Информационные процессы, Том 12, № 4, 2012, стр. 408–437.

ИНФОРМАЦИОННОЕ ВЗАИМОДЕЙСТВИЕ

Задача совместимости свойств формальных грамматик

В.А.Серебряков*, С.Ю.Соловьев**

**Вычислительный центр РАН, Москва, Россия*
***МГУ имени М.В.Ломоносова, факультет ВМК, Москва, Россия*
Поступила в редколлегию 25.09.2012

Аннотация—Рассматривается формальная постановка задачи существования контекстно-свободных грамматик, обладающая заданными набором свойств. Сформулированы достаточные условия разрешимости задачи. Исследованы шесть классов и девять свойств, для которых установлена возможность их совмещения в одной грамматике.

1. ВВЕДЕНИЕ

В известной монографии Ахо и Ульмана [1] упоминаются более шестидесяти свойств/характеристик формальных грамматик; грамматики бывают “автоматные”, “без e-правил”, “общего вида” и т.д. В широком смысле под свойством грамматики понимается ее принадлежность заранее определенному классу. После выхода монографии прошло 40 лет, наука шагнула далеко вперед, и количество изученных классов увеличилось кратно. Исподволь, но настойчиво стала заявлять о себе проблема совместимости в одной грамматике нескольких свойств. Скажем, левая факторизация способна породить цепные правила вывода [1], а устранение цепных правил вывода может потребовать повторной левой факторизации, и т.д. Есть ли выход из этого круга? Можно ли надеяться на существование грамматик, не имеющих цепных правил и не требующей левой факторизации? Этот конкретный вопрос для пары свойств подводит к постановке общей задачи совместимости.

Задача совместимости свойств. *Для всех грамматик из класса* Γ *доказать существование в том же классе* Γ *эквивалентных грамматик, удовлетворяющих заранее перечисленным свойствам.*

В настоящей работе задача совместимости свойств рассматривается как задача существования, которая не обязана иметь алгоритм решения.

2. МОДЕЛЬ СОВМЕСТИМОСТИ СВОЙСТВ

В соответствии с терминологией алгебраических систем [3] будем рассматривать модели $M =< \Gamma; =, ..., >$, где Γ – основное множество, на котором определено симметричное, рефлексивное и транзитивное отношение эквивалентности $=$. Каждый элемент $G \in \Gamma$ входит в единственный класс эквивалентности $[G] \stackrel{def}{=} \{G' \in \Gamma \mid G' = G\}$. Совокупность классов эквивалентности называется фактор-множеством [3] и обозначается $\Gamma/=$. По отношению к фактор-множеству все подмножества из Γ подразделяются на экстенсионалы свойств и подвиды.

Для основного множества Γ подмножество Γ_{ext} называется экстенсионалом свойства (экстенсионалом), если Γ_{ext} имеет общие элементы со всеми классами эквивалентности:

$$\forall\, E \in \Gamma/= \quad \text{выполняется} \quad E \cap \Gamma_{ext} \neq \emptyset; \tag{1}$$

Для основного множества Γ подмножество Γ_{sub} называется подвидом, если

$$\exists\, E \in \Gamma/= \quad \text{такой, что} \quad E \cap \Gamma_{sub} \;=\; \emptyset.$$

С целью унификации терминологии в дальнейшем будем полагать, что основное множество Γ также является подвидом.

На рисунке 1 основное множество Γ изображено в виде прямоугольника, а классы эквивалентности – в виде его полос. Экстенсионал свойства Γ_{ext} и подвид Γ_{sub} изображены прямоугольниками со скошенными углами.

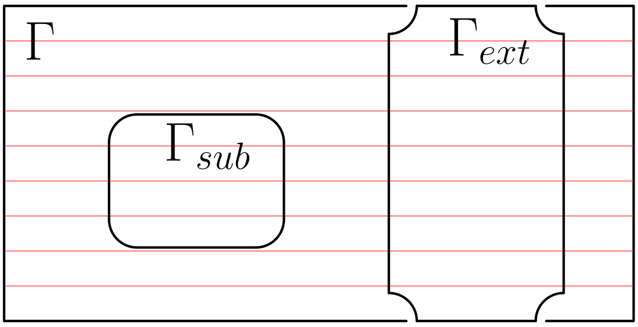

Рис. 1. Подвид Γ_{sub} и экстенсионал свойства Γ_{ext}

Констатирующая часть задачи совместимости свойств формулируется в виде модели $M =< \Gamma; =, \Gamma_1, ..., \Gamma_m >$, где

Γ – основное множество, элементы которого будем называть грамматиками;
= – отношение эквивалентности;
m – фиксированное число, $m \geq 0$;
Γ_i – экстенсионалы свойств для Γ, $i = 1, ..., m$;
иногда на местах свойств будем писать $\ddot{\Gamma}_i$, что означает $\Gamma \cap \Gamma_i$.

Отдельно взятое подмножество Γ_i интерпретируется как совокупность грамматик основного множества, удовлетворяющих некоторому свойству. Считается, что (за пределами рассматриваемой модели) каждое свойство зафиксировано в виде некоторого одноместного предиката.

Совокупность экстенсионалов свойств $\Gamma_1, ..., \Gamma_m$ позволяет определить подмножество грамматик, одновременно удовлетворяющих всем свойствам модели M:

$$sp(M) \stackrel{def}{=} \Gamma \cap \Gamma_1 \cap ... \cap \Gamma_m.$$

Определение. Задача совместимости свойств имеет решение для модели M, если подмножество $E = sp(M)$ удовлетворяет условию (1). // Иными словами, задача совместимости свойств имеет решение, если для любой грамматики G из Γ найдется эквивалентная ей грамматика $G' \in sp(M)$, для которой

$$G' \in \Gamma_1, \quad G' \in \Gamma_2, \quad ..., \quad G' \in \Gamma_m.$$

Как следует из определения $sp(M)$, задача совместимости свойств имеет решение при $m = 0$. Кроме того, справедливо

Утверждение 1. *Если для модели* $< \Gamma; =, \Gamma_1, ..., \Gamma_m >$ *задача совместимости свойств имеет решение, то для любой модели* $< \Gamma; =, \Gamma_i^{(1)}, ..., \Gamma_i^{(k)} >$, *где* $\{\Gamma_i^{(1)}, ..., \Gamma_i^{(k)}\} \subseteq \{\Gamma_1, ..., \Gamma_m\}$, *задача совместимости свойств также имеет решение.*

Определение. Экстенсионал Γ_1 в модели $< \Gamma; =, \Gamma_1, ..., >$ называется фиктивным, если $\Gamma \subseteq \Gamma_1$. // Добавление в модель или исключение из модели фиктивных экстенсионалов не сказывается на существовании решения задачи совместимости свойств.

Остановимся на некоторых других достаточных условиях существования решения. Будем рассматривать экстенсионалы независимо друг от друга, а их способности к сосуществованию "передоверим" некоторому функционалу $\Upsilon : \Gamma \to \mathbb{N}$, который каждой грамматике основного множества сопоставляет натуральное число. В дальнейшем функционалы такого рода будем называть кросс–функционалами. Основное множество Γ с фиксированным кросс–функционалом Υ, будем обозначать $\Gamma \diamond \Upsilon$.

Свойства первого рода. Будем говорить, что в модели $< \Gamma \diamond \Upsilon; =, \Gamma_1, ... >$ подмножество Γ_1 является экстенсионалом свойства первого рода, если
(1) Γ_1 является экстенсионалом свойства; и
(2) для любой грамматики $G \in \Gamma \backslash \Gamma_1$ существует грамматика $G' \in [G]$ такая, что $\Upsilon(G) > \Upsilon(G')$.

Свойства второго рода. Будем говорить, что в модели $< \Gamma \diamond \Upsilon; =, \Gamma_1, ... >$ подмножество Γ_1 является экстенсионалом свойства второго рода, если
(1) Γ_1 является экстенсионалом свойства; и
(2) для любой грамматики $G \in \Gamma \backslash \Gamma_1$ существует грамматика $G' \in [G]$ такая, что $\Upsilon(G) \geq \Upsilon(G')$ и $G' \in \Gamma_1$.

Утверждение 2. *Задача совместимости свойств имеет решение для модели*

$$M =< \Gamma \diamond \Upsilon; =, \Gamma_1, ..., \Gamma_m >,$$

если
– подмножества $\Gamma_1, ..., \Gamma_{m-1}$ являются экстенсионалами свойств первого рода, а
– подмножество Γ_m является экстенсионалом свойства второго рода.

При $m = 1$ утверждение 2 очевидно. Для доказательства утверждения при $m > 1$ предположим обратное:

$$\exists\, G_1 \in \Gamma \quad \text{такая, что} \quad [G_1] \cap sp(M) = \emptyset. \tag{2}$$

Методом математической индукции покажем, что тогда в классе эквивалентности $[G_1]$ найдется бесконечная последовательность грамматик

$$G_1,\ G_2,\ G_3,\ \ldots, \tag{3}$$

порождающая монотонно убывающую последовательность чисел $\Upsilon(G_1),\ \Upsilon(G_2),\ \Upsilon(G_3),\ \ldots$.

Базис индукции. Очевидно $G_1 \in [G_1]$.

Шаг индукции. Пусть $\{ G_1, ..., G_n \} \subseteq [G_1]$ и $\Upsilon(G_1) > ... > \Upsilon(G_n)$ для фиксированного $n \geq 1$. Из $G_n \cap sp(M) = \emptyset$ следует, что $G_n \in \Gamma \setminus \Gamma_i$ для некоторого i. При этом возможны два случая: $i < m$ или $i = m$.

Если $i < m$, то Γ_i есть экстенсионал свойства первого рода, и для G_n существует эквивалентная ей грамматика G_{n+1} такая, что $\Upsilon(G_n) > \Upsilon(G_{n+1})$.

Если $i = m$, то Γ_i есть экстенсионал свойства второго рода, и для G_n существует эквивалентная ей грамматика G'_n такая, что $\Upsilon(G_n) \geq \Upsilon(G'_n)$ и $G'_n \in \Gamma_m$.
Поскольку $G'_n \in [G_n] = [G_1]$, то согласно (2) имеем: $G'_n \in \Gamma \backslash \Gamma_j$ для некоторого $j \neq m$.
А следовательно, Γ_j есть экстенсионал свойства первого рода, и для G'_n существует эквивалентная ей грамматика G_{n+1} такая, что $\Upsilon(G'_n) > \Upsilon(G_{n+1})$.

Откуда $\Upsilon(G_n) \geq \Upsilon(G'_n) > \Upsilon(G_{n+1})$.

Итак, из факта существования первых n членов последовательности (3) следует существование всех последующих ее членов. Поскольку бесконечной монотонно убывающей последовательности натуральных чисел не существует, то предположение (2) ложно.

Утверждение 2 доказано, а из доказательства последовательно вытекают следующие утверждения 3, 4, 5 и 6.

Утверждение 3. *Задача совместимости свойств имеет решение для модели*

$$M =< \Gamma\diamond\Upsilon; =, \Gamma_1, ..., \Gamma_m >,$$

если подмножества $\Gamma_1, ..., \Gamma_m$ являются экстенсионалами свойств первого рода.

