хотелось бы рассматривать как структуру известной части языка $L(G)$ [6], можно построить правила грамматики G . Сказанное справедливо если G – КС-грамматика в виде, удовлетворяющем требованиям 1-5 и 6, если известно число $i = chis(G)$, и если $[\mathcal{T} \approx t_{n/S}]_{dp(i) \rightarrow \{U\}}$ для некоторого $n \geq main(G) + i$.

Если удалось подтвердить все три "если", то для построения правил КС-грамматики достаточно воспользоваться замечаниями 5.3 и 5.4, при этом одноименными следует считать p -сети, для которых справедливо (2). В противном случае можно построить несколько комплектов правил для разных значений параметра i и проверить три "если" для каждого комплекта. Те комплекты правил, которые пройдут проверку, могут считаться правдоподобными.

Вопрос о происхождении сетей \mathcal{T} в этой работе не рассматривался.

СПИСОК ЛИТЕРАТУРЫ

1. Ахо А., Ульман Дж. *Теория синтаксического анализа, перевода и компиляции*. М.: Мир, 1978, тт. 1,2.
2. Соловьев С.Ю. Эквивалентные преобразования контекстно-свободных грамматик. *Информационные процессы*, 2010, том 10, No.3, стр. 292-302.
3. Агафонов В.Н. О методах разбора "по текущему символу". В кн. [1]. М.: Мир, 1978, том 1, стр.408-420.
4. Яблонский С.В. *Введение в дискретную математику*. М.: Наука, 1986.
5. Paull M.C., Unger S.H. Structural Equivalence of Context-Free Grammars. *Journal of Computer and System Sciences*, 1968, vol. 2, pp. 427-463.
6. Соловьев С.Ю. *Методы восстановления контекстно-свободных грамматик*. М.: МГУ, 1981.