Утверждение 4. *Если в модели $< \Gamma\diamond\Upsilon; =, \Gamma_1 >$ подмножество Γ_1 является экстенсионалом свойства первого рода, то для любой грамматики $G \in \Gamma \setminus \Gamma_1$ существует грамматика $G' \in [G]$ такая, что $\Upsilon(G) > \Upsilon(G')$ и $G' \in \Gamma_1$.*

Утверждение 5. *Если в модели $< \Gamma\diamond\Upsilon; =, \Gamma_1, ..., \Gamma_m >$ подмножества $\Gamma_1, ..., \Gamma_m$ есть экстенсионалы свойств первого рода, то $\Gamma_1, ..., \Gamma_m$ являются также экстенсионалами свойств второго рода. (Обратное неверно.)*

Утверждение 6. *Если для некоторого кросс-функционала Υ для модели $< \Gamma\diamond\Upsilon; =, \Gamma_1, ..., \Gamma_m >$ задача совместимости свойств разрешима, то задача совместимости свойств разрешима для модели $< \Gamma; =, \Gamma_1, ..., \Gamma_m >$.*

В дальнейшем изложении:

- в качестве основных множеств рассматриваются КС-грамматики [1] и некоторые их подвиды ▷▷ раздел 3
- эквивалентность грамматик понимается стандартно ▷▷ определение 2
- исследуются девять конкретных свойств грамматик ▷▷ разделы 5–7
- кросс–функционал фиксируется для всех моделей. ▷▷ раздел 4

3. МНОЖЕСТВО КОНТЕКСТНО-СВОБОДНЫХ ГРАММАТИК

Определение 1. Контекстно-свободной грамматикой (КС–грамматикой) называется четверка $G =< N, \Sigma, P, S >$, где

N – алфавит нетерминальных символов (нетерминалов);
Σ – непересекающийся с N алфавит терминальных символов (терминалов);
P – конечное множество правил вывода вида $A \to \alpha$,
где $A \in N$, и α – цепочка символов из $(N \cup \Sigma)^*$;
S – выделенный символ из N, именуемый начальным символом.

В последующих выкладках будем также полагать действующими соглашениях о символах, правилах и выводимости.

Соглашения об использовании символов:

A, B, C, S – нетерминальные символы из N, причем S – начальный символ;
a, b, c, d – терминальные символы из Σ^*;
X, Y, Z – либо терминальные, либо нетерминальные символы из $N \cup \Sigma$;

$\alpha, \beta, \gamma, \sigma$ – цепочки символов из $(N \cup \Sigma)^*$;
$x,\ y,\ z$ – цепочки символов из Σ^* (предложения);
$|\alpha|$ – количество символов в цепочке α (длина цепочки α);
e – цепочка нулевой длины (пустая цепочка).

Соглашения о правилах вывода:
○ правило $A \to \alpha$ называется A-правилом;
○ $R(G, A) \overset{def}{=} \{\alpha \,|\, A \to \alpha \in P\}$ – множество альтернатив нетерминала A;
○ $A \to \alpha_1 \mid ... \mid \alpha_n$ есть совокупность правил $A \to \alpha_1,\ ...,\ A \to \alpha_n$.

Соглашения о выводимости:
запись $\alpha \Rightarrow_G \beta \iff$ "цепочка β непосредственно выводима из цепочки α",
если $\alpha = \alpha_1 A \alpha_2,\quad \beta = \alpha_1 \gamma \alpha_2$ и $A \to \gamma \in P$;
запись $\alpha \Rightarrow_G^* \beta \iff$ "вывод цепочки β из цепочки α",
если $\alpha = \beta$ или $\alpha \Rightarrow_G \alpha_1 \Rightarrow_G \alpha_2 \Rightarrow_G ...\alpha_n \Rightarrow_G \beta$;
дерево вывода есть представление вывода $A \Rightarrow_G^* \beta$ в виде графа [1];
$L(G, \alpha) \overset{def}{=} \{\, x \,|\, \alpha \Rightarrow_G^* x \,\}$;
$L(G) \overset{def}{=} L(G, S)$ – язык, порождаемый грамматикой G.

Принятые соглашения позволяют в некоторых случаях задавать грамматики перечислением правил вывода P. При этом считается, что S – основной символ, $N = \{\, A \,|\, A \to \alpha \in P \,\}$, а множество Σ состоит из символов a, b, c, d, встречающихся в правых частях правил вывода.

Определение 2. КС-грамматики G и G' называются эквивалентными, если они порождают один и тот же язык:

$$G = G' \quad \iff \quad L(G) = L(G').$$

3.1. Наименования правил и символов

Будем полагать известной КС–грамматику $G =< N, \Sigma, P, S >$. Определим некоторые разновидности правил из P и символов из $N \cup \Sigma$.

Правило вывода $A \to \alpha$ *называется*
– e–правилом, если $\alpha = e$;
– цепным правилом, если $\alpha \in N$.
Грамматика G_X из приложения А содержит два e–правила $A \to e$ и $D \to e$, а также два цепных правила $S \to C$, $A \to D$.

Символ X *называется* бесполезным, если в грамматике G, не существует вывода $S \Rightarrow_G^* xAz \Rightarrow_G^* xyz$. Другими словами:
– терминал a является бесполезным, если a не входит ни в одно предложение из $L(G)$; а
– нетерминал A является бесполезным, если в G не существует вывода $S \Rightarrow_G^* xAz \Rightarrow_G^+ xyz$.

Введем ряд специальных обозначений для КС-грамматик.

Для фиксированного нетерминала A, множество $R(G, A)$ которого состоит из альтернатив $\alpha_1, ..., \alpha_n$, *обозначим*:

$$prefix(G, A) \overset{def}{=} argmax\{|\alpha| : \alpha\beta_1 = \alpha_1, ..., \alpha\beta_n = \alpha_n,\ \beta_1 \neq e, ..., \beta_n \neq e\},$$

$$suffix(G, A) \overset{def}{=} argmax\{|\alpha| : \beta_1\alpha = \alpha_1, ..., \beta_n\alpha = \alpha_n,\ \beta_1 \neq e, ..., \beta_n \neq e\}.$$

В грамматике G_A из раздела 3.3 $prefix(G_A, A) = c,\ suffix(G_A, B) = dd$.

Для фиксированного X из $N \cup \Sigma$ *обозначим* $LD(X)$ подмножество нетерминалов из N, удовлетворяющее двум условиям:
(1) $X \notin LD(X)$;
(2) $\forall A \in LD(X)$ все A–правила имеют вид
либо $A \to X\alpha$ и $\alpha \neq e$,
либо $A \to B\alpha$ и $B \in LD(X)$.

В грамматике G^L из приложения В в качестве $LD(X)$ могут использоваться $LD(C) = \{B\}$, $LD(B) = \{A\}$ или $LD(C) = \{A, B\}$.

Аналогично для фиксированного X из $N \cup \Sigma$ *обозначим* $RD(X)$ подмножество нетерминалов из N, удовлетворяющее двум условиям:
(1) $X \notin RD(X)$;
(2) $\forall A \in RD(X)$ все A–правила имеют вид
либо $A \to \alpha X$ и $\alpha \neq e$,
либо $A \to \alpha B$ и $B \in RD(X)$.

Нетерминал A *называется*
– простым, если в грамматике G имеется ровно одно A–правило;
– рекурсивным[1], если в грамматике G существует вывод $A \Rightarrow_G^* \alpha A \beta$;
– нерекурсивным, если в грамматике G не существует вывод $A \Rightarrow_G^* \alpha A \beta$;
– ЛНФ–нетерминалом[2], если $prefix(G, A) \neq e$;
– ПНФ–нетерминалом[3], если $suffix(G, A) \neq e$;
– делимым слева, если $A \in LD(X)$ для некоторых X и $LD(X)$;
– делимым справа, если $A \in RD(X)$ для некоторых X и $RD(X)$.

Пара нетерминалов A и B *называется* нетерминалами–синонимами, если $L(G, A) = L(G, B)$. При этом нетерминал A называется синонимом нетерминала B, а нетеринал B – синонимом нетерминала A.

3.2. Подвиды КС–грамматик

Определение 3. КС–грамматика $G =< N, \Sigma, P, S >$ называется КС#грамматикой, если
– в P отсутствуют e-правила; и
– в P имеется правило $S \to \#$, причем терминал # в других правилах не встречается.

В КС#грамматике G, порождающей более одного предложения, основной символ S
– не является простым;
– не является рекурсивным;
– не является ЛНФ–нетерминалом;
– не является ПНФ–нетерминалом;
– не является делимым слева нетерминалом;
– не является делимым справа нетерминалом;
– не имеет синонимов в G.

КС#грамматики представляют интерес в качестве “почти эквивалентного” представления КС–грамматик общего вида. Связь между КС– и КС#грамматиками устанавливает очевидное

Утверждение 7. *Для любой КС–грамматики $G =< N, \Sigma, P, S >$, не использующей символ #, существует такая КС#грамматика $G_1 =< N \cup \{ S_1 \}, \Sigma \cup \{\#\}, P_1, S_1 >$, что*

$$L(G) \setminus \{e\} = L(G_1) \setminus \{\#\}.$$

[1] простой рекурсивный нетерминал является бесполезным символом.
[2] то есть нетерминалом, допускающим левую неукорачивающую факторизацию
[3] то есть нетерминалом, допускающим правую неукорачивающую факторизацию

Грамматика G_1 из утверждения 7 имеет вид:

$$P_1 = \{\, S_1 \to S,\, S_1 \to \# \,\} \ \cup \bigcup_{A \to \alpha \in P} \{\, A \to \beta \mid \beta \in K(\alpha) \,\} \ \setminus \ \{\, A \to e \mid A \in N \,\},$$

$$\text{где} \ \ K(\alpha) = \begin{cases} K(\alpha_1)\, \{e,\, B\}\, K(\alpha_2), & \text{если } \alpha = \alpha_1 B \alpha_2 \ \text{ и } \ B \Rightarrow^{+}_{G} e; \\ \alpha & \text{иначе.} \end{cases}$$

В рекурсивном определении множества цепочек $K(\alpha)$ используется операция конкатенации трех языков: языка $K(\alpha_1)$, языка $\{e, B\}$ и языка $K(\alpha_2)$. Заметим, что переход от G к G_1 выполняется однозначно, а обратный переход возможен только при наличии априорной информации о выполнимости или невыполнимости соотношения $e \in L(G)$. Строго говоря, при обратном переходе можно построить не исходную грамматику G, а ее эквивалент – грамматику $G_2 =< N \cup \{\, S_1 \,\},\, \Sigma,\, P_2,\, S_1 >$, где

$$P_2 = \begin{cases} \{\, S_1 \to e \,\} \ \cup \ P_1 \ \setminus \ \{\, S_1 \to \# \,\}, & \text{если } e \in L(G); \\ P_1 \ \setminus \ \{\, S_1 \to \# \,\} & \text{иначе.} \end{cases}$$

Соотношения между КС– и КС#грамматиками представлены на рисунке 2.

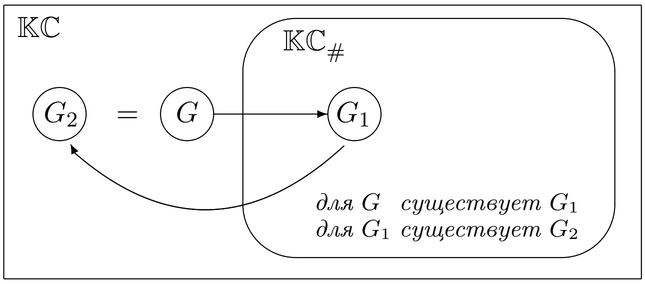

Рис. 2. КС– и КС# грамматики

Определение 4. КС–грамматика G называется однозначной КС–грамматикой, если для каждого предложения из $L(G)$ существует ровно одно дерево вывода.

Определение 5. КС#грамматика G называется однозначной КС#грамматикой, если для каждого предложения из $L(G)$ существует ровно одно дерево вывода.

Существование единственного дерева вывода равносильно существованию единственного левого и единственного правого выводов [1] предложения.

Определение 6. КС–грамматика $G =< N, \Sigma, P, S >$ называется разделенной КС–грамматикой, если для каждого $A \in N$ все его альтернативы начинаются различными терминальными символами.

Определение 7. КС#грамматика $G =< N, \Sigma, P, S >$ называется разделенной КС#грамматикой, если для каждого $A \in N$ все его альтернативы начинаются различными терминальными символами.

В качестве основных множеств формальных моделей выберем шесть подвидов КС–грамматик:

(1) $\mathbb{KC}$ – множество КС–грамматик общего вида, удовлетворяющих определению 1;

(2) $\mathbb{KC}^{\text{O}}$ – множество однозначных КС–грамматик, удовлетворяющих определению 4;

(3) $\mathbb{KC}^{\text{P}}$ – множество разделенных КС–грамматик, удовлетворяющих определению 6.

(4) $\mathbb{KC}_{\#}$ – множество КС#грамматик, удовлетворяющих определению 3;

(5) $\mathbb{KC}^{\text{O}}_{\#}$ – множество однозначных КС#грамматик, удовлетворяющих определению 5;

(6) $\mathbb{KC}^{\text{P}}_{\#}$ – множество разделенных КС#грамматик, удовлетворяющих определению 7.

Соотношения между подвидами грамматик представлены на рисунке 3.

$$\mathbb{KC}^{\text{P}} \subset \mathbb{KC}^{\text{O}} \subset \mathbb{KC}$$

$$\mathbb{KC}^{\text{P}}_{\#} \subset \mathbb{KC}^{\text{O}}_{\#} \subset \mathbb{KC}_{\#} \subset \mathbb{KC}$$

$$\mathbb{KC}^{\text{O}}_{\#} = \mathbb{KC}_{\#} \cap \mathbb{KC}^{\text{O}}$$

$$\mathbb{KC}^{\text{P}}_{\#} = \mathbb{KC}_{\#} \cap \mathbb{KC}^{\text{P}}$$

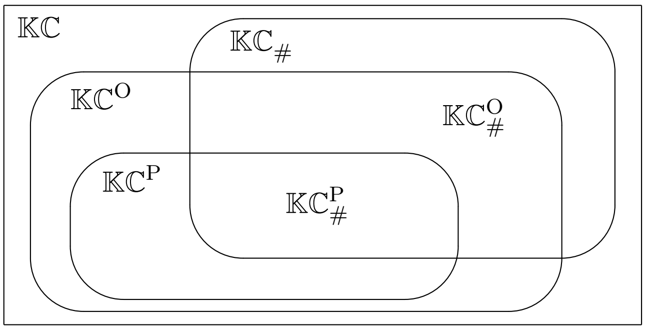

Рис. 3. Подвиды КС–грамматик

Соотношения между подвидами грамматик представлены на рисунке 3.

3.3. Кросс–функционал Υ_3

Пусть $G =< N, \Sigma, P, S >$ – КС–грамматика. Положим

$$\Upsilon_3(G) \overset{def}{=} \|N\| + \|\Sigma\| + \sum_{A \in N} \Lambda(G, A),$$

где $\|N\|$ и $\|\Sigma\|$ – количество символов в алфавитах N и Σ, а

$$\Lambda(G, A) = \begin{cases} 1 + min\{\, |x| :\ x \in L(G, A)\, \}, & \text{если } L(G, A) \neq \emptyset \\ 0 & \text{иначе.} \end{cases}$$

Например, для грамматики

$$\begin{array}{llcrcl} G_A : & S & \to & aaA & | & bbB \\ & A & \to & cAd & | & cd \\ & B & \to & cBdd & | & cdd \end{array}$$

имеем: $N = \{\, S,\, A,\, B\, \}$ и $\|N\| = 3$,
$\Sigma = \{\, a,\, b,\, c,\, d\, \}$ и $\|\Sigma\| = 4$,
$min\{\, |x| :\ x \in L(G_A, B)\, \} = 3$,
$min\{\, |x| :\ x \in L(G_A, A)\, \} = 2$,
$min\{\, |x| :\ x \in L(G_A, S)\, \} = 4$, и окончательно:

$$\Upsilon_3(G_A) = 3 + 4 + (1 + 3) + (1 + 2) + (1 + 4) = 19.$$

Заметим, что в [2] с аналогичной целью использовался более “простой” кросс–функционал $h(G)$.

Приведем ряд простых утверждений, устанавливающих связь между значениями кросс–функционала для КС-грамматик $G_1 =< N_1, \Sigma_1, P_1, S_1 >$ и $G_2 =< N_2, \Sigma_2, P_2, S_2 >$. Во всех утверждениях будем предполагать существование отображения $\varphi\colon N_2 \to N_1$.

Утверждение 8. *Если* $\|N_1 \cup \Sigma_1\| > \|N_2 \cup \Sigma_2\|$ *и* $\forall A \in N_2$ *выполняется равенство* $L(G_2, A) = L(G_1, \varphi(A))$, *то* $\Upsilon_3(G_1) > \Upsilon_3(G_2)$.

Утверждение 9. *Если* $\|N_1 \cup \Sigma_1\| \geq \|N_2 \cup \Sigma_2\|$ *и* $\forall A \in N_2$ *выполняется равенство* $L(G_2, A) = L(G_1, \varphi(A))$, *то* $\Upsilon_3(G_1) \geq \Upsilon_3(G_2)$.

Утверждение 10. *Если*
–а– $N_1 = N_2$ *и* $\Sigma_1 = \Sigma_2$; *и*
–б– φ *– взаимно однозначное отображение; и*
–в– существует непустое подмножество нетерминалов $N_0 \subseteq N_1$; *и*
–г– существует язык L_0, *для которого* $min\{\,|x| : x \in L_0\,\} \geq 1$; *и*
–д– $\forall B \notin N_0$ *выполняется равенство* $L(G_1, B) = \quad L(G_2, \varphi^{-1}(B))$; *и*
–е– $\forall B \in N_0$ *выполняется равенство* $L(G_1, B) = L_0\, L(G_2, \varphi^{-1}(B))$ *см.*[4],
то $\Upsilon_3(G_1) > \Upsilon_3(G_2)$.

4. СВОЙСТВА КС–ГРАММАТИК

Приведем свойства грамматик, существенные для порождения бесконечных КС–языков. Перечень этих свойств, определенных через экстенсионалы, включает девять позиций:

$\Gamma^{\text{без}}_{\text{ЦП}}$ – КС—грамматики, которые не содержат цепных правил;

$\Gamma^{\text{без}}_{\text{БС}}$ – КС—грамматики, которые не содержат бесполезных символов;

$\Gamma^{\text{без}}_{\text{НР}}$ – КС—грамматики, которые не содержат нерекурсивных нетерминалов[5];

$\Gamma^{\text{без}}_{\text{СН}}$ – КС—грамматики, которые не содержат нетерминалов–синонимов;

$\Gamma^{\text{без}}_{\text{ПР}}$ – КС#грамматики, которые не содержат простых нетерминалов;

$\Gamma^{\text{без}}_{\text{ЛФ}}$ – КС#грамматики, которые не содержат ЛНФ–нетерминалов;

$\Gamma^{\text{без}}_{\text{ПФ}}$ – КС#грамматики, которые не содержат ПНФ–нетерминалов;

$\Gamma^{\text{без}}_{\text{ДЛ}}$ – КС#грамматики, которые не содержат делимых–слева нетерминалов;

$\Gamma^{\text{без}}_{\text{ДП}}$ – КС#грамматики, которые не содержат делимых–справа нетерминалов.

Хотя приведенный перечень не является исчерпывающим, он вполне пригоден для иллюстрации нетривиального решения задачи совместимости свойств.

Исследование перечисленных экстенсионалов вида $\Gamma^{\text{без}}_{\text{XX}}$ сводится к доказательству некоторого утверждения–о–существовании в заданном подвиде Γ грамматик с указанными свойствами. Типичное утверждение о существовании имеет вид:

Если $G \in \Gamma \setminus \Gamma^{\textit{без}}_{XX}$,
то существует грамматика $G' \in [G] \in \Gamma/\!=$,
для которой $\Upsilon_3(G) > \Upsilon_3(G')$ *или* $\Upsilon_3(G) \geq \Upsilon_3(G')$.

Типичное доказательство утверждения о существовании распадается на шесть этапов.

Этап 1. Сопоставить грамматике G из $\Gamma \setminus \Gamma^{\text{без}}_{\text{XX}}$ некоторую грамматику G'.
Этап 2. Доказать $G = G'$.
Этап 3. Доказать (если это необходимо) $G' \in \{\,\mathbb{KC}, \mathbb{KC}_\#\,\}$.
Этап 4. Доказать (если это необходимо) $G' \in \{\,\mathbb{KC}^{\text{O}}, \mathbb{KC}^{\text{O}}_\#\,\}$.
Этап 5. Доказать (если это необходимо) $G' \in \{\,\mathbb{KC}^{\text{P}}, \mathbb{KC}^{\text{P}}_\#\,\}$.
Этап 6. Доказать $\Upsilon_3(G) > \Upsilon_3(G')$ или $\Upsilon_3(G) \geq \Upsilon_3(G')$.

[4] В равенстве –е– порядок конкатенирующих языков L_0 и $L(G_2, \varphi^{-1}(B))$ может быть изменен на обратный.
[5] За исключением, быть может, основного символа.

В общем виде этап 1 выглядит как некоторое преобразование $G' = f(G, \chi)$, использующее дополнительную информацию $\chi = \chi(G)$. Часто χ без труда вычисляется по известной грамматике G, но в отдельных случаях – см., например, раздел 4.4 – не имеет универсального алгоритма вычисления. Это обстоятельство позволяет говорить о существовании грамматики G', но не предоставляет алгоритма перехода от G к G'.

В дальнейшем изложении для описания этапов 1 потребуется операция одновременной замены в цепочке α подцепочек β на γ (считается $\beta \neq e$). Результат операции есть цепочка $\alpha[\beta /\!/ \gamma]$:

$$\alpha\,[\beta /\!/ \gamma] \stackrel{def}{=} \begin{cases} \alpha_0\gamma\alpha_1', & \text{если } \alpha = \alpha_0\beta\alpha_1 \ (\alpha_0 \text{ – минимальной длины) и } \alpha_1' = \alpha_1[\beta /\!/ \gamma]; \\ \alpha, & \text{иначе.} \end{cases}$$

Для совокупности замен определим

$$\alpha\,[\beta_1 /\!/ \gamma_1] \ldots [\beta_k /\!/ \gamma_k] \stackrel{def}{=} \Big(\alpha\,[\beta_1 /\!/ \gamma_1] \ldots [\beta_{k-1} /\!/ \gamma_{k-1}]\Big)\,[\beta_k /\!/ \gamma_k].$$

Если, например, $\alpha = A + B * A$, то $\alpha\,[A /\!/ aA] = aA + B * aA$,
$\alpha\,[A /\!/ aA]\,[B /\!/ bA] = \Big(aA + B * aA\Big)\,[B /\!/ bA] = aA + bA * aA$,
$\alpha\,[B /\!/ bA]\,[A /\!/ aA] = \Big(A + bA * aA\Big)\,[A /\!/ aA] = aA + baA * aaA$.

Зафиксируем также стандартную аргументацию этапов 2–5, равноприменимую ко всем рассматриваемым свойствам грамматик.

Этап 2 $\gg$ Доказательство приводится в [1, 4] или в ином источнике.
Этап 3 $\gg$ Доказательство очевидно.
Этап 4 $\gg$ Доказательство следует из взаимно однозначного соответствия между выводами предложений в G и G'.
Этап 5 $\gg$ Доказательство следует из особенностей этапа 1.

Наличие стандартной аргументации позволяет приводить в описаниях свойств только их уникальные особенности.

4.1. Грамматики без цепных правил $\Gamma^{\textit{без}}_{\textit{ЦП}}$

В качестве основного множества Γ рассматривается один из четырех подвидов КС–грамматик:

$$\Gamma \in \{\, \mathbb{KC},\, \mathbb{KC}^{\mathrm{O}},\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\mathrm{O}}_{\#} \,\}. \tag{4}$$

По определению множество $\Gamma^{\text{без}}_{\text{ЦП}}$ образуют КС–грамматики, не содержащие цепных правил. $\ggg$ Если $G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ЦП}}$, то в грамматике G имеется некоторое количество цепных правил[6].

Утверждение 11. *Если* Γ *удовлетворяет (4) и* $G \in \Gamma \setminus \Gamma^{\textit{без}}_{\textit{ЦП}}$, *то существует грамматика* $G' \in [G] \in \Gamma/{=}$, *для которой* $\Upsilon_3(G) \geq \Upsilon_3(G')$ *и* $G' \in \Gamma^{\textit{без}}_{\textit{ЦП}} \cap \Gamma$.

Грамматика G' из утверждения 11 определяется как $< N, \Sigma, P', S >$, причем

$$P' = \{\, B \to \alpha \,|\, A \to \alpha \in P,\ \alpha \notin N,\ B \Rrightarrow^* A \,\}, \quad \text{где} \quad B \Rrightarrow A \iff B \to A \in P.$$

Соотношение $\Upsilon_3(G) \geq \Upsilon_3(G')$ следует из утверждения 9 и из равенств $L(G, A) = L(G', A)$ для $A \in N$.

[6] $\chi(G)$ – множество цепных правил грамматики G.

$\boxed{\textit{Вывод}}$ Если подвид Γ удовлетворяет условию (4), то в модели $< \Gamma\diamond\Upsilon_3;\, =,\, \ddot{\Gamma}^{\text{без}}_{\text{ЦП}},\, ... >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{ЦП}}$ является экстенсионалом свойства второго рода.

4.2. Грамматики без бесполезных символов $\Gamma^{\textit{без}}_{\textit{БС}}$

В качестве основного множества Γ рассматривается один из шести подвидов КС–грамматик:

$$\Gamma \in \{\, \mathbb{KC},\, \mathbb{KC}^{\text{O}},\, \mathbb{KC}^{\text{P}},\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\text{O}}_{\#},\, \mathbb{KC}^{\text{P}}_{\#} \,\} \tag{5}$$

По определению множество $\Gamma^{\text{без}}_{\text{БС}}$ образуют КС–грамматики, не содержащие бесполезных символов. Каждой КС–грамматике $G =< N,\, \Sigma,\, P,\, S >$ поставим в соответствие КС–грамматику $G_+ =< N_+,\, \Sigma_+,\, P_+,\, S >$, где

$$P_+ = \{\, A \to \alpha \in \widehat{P} \mid S \Rrightarrow^* A \,\}, \quad \text{где} \quad \widehat{P} = \{\, A \to \alpha \in P \mid \alpha \Rightarrow^*_G x \,\} \ \text{ и } \ B \Rrightarrow C \iff B \to \beta C \gamma \in \widehat{P},$$
$$N_+ = \{\, A \in N \mid A \to \alpha \in P_+ \,\},$$
$$\Sigma_+ = \{\ a \in \Sigma \mid A \to \beta a \gamma \in P_+ \,\}.$$

Множество $\Gamma^{\text{без}}_{\text{БС}}$ образуют те и только те грамматики $G =< N, \Sigma, P, S >$, для которых $N = N_+$ и $\Sigma = \Sigma_+$. $\ggg$ Если $G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{БС}}$, то $N_+ \cup \Sigma_+ \subset N \cup \Sigma$.

Утверждение 12. *Если* Γ *удовлетворяет (5) и* $G \in \Gamma \setminus \Gamma^{\textit{без}}_{\textit{БС}}$, *то* $G_+ \in [G] \in \Gamma/=$, $\Upsilon_3(G) > \Upsilon_3(G')$, *а кроме того* $G_+ \in \Gamma^{\textit{без}}_{\textit{БС}} \cap \Gamma$.

Неравенство $\Upsilon_3(G) > \Upsilon_3(G')$ вытекает из утверждения 8, поскольку

$$N_+ \cup \Sigma_+ \subset N \cup \Sigma \qquad \text{и} \qquad P_+ \subseteq P.$$

$\boxed{\textit{Вывод}}$ Если подвид Γ удовлетворяет условию (5), то в модели $< \Gamma\diamond\Upsilon_3;\, =,\, \ddot{\Gamma}^{\text{без}}_{\text{БС}},\, ... >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{БС}}$ является экстенсионалом свойства первого рода.

4.3. Грамматики без нерекурсивных нетерминалов $\Gamma^{\textit{без}}_{\textit{НР}}$

В качестве основного множества Γ рассматривается один из шести подвидов КС–грамматик:

$$\Gamma \in \{\, \mathbb{KC},\, \mathbb{KC}^{\text{O}},\, \mathbb{KC}^{\text{P}},\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\text{O}}_{\#},\, \mathbb{KC}^{\text{P}}_{\#} \,\} \tag{6}$$

По определению множество $\Gamma^{\text{без}}_{\text{НР}}$ образуют КС–грамматики $< N, \Sigma, P, S >$, в которых все нетерминалы из $N \backslash \{S\}$ являются рекурсивными. $\ggg$ Если $G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{НР}}$, то в грамматике G найдется нерекурсивный нетерминал $\dot{A}$, отличный от S.

Утверждение 13. *Если* Γ *удовлетворяет (6) и* $G \in \Gamma \backslash \Gamma^{\textit{без}}_{\textit{НР}}$, *то существует грамматика* $G' \in [G] \in \Gamma/=$, *для которой* $\Upsilon_3(G) > \Upsilon_3(G')$.

Грамматика G' из утверждения 13 определяется как $< N',\, \Sigma,\, P',\, S >$, где

$$N' = N \setminus \{\dot{A}\} \qquad \text{и} \qquad P' = \bigcup_{\alpha \in R(G, \dot{A})} \{\, B \to \beta[\dot{A} /\!\!/ \alpha] :\ B \neq \dot{A},\ B \to \beta \in P \,\}$$

Неравенство $\Upsilon_3(G) > \Upsilon_3(G')$ следует из утверждения 8, поскольку $N' \subset N$ и $L(G, B) = L(G', B)$ для $B \in N'$.

Заметим, что в соответствии с утверждением 4 в КС#грамматиках можно полностью избавиться от нерекурсивных нетерминалов:

$$\forall\, G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{НР}} \qquad [G] \cap \Gamma^{\text{без}}_{\text{НР}} \cap \Gamma \neq \emptyset.$$

$\boxed{\textit{Вывод}}$ Если подвид Γ удовлетворяет условию (6), то в модели $< \Gamma\diamond\Upsilon_3;\, =,\, \ddot{\Gamma}^{\text{без}}_{\text{НР}},\, ... >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{НР}}$ является экстенсионалом свойства первого рода.

4.4. Грамматики без синонимов $\Gamma^{\textit{без}}_{\textit{CH}}$

В качестве основного множества Γ рассматривается один из шести подвидов КС–грамматик:

$$\Gamma \in \{\, \mathbb{KC},\, \mathbb{KC}^{\mathrm{O}},\, \mathbb{KC}^{\mathrm{P}},\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\mathrm{O}}_{\#},\, \mathbb{KC}^{\mathrm{P}}_{\#} \,\} \tag{7}$$

Определение 8. КС–грамматика $G =< N, \Sigma, P, S >$ называется КС–грамматикой без синонимов, если выполняется одно из двух: либо $N = \{S\}$, либо для любой пары различных нетерминалов A и B выполняется неравенство $L(G, A) \neq L(G, B)$. // По определению множество $\Gamma^{\text{без}}_{\text{CH}}$ образуют КС–грамматики без синонимов.

В приложении А (раздел А.4) доказывается

Утверждение 14. *Если* Γ *удовлетворяет (7) и* $G \in \Gamma \setminus \Gamma^{\textit{без}}_{\textit{CH}}$*, то существует грамматика* $G' \in [G] \in \Gamma/{=}$*, для которой* $\Upsilon_3(G) > \Upsilon_3(G')$*.*

Вывод Если подвид Γ удовлетворяет условию (7), то в модели $< \Gamma\diamond\Upsilon_3;\, =,\, \ddot{\Gamma}^{\text{без}}_{\text{CH}},\, \ldots >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{CH}}$ является экстенсионалом свойства первого рода.

4.5. Грамматики без простых нетерминалов $\Gamma^{\textit{без}}_{\textit{ПР}}$

В качестве основного множества Γ рассматривается один из трех подвидов КС–грамматик:

$$\Gamma \in \{\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\mathrm{O}}_{\#},\, \mathbb{KC}^{\mathrm{P}}_{\#} \,\} \tag{8}$$

По определению множество $\Gamma^{\text{без}}_{\text{ПР}}$ образуют КС#грамматики, в которых отсутствуют простые нетерминалы. $\ggg$ Если $G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ПР}}$, то в грамматике G найдется нетерминал $\dot{A}$, отличный от S, для которого все множество $\dot{A}$–правил исчерпывается одним правилом $\dot{A} \to \dot{\alpha}$.

Утверждение 15. *Если* Γ *удовлетворяет (8) и* $G \in \Gamma \setminus \Gamma^{\textit{без}}_{\textit{ПР}}$*, то существует грамматика* $G' \in [G] \in \Gamma/{=}$*, для которой* $\Upsilon_3(G) > \Upsilon_3(G')$*.*

Грамматика G' из утверждения 15 определяется как $< N', \Sigma, P', S >$, где

$$N' = N \setminus \{\dot{A}\} \qquad \text{и} \qquad P' = \begin{cases} \{\, B \to \beta \in P \;\;:\; B \neq \dot{A},\; \beta \neq \beta_0 \dot{A} \beta_1 \,\}, & \text{если } \dot{\alpha} = \alpha_0 \dot{A} \alpha_1; \\ \{\, B \to \beta[\dot{A} /\!\!/ \dot{\alpha}] :\; B \neq \dot{A},\; B \to \beta \in P \,\} & \text{иначе.} \end{cases}$$

Неравенство $\Upsilon_3(G) > \Upsilon_3(G')$ следует из утверждения 8, поскольку $N' \subset N$ и $L(G, B) = L(G', B)$ для $B \in N'$.

Заметим, что в соответствии с утверждением 4 в КС–грамматиках можно полностью избавиться от простых нетерминалов:

$$\forall\, G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ПР}} \qquad [G] \cap \Gamma^{\text{без}}_{\text{ПР}} \cap \Gamma \neq \emptyset.$$

Вывод Если подвид Γ удовлетворяет условию (8), то в модели $< \Gamma\diamond\Upsilon_3;\, =,\, \ddot{\Gamma}^{\text{без}}_{\text{ПР}},\, \ldots >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{ПР}}$ является экстенсионалом свойства первого рода.

4.6. Грамматики без ЛНФ-нетерминалов $\Gamma^{\textit{без}}_{\textit{ЛФ}}$

В качестве основного множества Γ рассматривается один из двух подвидов КС#грамматик:

$$\Gamma \in \{\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\mathrm{O}}_{\#} \,\} \tag{9}$$

По определению множество $\Gamma^{\text{без}}_{\text{ЛФ}}$ образуют КС#грамматики, в которых отсутствуют ЛНФ–нетерминалы. $\ggg$ Если $G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ЛФ}}$, то в грамматике G существует нетерминал $\dot{A}$, отличный от S, для которого $\dot{\alpha} = prefix(G, \dot{A}) \neq e$. В [2] установлено

Утверждение 16. *Если* Γ *удовлетворяет (9) и* $G \in \Gamma \setminus \Gamma^{\textit{без}}_{\textit{ЛФ}}$*, то существует грамматика* $G' \in [G] \in \Gamma/\!=$*, для которой* $\Upsilon_3(G) > \Upsilon_3(G')$.

Упомянутая в утверждении 16 грамматика G' имеет вид $< N, \Sigma, P', S >$, где

$$P' = \begin{cases} \{\, B\to\beta \in P \quad :\ B \neq \dot{A},\ \beta \neq \beta_0 \dot{A} \beta_1 \,\}, & \text{если } \dot{\alpha} = \alpha_0 \dot{A} \alpha_1; \\ \{\, B\to\beta[\dot{A}/\!\!/\dot{\alpha}\dot{A}] :\ B \neq \dot{A},\ B\to\beta \in P \,\} \ \cup\ \{\, \dot{A}\to\gamma[\dot{A}/\!\!/\dot{\alpha}\dot{A}] :\ \dot{A}\to\dot{\alpha}\gamma \in P \,\} & \text{иначе.} \end{cases}$$

Неравенство $\Upsilon_3(G) > \Upsilon_3(G')$ следует из утверждения 10, случай $N_0 = \{\dot{A}\}$, $L_0 = \{\, x \mid \dot{\alpha} \Rightarrow^*_G x \,\}$.

Заметим, что в соответствии с утверждением 4 в КС#грамматиках можно полностью избавиться от ЛНФ–нетерминалов:

$$\forall\, G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ЛФ}} \qquad [G] \cap \Gamma^{\text{без}}_{\text{ЛФ}} \cap \Gamma \neq \emptyset.$$

$\boxed{\textit{Вывод}}$ Если подвид Γ удовлетворяет условию (9), то в модели $< \Gamma \diamond \Upsilon_3;\ =,\ \ddot{\Gamma}^{\text{без}}_{\text{ЛФ}},\ \ldots >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{ЛФ}}$ является экстенсионалом свойства первого рода.

4.7. Грамматики без ПНФ-нетерминалов $\Gamma^{\textit{без}}_{\textit{ПФ}}$

В качестве основного множества Γ рассматривается один из трех подвидов КС–грамматик:

$$\Gamma \in \{\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\mathrm{O}}_{\#},\, \mathbb{KC}^{\mathrm{P}}_{\#} \,\} \tag{10}$$

По определению множество $\Gamma^{\text{без}}_{\text{ПФ}}$ образуют КС#грамматики, в которых отсутствуют ПНФ–нетерминалы. $\ggg$ Если $G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ПФ}}$, то в грамматике G найдется нетерминал $\dot{A}$, отличный от S, для которого $\dot{\alpha} = suffix(G, \dot{A}) \neq e$. В [2] установлено

Утверждение 17. *Если* Γ *удовлетворяет (10) и* $G \in \Gamma \setminus \Gamma^{\textit{без}}_{\textit{ПФ}}$*, то существует грамматика* $G' \in [G] \in \Gamma/\!=$*, для которой* $\Upsilon_3(G) > \Upsilon_3(G')$.

Упомянутая в утверждении 17 грамматика G', в которой нетерминал $\dot{A}$ не является ПНФ–нетерминалом, имеет вид $< N, \Sigma, P', S >$, где

$$P' = \begin{cases} \{\, B\to\beta \in P \quad :\ B \neq \dot{A},\ \beta \neq \beta_0 \dot{A} \beta_1 \,\}, & \text{если } \dot{\alpha} = \alpha_0 A \alpha_1; \\ \{\, B\to\beta[\dot{A}/\!\!/\dot{A}\dot{\alpha}] :\ B \neq \dot{A},\ B\to\beta \in P \,\} \ \cup\ \{\, A\to\gamma[\dot{A}/\!\!/\dot{A}\dot{\alpha}] :\ A\to\gamma\dot{\alpha} \in P \,\} & \text{иначе.} \end{cases}$$

Неравенство $\Upsilon_3(G) > \Upsilon_3(G')$ следует из утверждения 10, случай $N_0 = \{\dot{A}\}$, $L_0 = \{\, x \mid \dot{\alpha} \Rightarrow^*_G x \,\}$.

Заметим, что в соответствии с утверждением 4 в КС#грамматиках можно полностью избавиться от ПНФ–нетерминалов:

$$\forall\, G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ПФ}} \qquad [G] \cap \Gamma^{\text{без}}_{\text{ПФ}} \cap \Gamma \neq \emptyset.$$

$\boxed{\textit{Вывод}}$ Если подвид Γ удовлетворяет условию (9), то в модели $< \Gamma \diamond \Upsilon_3;\ =,\ \ddot{\Gamma}^{\text{без}}_{\text{ПФ}},\ \ldots >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{ПФ}}$ является экстенсионалом свойства первого рода.

4.8. Грамматики без делимых слева нетерминалов $\Gamma^{\textit{без}}_{\textit{ДЛ}}$

В качестве основного множества Γ рассматривается один из двух подвидов КС#грамматик:

$$\Gamma \in \{\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\mathrm{O}}_{\#} \,\} \tag{11}$$

По определению множество $\Gamma^{\text{без}}_{\text{ДЛ}}$ образуют КС#грамматики, в которых отсутствуют делимые слева нетерминалы. В приложении В доказывается[7]

Утверждение 18. *Если* Γ *удовлетворяет (11) и* $G \in \Gamma \backslash \Gamma^{\textit{без}}_{\textit{ДЛ}}$, *то существует грамматика* $G_W \in [G] \in \Gamma/\!=$, *для которой* $\Upsilon_3(G) > \Upsilon_3(G_W)$.

Заметим, что в соответствии с утверждением 4 в КС#грамматиках можно полностью избавиться от делимых слева нетерминалов:

$$\forall\, G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ДЛ}} \qquad [G] \cap \Gamma^{\text{без}}_{\text{ДЛ}} \cap \Gamma \neq \emptyset.$$

$\boxed{\textit{Вывод}}$ Если подвид Γ удовлетворяет условию (11), то в модели $< \Gamma\diamond\Upsilon_3;\, =,\, \ddot{\Gamma}^{\text{без}}_{\text{ДЛ}},\, ... >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{ДЛ}}$ является экстенсионалом свойства первого рода.

4.9. Грамматики без делимых справа нетерминалов $\Gamma^{\textit{без}}_{\textit{ДП}}$

В качестве основного множества Γ рассматривается один из трех подвидов КС–грамматик:

$$\Gamma \in \{\, \mathbb{KC}_{\#},\, \mathbb{KC}^{\mathrm{O}}_{\#},\, \mathbb{KC}^{\mathrm{P}}_{\#} \,\} \tag{12}$$

По определению множество $\Gamma^{\text{без}}_{\text{ДП}}$ образуют КС#грамматики, в которых отсутствуют делимые справа нетерминалы. В приложении В доказывается[8]

Утверждение 19. *Если* Γ *удовлетворяет (12) и* $G \in \Gamma \backslash \Gamma^{\textit{без}}_{\textit{ДП}}$, *то существует грамматика* $G_V \in [G] \in \Gamma/\!=$, *для которой* $\Upsilon_3(G) > \Upsilon_3(G_V)$.

Заметим, что в соответствии с утверждением 4 в КС#грамматиках можно полностью избавиться от делимых справа нетерминалов:

$$\forall\, G \in \Gamma \setminus \Gamma^{\text{без}}_{\text{ДП}} \qquad [G] \cap \Gamma^{\text{без}}_{\text{ДП}} \cap \Gamma \neq \emptyset.$$

$\boxed{\textit{Вывод}}$ Если подвид Γ удовлетворяет условию (12), то в модели $< \Gamma\diamond\Upsilon_3;\, =,\, \ddot{\Gamma}^{\text{без}}_{\text{ДП}},\, ... >$ подмножество $\ddot{\Gamma}^{\text{без}}_{\text{ДП}}$ является экстенсионалом свойства первого рода.

5. НОРМАЛЬНЫЕ ФОРМЫ КС-ГРАММАТИК

Результаты предыдущего раздела сводятся в единую таблицу.

[7] Утверждения 39 – 41.
[8] Утверждения 42 – 44.

Экстенсионал	Род	Основные множества					
		$\mathbb{KC}$	$\mathbb{KC}^{\text{O}}$	$\mathbb{KC}^{\text{P}}$	$\mathbb{KC}_{\#}$	$\mathbb{KC}^{\text{O}}_{\#}$	$\mathbb{KC}^{\text{P}}_{\#}$
$\Gamma^{\text{без}}_{\text{ЦП}}$	2	+	+	ф_1	+	+	ф_2
$\Gamma^{\text{без}}_{\text{БС}}$	1	+	+	+	+	+	+
$\Gamma^{\text{без}}_{\text{НР}}$	1	+	+	+	+	+	+
$\Gamma^{\text{без}}_{\text{СН}}$	1	+	+	+	+	+	+
$\Gamma^{\text{без}}_{\text{ПР}}$	1				+	+	+
$\Gamma^{\text{без}}_{\text{ЛФ}}$	1				+	+	ф_3
$\Gamma^{\text{без}}_{\text{ПФ}}$	1				+	+	+
$\Gamma^{\text{без}}_{\text{ДЛ}}$	1				+	+	ф_4
$\Gamma^{\text{без}}_{\text{ДП}}$	1				+	+	+

В таблице существенные свойства отмечены знаком $+$. В четырех случаях, отмеченных символами ф_i, свойства основных множеств оказываются фиктивными: $\mathbb{KC}^{\text{P}} \subseteq \Gamma^{\text{без}}_{\text{ЦП}}$, $\mathbb{KC}^{\text{P}}_{\#} \subseteq \Gamma^{\text{без}}_{\text{ЦП}}$, $\mathbb{KC}^{\text{P}}_{\#} \subseteq \Gamma^{\text{без}}_{\text{ЛФ}}$ и $\mathbb{KC}^{\text{P}}_{\#} \subseteq \Gamma^{\text{без}}_{\text{ДЛ}}$. В соответствии с утверждением 2 все полученные результаты сводятся в

Утверждение 20. *Задача совместимости свойств имеет решение для моделей*

$$
\begin{aligned}
&< \mathbb{KC};\ \ =, \Gamma^{\textit{без}}_{\textit{ЦП}}, \Gamma^{\textit{без}}_{\textit{БС}},\ \Gamma^{\textit{без}}_{\textit{НР}},\ \Gamma^{\textit{без}}_{\textit{СН}} >,\\
&< \mathbb{KC}^{O}; =, \ddot{\Gamma}^{\textit{без}}_{\textit{ЦП}}, \ddot{\Gamma}^{\textit{без}}_{\textit{БС}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{НР}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{СН}} >,\\
&< \mathbb{KC}^{P}; =, \ddot{\Gamma}^{\textit{без}}_{\textit{ЦП}}, \ddot{\Gamma}^{\textit{без}}_{\textit{БС}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{НР}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{СН}} >,\\
&< \mathbb{KC}_{\#}; =, \ddot{\Gamma}^{\textit{без}}_{\textit{ЦП}}, \ddot{\Gamma}^{\textit{без}}_{\textit{БС}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{НР}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{СН}},\ \Gamma^{\textit{без}}_{\textit{ПР}},\ \Gamma^{\textit{без}}_{\textit{ЛФ}},\ \Gamma^{\textit{без}}_{\textit{ПФ}},\ \Gamma^{\textit{без}}_{\textit{ДЛ}},\ \Gamma^{\textit{без}}_{\textit{ДП}} >,\\
&< \mathbb{KC}^{O}_{\#}; =, \ddot{\Gamma}^{\textit{без}}_{\textit{ЦП}}, \ddot{\Gamma}^{\textit{без}}_{\textit{БС}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{НР}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{СН}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ПР}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ЛФ}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ПФ}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ДЛ}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ДП}} >,\\
&< \mathbb{KC}^{P}_{\#}; =, \ddot{\Gamma}^{\textit{без}}_{\textit{ЦП}}, \ddot{\Gamma}^{\textit{без}}_{\textit{БС}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{НР}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{СН}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ПР}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ЛФ}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ПФ}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ДЛ}},\ \ddot{\Gamma}^{\textit{без}}_{\textit{ДП}} > .
\end{aligned}
$$

В более привычном виде это утверждение можно переформулировать так:

Утверждение 21. *Для любой КС–грамматики G существует эквивалентная ей КС–грамматика G_2, которая не содержат (a) цепных правил, (b) бесполезных символов, (c) нерекурсивных нетерминалов и (d) простых нетерминалов.*

Утверждение 22. *Для любой однозначной КС–грамматики G существует эквивалентная ей однозначная КС–грамматика G_2, которая не содержат (a) цепных правил, (b) бесполезных символов, (c) нерекурсивных нетерминалов и (d) простых нетерминалов.*

Утверждение 23. *Для любой разделенной КС–грамматики G существует эквивалентная ей разделенная КС–грамматика G_2, которая не содержат (a) цепных правил, (b) бесполезных символов, (c) нерекурсивных нетерминалов и (d) простых нетерминалов.*

Утверждение 24. *Для любой КС#грамматики G существует эквивалентная ей КС#грамматика G_2, которая не содержат (a) цепных правил, (b) бесполезных символов, (c) нерекурсивных нетерминалов, (d) простых нетерминалов, (e) нетерминалов–синонимов, (f) ЛНФ–нетерминалов, (g) ПНФ–нетерминалов, (h) делимых–слева нетерминалов, (i) делимых–справа нетерминалов.*

Утверждение 25. *Для любой однозначной КС#грамматики G существует эквивалентная ей однозначная КС#грамматика G_2, которая не содержат (a) цепных правил, (b) бесполезных символов, (c) нерекурсивных нетерминалов, (d) простых нетерминалов, (e) нетерминалов–синонимов, (f) ЛНФ–нетерминалов, (g) ПНФ–нетерминалов, (h) делимых–слева нетерминалов, (i) делимых–справа нетерминалов.*

Утверждение 26. *Для любой разделенной КС#грамматики G существует эквивалентная ей разделенная КС#грамматика G_2, которая не содержат (a) цепных правил, (b) бесполезных символов, (c) нерекурсивных нетерминалов, (d) простых нетерминалов, (e) нетерминалов–синонимов, (f) ЛНФ–нетерминалов, (g) ПНФ–нетерминалов, (h) делимых–слева нетерминалов, (i) делимых–справа нетерминалов.*

6. ЗАКЛЮЧЕНИЕ

Проведенное исследование показывает, что алгебраический подход к задаче совместимости свойств грамматик оказывается вполне плодотворным. Вместе с тем, определяющую роль в этом подходе принадлежит кросс-функционалу Υ. Для каждого нестандартного набора свойств кросс–функционал необходимо заново сконструировать и исследовать.

ПРИЛОЖЕНИЕ

A. КС–ГРАММАТИКИ БЕЗ СИНОНИМОВ

Как следует из определения 8, КС–грамматика G не является КС–грамматикой без синонимов, если для некоторых A и B из N, $A \neq B$, выполняется равенство $L(G, A) = L(G, B)$. С практической точки зрения интерес представляют только такие пары нетерминалов синонимов, для которых

$$L(G, A) = L(G, B) \neq \emptyset \tag{A.1}$$

Покажем, что для каждого КС–языка существует порождающая его КС–грамматика без синонимов. С этой целью докажем возможность удаления из КС–грамматике одного нетерминала из известной пары синонимов. Конструктивная часть доказательства связана со специальными преобразованиями деревьев вывода.

A.1. Дополнительная терминология деревьев вывода

Зафиксируем алфавит терминальных символов Σ, и будем рассматривать деревья с помеченными вершинами. В качестве меток могут использоваться как символы из Σ, так и другие символы, именуемые нетерминальными символами. Формальными деревьями [вывода] будем считать такие помеченные деревья, в которых:
– концевые вершины (листья; терминальные вершины) помечены символами из Σ;
– все внутренние (нетерминальные) вершины помечены нетерминальными символами;
– на множестве непосредственных потомков каждой нетерминальной вершины зафиксирован порядок их перечисления.

Общий для нескольких грамматик алфавит терминальных символов Σ, определяет класс формальных деревьев, некоторые из которых являются деревьями вывода в соответствующих грамматиках, а другие исполняют роли "переходных" форм.

Пусть $A \Rightarrow_G^+ x$ – вывод некоторого предложения x в грамматике $G =< N, \Sigma, P, S >$. Этому выводу однозначно соответствует дерево вывода $DT(A \Rightarrow_G^+ x)$. В левой части рисунка 4 приводится одно из возможных деревьев вывода предложения acc в грамматике

$$
\begin{array}{lllcccccc}
G_X : & S & \to & & & C & \mid & AD \\
 & A & \to & e & \mid & a & \mid & D \\
 & D & \to & e & \mid & b & \mid & DC \\
 & C & \to & & & c & \mid & AC
\end{array}
$$

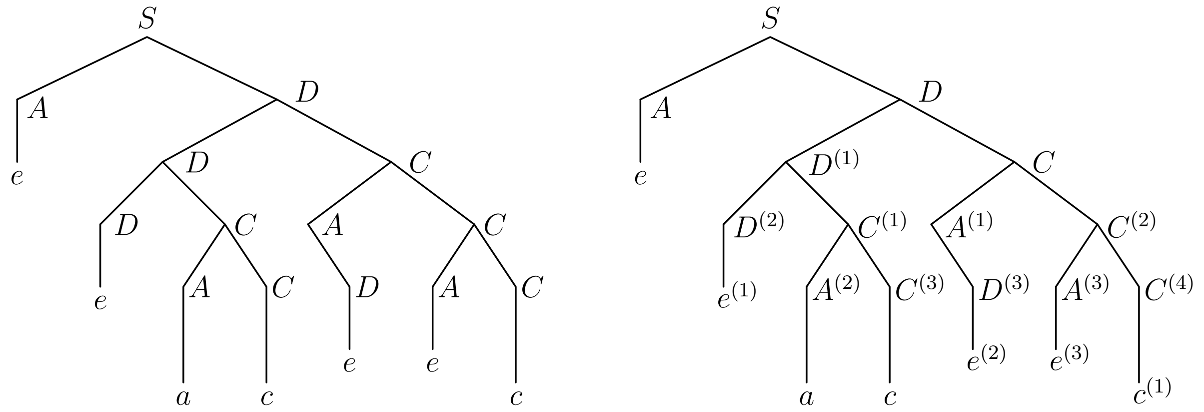

Рис. 4. Пример дерева вывода T_{acc}

Вершину дерева вывода, помеченную символом X, будем обозначать тем же символом X, а в случае коллизии меток будем различать вершины с помощью верхних индексов: $X^{(индекс)}$ и т.д. В правой части рисунка 4 приводится дерево вывода, в котором учтено соглашение об использовании верхних индексов.

Для известного дерева вывода $T = DT(A \Rightarrow_G \beta \Rightarrow_G^* \alpha)$ будем обозначать:
$v(T)$ – корень $A^{(x)}$ дерева T;
$s(T)$ – цепочку символов α, соответствующую меткам концевых вершин дерева T;
$p(T)$ – связанное с корнем дерева правило вывода $A \to \beta$.

Каждая вершина X дерева вывода T однозначно порождает поддерево $ST(T, X)$, состоящее из самой вершины X и всех ее потомков, а также из связывающих эти вершины ребер. Частным случаем поддерева является само это дерево: $T \equiv ST(T, v(T))$. Для упрощения записи согласимся использовать:
о б о з н а ч е н и е $s(T, X)$ вместо $s(\ ST(T, X)\ )$ и
о б о з н а ч е н и е $p(T, X)$ вместо $p(\ ST(T, X)\ )$.

Будем полагать определенной операцию замены в дереве вывода T поддерева $ST(T, D)$ на заданное дерево вывода T'. Не ограничивая общности можно считать, что в дереве T' все вершины имеют уникальные верхние индексы, не встречающиеся в других деревьях вывода. Результат замены – новое дерево вывода T_x записывается как $T \blacktriangleleft D \lessdot T'$. На рисунке 5 приводится пример замены для дерева T_{acc}, изображенного на рисунке 4. Дерево $T_x = T \blacktriangleleft D \lessdot T'$ является деревом вывода в грамматике G, если T и T' – деревья вывода в G, а вершины D и $v(T')$ помечены одним нетерминалом.

Если T_1, ..., T_n – некоторые формальные деревья, а D_1, ..., D_n – вершины дерева T, не связанные отношениями "предок–потомок", то запись $T \blacktriangleleft D_1 \lessdot T_1 \blacktriangleleft D_2 \lessdot T_2 \ldots \blacktriangleleft D_n \lessdot T_n$

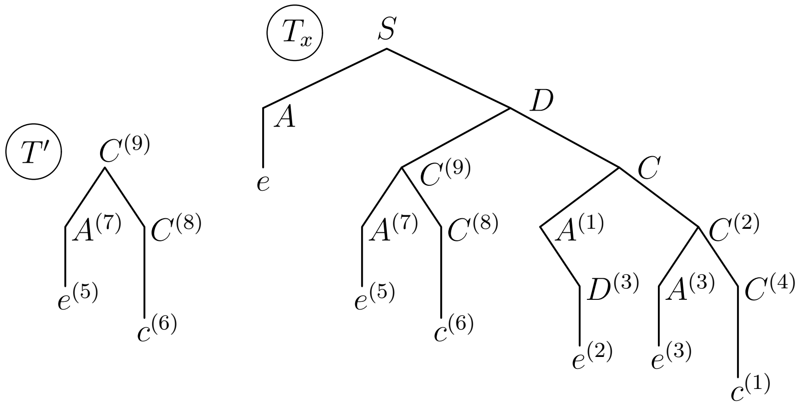

Рис. 5. Пример замены $T_x = T_{acc} \blacktriangleleft D^{(1)} \lessdot T'$

означает формальное дерево вывода, полученное следующим образом:

$$\Big(... \Big((T \blacktriangleleft D_1 \lessdot T_1) \blacktriangleleft D_2 \lessdot T_2 \Big) \; ... \Big) \; \blacktriangleleft D_n \lessdot T_n.$$

Если для некоторой грамматики $G =< N, \Sigma, P, S >$ известно, что $L(G, A) = L(G, B)$, то
$\triangleright$ для любого вывода $A \Rightarrow^+_G x$ существует вывод $B \Rightarrow^+_G x$, и
$\triangleright$ для любого дерева вывода $DT(A \Rightarrow^+_G x)$ существует дерево вывода $DT(B \Rightarrow^+_G x)$.

Если в формальном дереве T имеется вершина $A^{(x)}$ и $ST(T, A^{(x)})$ – есть дерево вывода в грамматике G, то существуют вывод $B \Rightarrow^+_G s(T, A^{(x)})$ и дерево вывода $DT(B \Rightarrow^+_G s(T, A^{(x)}))$, которое будем обозначать как $B_G(T, A^{(x)})$.

Аналогично, определим дерево $A_G(T, B^{(y)})$ как некоторое дерево $DT(A \Rightarrow^+_G s(T, B^{(y)}))$, построенное для вершины $B^{(y)}$ дерева T при условии $B \Rightarrow^+_G s(T, B^{(y)})$. Деревья A_G и B_G есть альтернативные выводы одного и того же предложения в грамматике G.

A.2. Универсальный подход к устранению синонимии

Пусть в КС–грамматике $G =< N, \Sigma, P, S >$ для двух различных нетерминалов A и B выполняется равенство $L(G, A) = L(G, B)$. В настоящем разделе используются следующие обозначения:
$\triangleright$ $AB \stackrel{def}{=} \{ A, B \}$;
$\triangleright$ E – новый нетерминальный символ; по определению $E \notin N$;
$\triangleright$ G_E – КС–грамматика $< N_E, \Sigma, P_E, S >$, в которой

$N_E = \{E\} \cup N \setminus AB$,

$P_E = \{ D[A /\!\!/ E][B /\!\!/ E] \to \alpha[A /\!\!/ E][B /\!\!/ E] : \; D \to \alpha \in P \}$.

Например:

$$\begin{array}{llllllllllll} \overline{G}: & S & \to & S + A & \mid & A * B & \qquad \overline{G}_E: & S & \to & S + E & \mid & E * E \\ & A & \to & a & \mid & a\,B & & E & \to & a & \mid & a\,E \;\mid\; E\,a \\ & B & \to & a & \mid & A\,a & & & & & & \end{array}$$

$$L(\overline{G}, A) = L(\overline{G}, B) = \{ a^n \mid n > 0 \}$$

Утверждение 27. *Для любого* $D \in N \backslash AB$ *имеет место* $L(G, D) \subseteq L(G_E, D)$, *а также* $L(G, B) \subseteq L(G_E, E)$.

Выберем некоторый нетерминал D из $N \backslash \{A\}$, Для доказательства утверждения достаточно заметить, что любое дерево вывода T', полученное из $DT(D \Rightarrow^+_G x)$ заменой меток A и B на метку E, является деревом вывода в грамматике G_E. А если при этом $D = B$, то корень дерева T' имеет метку E.

Для каждого правила $C \to \beta$ из P_E зафиксируем правило–пробраз $D \to \alpha$ из P:

$$D[A /\!\!/ E][B /\!\!/ E] = C,$$
$$\alpha[A /\!\!/ E][B /\!\!/ E] = \beta.$$

Утверждение 28. *Для любого $D \in N \backslash AB$ имеет место $L(G, D) \supseteq L(G_E, D)$, а также $L(G, A) \supseteq L(G_E, E)$.*

Для доказательства выберем
(а) некоторый нетерминал D из N_E,
(б) некоторое предложение x из $L(G_E, D)$ и
(в) некоторое дерево вывода $T = DT(D \Rightarrow^+_{G_E} x)$.
и покажем, что в грамматике G существует дерево вывода $T' = DT(D \Rightarrow^+_G x)$.

Перенумеруем нетерминальные вершины $D_1, D_2, ..., D_q$ дерева T таким образом, что номер каждой нетерминальной вершины превосходит номера ее непосредственных потомков.

Доказательство утверждения 28 связано с поэтапным преобразованием дерева T:

$$T_0 = T, \quad \Big[\text{от } T_0 \text{ к } T_1\Big], \quad \Big[\text{от } T_1 \text{ к } T_2\Big] \quad \text{и т.д.} \quad \Big[\text{от } T_{q-1} \text{ к } T_q\Big], \quad T' = T_q$$

Обозначим $y = s(T_j, D_j)$. На каждом этапе преобразования $\Big[\text{от } T_j \text{ к } T_{j+1}\Big]$ поддерево $ST(T_j, D_j)$ заменяется
либо на некоторое поддерево $DT(A \Rightarrow^+_G y)$, если $D_j = E^{(x)}$;
либо на некоторое поддерево $DT(C \Rightarrow^+_G y)$, если $D_j = C^{(z)}$, $C \neq E$.
Полагая это свойство установленным для всех деревьев $T_1, \dots T_j$, построим дерево T_{j+1}.

(1) Принятый порядок рассмотрения вершин гарантирует, что в дереве T_j для непосредственных потомков вершины D_j –суть– вершин

$$X_1^{(j_1)}, X_2^{(j_2)}, ..., X_n^{(j_n)} \tag{A.2}$$

выполняется одно из двух:
или $X_k^{(j_k)}$ – терминальная вершина, то есть $X_k \in \Sigma$;
или $ST(T_j, X_k^{(j_k)})$ – дерево вывода в грамматике G, причем $X_k \in N \setminus \{B\}$.

(2) В дереве T вершина D_j имеет в качестве непосредственных потомков вершины

$$Y_1^{(j_1)}, Y_2^{(j_2)}, ..., Y_n^{(j_n)},$$

соответственно $p(T, D_j)$ имеет вид $D \to Y_1\, Y_2 \dots Y_n$, где D – метка вершины D_j, $D_j \in N_E$.

(3) Правилу $p(T, D_j)$ соответствует прообраз вида

$$D' \to Z_1\, Z_2 \dots Z_n, \qquad \text{где} \qquad D' \in N.$$

(4) Посимвольное сравнение цепочек $X_1\, X_2 \dots X_n$, $Y_1\, Y_2 \dots Y_n$ и $Z_1\, Z_2 \dots Z_n$ приводит к одному из трех результатов:
(р1) $X_k = Y_k = Z_k$, если $X_k \in \Sigma \cup N \setminus AB$,
(р2) $X_k = A,\ Y_k = E,\ Z_k = A$,
(р3) $X_k = A,\ Y_k = E,\ Z_k = B$.

(5) Обозначим

$\triangleright$ $A^{(k_1)}$, ..., $A^{(k_m)}$ все вершины из списка (A.2), для которых выполняется условие (p3);
$\triangleright$ $T'_{j+1} \ = \ T_j \blacktriangleleft A^{(k_1)} \lessdot B_G(T_j, A^{(k_1)}) \ \blacktriangleleft \ \ldots \ \blacktriangleleft A^{(k_m)} \lessdot B_G(T_j, A^{(k_m)})$;
$\triangleright$ u – уникальный индекс, не встречающийся в деревьях T_j и T'_{J+1};
$\triangleright$ T^A_{j+1} – дерево T'_{j+1}, в котором вершина D_j переименована в $A^{(u)}$;
$\triangleright$ T^B_{j+1} – дерево T'_{j+1}, в котором вершина D_j переименована в $B^{(u)}$.

(6) Окончательно дерево T_{j+1} определяется так:
(v1) если $D' = A$, то $T_{j+1} = T^A_{j+1}$;
(v2) если $D' = B$, то $T_{j+1} = T^B_{j+1} \blacktriangleleft B^{(u)} \lessdot A_G(T^B_{j+1}, B^{(u)})$;
(v3) если $D' \neq AB$, то $T_{j+1} = T'_{j+1}$.

(7) Во всех трех вариантах поддерево, построенное в T_{j+1} на позиции вершины D_j, является деревом вывода в грамматике G.

Описанная процедура построения T_{j+1} вполне применима к дереву T_1, в этом случае все $Y_1^{(j_1)}$, ..., $Y_n^{(j_n)}$ – терминальные вершины.

Полученное в результате преобразования дерево T_q является деревом вывода в грамматике G. Если на последнем этапе применялись варианты (v1) и (v2), то метка $v(T)$ совпадает с меткой $v(T_q)$, откуда вытекает $L(G, D) \supseteq L(G_E, D)$ для $D \in N \backslash AB$. Если на последнем этапе преобразования применялся вариант (v2), то $v(T) = E^{(r)}$, $v(T_q) = A^{(r)}$, и, следовательно, $L(G, A) \supseteq L(G_E, E)$.
Утверждение доказано.

Очевидным образом утверждения 27 и 28 порождают

Утверждение 29. *Для любого $D \in N \backslash AB$ имеет место $L(G, D) = L(G_E, D)$, а также $L(G, B) = L(G_E, E)$.*

По построению грамматика G_E не содержит нетерминалы A и B. Корректно определим переименование в G_E нетерминала E на B. Полученная при таклм преобразовании грамматика имеет вид $G'_E =< N'_E, \Sigma, P'_E, S >$, где

$$N'_E = \begin{cases} \{S\} \cup N_E \setminus \{E\}, & \text{если } S \in AB; \\ \{B\} \cup N_E \setminus \{E\}, & \text{если } S \notin AB; \end{cases}$$

$$P'_E = \begin{cases} \{\, C[E/\!\!/S] \to \alpha[E/\!\!/S] :\ C \to \alpha \in P_E \,\}, & \text{если } S \in AB; \\ \{\, C[E/\!\!/B] \to \alpha[E/\!\!/B] :\ C \to \alpha \in P_E \,\}, & \text{если } S \notin AB. \end{cases}$$

Для грамматики $\overline{G}_E$ из предыдущего примера грамматика $\overline{G}'_E$ выглядит так:

$$\begin{array}{llllll} \overline{G}'_E : & S & \to & S + B & \mid\ B * B & \\ & B & \to & a & \mid\ a\,B & \mid\ B\,a \end{array}$$

Для вновь построенной грамматики G'_E утверждение (29) трансформируется в

Утверждение 30. *Для любой КС–грамматики $G =< N, \Sigma, P, S >$, в которой нетерминалы A и B являются синонимами, найдется грамматика $G'_E =< N'_E, \Sigma, P'_E, S >$ такая, что $L(G, C) = L(G'_E, C)$ для всех C из $N'_E \stackrel{def}{=} N \setminus \{A\}$. (В частности, $L(G) = L(G'_E)$.)*

A.3. Специальный подход к устранению синонимии

Описанное в предыдущем разделе построение

$$\left[\text{от } G \text{ к } G_E\right] \quad \text{и} \quad \left[\text{от } G_E \text{ к } G'_E\right]$$

зачастую не сохраняет однозначность, поэтому аналог утверждения 30 для однозначных грамматик необходимо обосновывать отдельно. С этой целью отметим два обстоятельства.

Первое. Если в грамматике G для пары нетерминалов существует вывод $A \Rightarrow_G^+ B \Rightarrow_G^+ B$, то G – неоднозначная грамматика. Соответственно для однозначной грамматики G выполняется $A \not\Rightarrow_G^+ B$ и/или $B \not\Rightarrow_G^+ A$.

Второе. В дереве вывода $T = DT(A \Rightarrow_G^+ x)$ теоретически возможны три случая:

–1– $|x| = 0$, то есть $x = e$;
–2– $|x| = 1$, и тогда в T существует – см. рисунок 6 – единственный путь от корня A к терминальной вершине x, непосредственным предком которой является некоторая вершина A_r;
–3– $|x| \geq 2$, и тогда в T существует – см. рисунок 6 – единственная вершина A_r, имеющая $k \geq 2$ непосредственных потомков $C_1, \ldots, C_k$, причем $s(T, A_r) = s(T, C_1) \ldots s(T, C_k)$ и $s(T, C_j) \neq e$, $j = 1 \ldots k$.

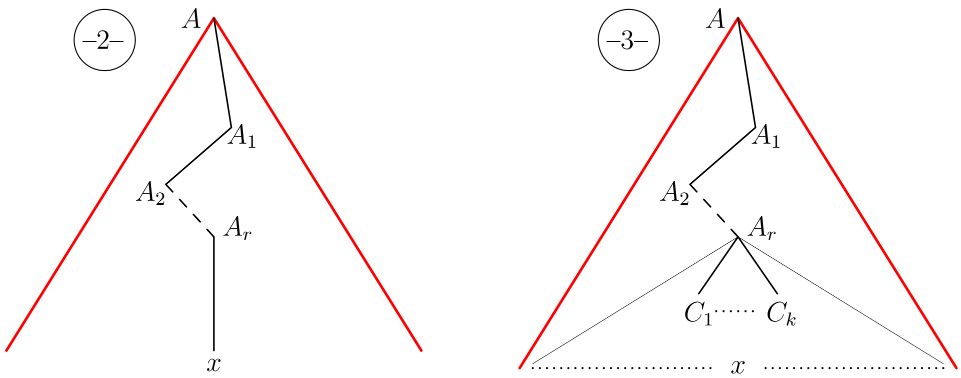

Рис. 6. Варианты деревьев вывода

В случаях –2– и –3– путь, соединяющий A и A_r, состоит из последовательности нетерминальных вершин A, A_1, ..., A_r, которую будем называть стволом дерева T и будем обозначать A+A_1+...+A_r В простейшем случае вершины A и A_r совпадают, и ствол дерева вырождается в корневую вершину A.

В дереве вывода $T = T_{acc}$ (рисунок 4)
случаю –1– соответствуют поддеревья
$ST(T, A)$, $ST(T, A^{(1)})$, $ST(T, A^{(3)})$, $ST(T, D^{(2)})$ и $ST(T, D^{(3)})$;
случаю –2– соответствуют поддеревья $ST(T, A^{(2)})$, $ST(T, C^{(4)})$, а также
$ST(T, C)$ *со стволом* C+$C^{(2)}$+$C^{(4)}$ и $ST(T, C^{(2)})$ *со стволом* $C^{(2)}$+$C^{(4)}$;
случаю –3– соответствуют поддеревья
$ST(T, S)$ *со стволом* S+D, $ST(T, D^{(1)})$ *со стволом* $D^{(1)}$+$C^{(1)}$ и $ST(T, D)$.

Предположим для КС–грамматики $G =< N, \Sigma, P, S >$ выполняются условия

$$L(G, A) = L(G, B) \text{ и } B \not\Rightarrow_G^+ A. \tag{A.3}$$

Определим грамматику G'_B следующим образом:

$$G'_B = \begin{cases} < N'_B,\ \Sigma,\ P'_B,\ S >, & \text{если } S \neq A; \\ < N'_B,\ \Sigma,\ P'_B,\ B > & \text{иначе.} \end{cases} \tag{A.4}$$

где $N'_B = N \setminus \{A\}$ и $P'_B = \{\, C \to \alpha[A/\!\!/B] \;:\; C \to \alpha \in P,\ C \neq A,\ L(G,\alpha) \neq \emptyset \,\}$.
Кроме того, определим – см. рисунок 7 – вспомогательные подмножества правил

$$Q^- = \{\, C \to \alpha \in P \;:\; L(G,\alpha) = \emptyset \,\} \;\cup\; \{\, A \to \alpha \in P \,\}, \qquad Q' = P'_B \setminus P \qquad \text{и}$$

$$Q^+ = \{\, C \to \beta' A \beta'' \in P \;:\; C \neq A,\ L(G, \beta' A \beta'') \neq \emptyset \,\} \qquad .$$

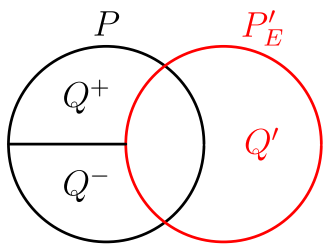

Рис. 7. Правила вывода грамматик G и G'

Множество Q^- образуют все правила грамматики G, которые не участвуют в построении грамматики G'. В общем случае правило из Q^+ содержит $n > 0$ нетерминалов A и имеет вид

$$C \to \alpha_0\, A\, \alpha_1 \,...\, \alpha_{n-1}\, A\, \alpha_n, \tag{A.5}$$

где $\alpha_i \in (\Sigma \cup N')^*$ для $i = 0, ..., n$.

Очевидно $Q' = \{\, C{\to}\alpha[A/\!\!/B] \;:\; C{\to}\alpha \in Q^+ \,\}$. То есть в любом правиле из Q' можно выделить $n > 0$ нетерминалов B, позволяющих представить это правило в виде

$$C \to \alpha_0\, B\, \alpha_1 \,...\, \alpha_{n-1}\, B\, \alpha_n \tag{A.6}$$

таком, что

$$C \to \alpha_0\, A\, \alpha_1 \,...\, \alpha_{n-1}\, A\, \alpha_n \in Q^+ \tag{A.7}$$

Правило (A.7) является прообразом правила (A.6). Наличие в грамматике G двух и более прообразов для одного и того же правила (A.6) означает неоднозначность грамматики G. Вообще говоря, количество нетерминалов B в правиле (A.6) может превосходить число n за счет символов B, "скрывающихся" в цепочках $\alpha_i \in (\Sigma \cup N'_B)^*$. В качестве примера приведем грамматику $\overline{G}'_B$, построенную для пары нетерминалов–синонимов A и B грамматики $\overline{G}$ из предыдущего раздела настоящего приложения.

$$\overline{G}'_B : \qquad \begin{array}{lclcl} S & \to & S + B & | & B * B \\ B & \to & a & | & B\, a \end{array}$$

Связанные с этой грамматикой множества выглядят так:
$Q^- = \{\, A \to a,\ A \to aB \,\}$,
$Q^+ = \{\, S \to S + A,\ S \to A * B,\ B \to Aa \,\}$,
$Q' \;= \{\, S \to S + B,\ S \to S * B,\ B \to Ba \,\}$,
прообразом правила $S \to S + B$ является правило $S \to S + A$,
прообразом правила $S \to S * B$ является правило $S \to A * B$,
прообразом правила $B \to Aa$ является правило $S \to Ba$.

Утверждение 31. *Для любого $D \in N'_B$ имеет место $L(G, D) \subseteq L(G'_B, D)$.*

Выберем
(а) некоторый нетерминал D из N'_B,
(б) некоторое предложение x из $L(G, D)$ и
(в) некоторое дерево вывода $T = DT(D \Rightarrow^+_G x)$
и покажем, что в грамматике G'_B существует дерево вывода $T' = DT(D \Rightarrow^+_{G'_B} x)$.

Поскольку $D \neq A$, то все вершины с метками A в дереве T являются непосредственными потомками некоторых вершин C причем $p(T, C) \in Q^+$.

Доказательство утверждения 31 связано с поэтапным преобразованием дерева T:

$$T_0 = T, \quad \Big[\text{от } T_0 \text{ к } T_1\Big], \quad \Big[\text{от } T_1 \text{ к } T_2\Big] \quad \text{и т.д.} \quad \Big[\text{от } T_{q-1} \text{ к } T_q\Big], \quad T' = T_q$$

Последовательность преобразований завершается, когда в очередном формальном дереве T_q не обнаруживается правил из Q^+. Рассмотрим две части этапа преобразования, состоящего в переходе $\Big[\text{от } T_j \text{ к } T_{j+1}\Big]$.

Часть 1/2. Выберем в дереве T_j вершину C, для которой

$$p(T_j, C) \in Q^+, \quad \text{но} \quad p(T_j, C') \notin Q^+ \quad \text{для всех предков} \quad C' \quad \text{вершины} \quad C. \tag{A.8}$$

Правило $p(T_j, C)$ имеет вид (A.5). Нетерминалам A из этого правила в дереве T_j соответствуют вершины $A^{(1)}, \ \ldots, \ A^{(n)}$ и поддеревья $T^{(1)} = ST(T_j, A^{(1)}), \ \ldots, \ T^{(n)} = ST(T_j, A^{(n)})$. Условие (A.8) гарантирует, что каждое $T^{(i)}$ есть дерево вывода в грамматике G, и ему можно сопоставить[9] дерево $T_1^{(i)} = B_G(T_j, A^{(i)})$. Из условия $B \not\Rightarrow^+_G A$ следует, что
п р и $y = e$ в дереве $T_1^{(i)}$ вообще отсутствуют вершины с метками A; а
п р и $y \neq e$ в дереве $T_1^{(i)}$ вершины $A^{(x)}$ могут находиться только вне стебля, причем для них выполняется неравенство

$$\Big|s(\ ST(T_1^{(i)}, A^{(x)})\ )\Big| \ < \ \Big|s(\ T^{(i)}\ )\Big| \tag{A.9}$$

Часть 2/2. Определим[10] дерево T_{j+1} следующим образом:

$$T_{j+1} = T_j \blacktriangleleft A^{(1)} \lessdot T_1^{(1)} \ \ \ldots \ \ \blacktriangleleft A^{(n)} \lessdot T_1^{(n)}$$

По построению $s(T_j) = s(T_{j+1})$, $p(T_{j+1}, C) \notin Q^+$ и во всех последующих преобразованиях вершина C не участвует.

Конечность процесса преобразования деревьев вывода гарантируют неравенства (A.9). Фактически в процессе преобразования в формальных деревьях правила из Q^+ и заменяются на правила из Q'. Отсюда следует, что полученное[11] при последней модификации дерево T_q является деревом вывода в грамматике G'. Утверждение доказано.

Обозначим $W'(T)$ – подмножество вершин C дерева вывода T, для которых

$$p(T, C) \in Q^+, \quad \text{но} \quad p(T, C') \notin Q^+ \quad \text{для всех потомков} \quad C' \quad \text{вершины} \quad C. \tag{A.10}$$

Утверждение 32. *Для любого $D \in N'_B$ имеет место $L(G, D) \supseteq L(G'_B, D)$.*

Выберем
(а) некоторый нетерминал D из N'_B,

[9] Если G – однозначная грамматика, то $T_1^{(i)}$ определяется однозначно.
[10] Если G – однозначная грамматика, то T_{j+1} по заданной вершине C определяется однозначно.
[11] Если G – однозначная грамматика, то T_q определяется единственным образом.

(б) некоторое предложение x из $L(G'_B, D)$ и
(в) некоторое дерево вывода $T = DT(D \Rightarrow^+_{G'_B} x)$
и покажем, что в грамматике G существует дерево вывода $T' = DT(D \Rightarrow^+_G x)$.

Доказательство утверждения 32 связано с поэтапным преобразованием дерева T:

$$T_0 = T, \quad \Big[\text{от } T_0 \text{ к } T_1\Big], \quad \Big[\text{от } T_1 \text{ к } T_2\Big] \quad \text{и т.д.} \quad \Big[\text{от } T_{q-1} \text{ к } T_q\Big], \quad T' = T_q$$

Преобразование завершается, когда в очередном дереве T_q не оказывается правил из Q'. Рассмотрим один этап преобразования –суть– переход от T_j к T_{j+1}.

Выберем в дереве T_j вершину C из множества $W(T_j)$.

Правило $p(T_j, C)$ имеет вид (A.6) и ему соответствует[12] прообраз – правило вида (A.7). По результатам сравнения правила $p(T_j, C)$ и его прообраза в дереве T_j среди непосредственных потомков вершины C обнаруживаются вершины $B^{(1)}$, ..., $B^{(n)}$, которые соответствуют терминалам A в прообразе. Порядок обхода вершин – условие (A.10) – гарантирует, что каждое поддерево $ST(T_j, B^{(i)})$ есть дерево вывода в грамматике G', что позволяет определить[13] T_{j+1} следующим образом:

$$T_{j+1} = T_j \blacktriangleleft B^{(1)} \lessdot A_G(T_j, B^{(1)}) \ \ \ldots \ \ \blacktriangleleft B^{(n)} \lessdot A_G(T_j, B^{(n)})$$

По построению $s(T_j) = s(T_{j+1})$ и $p(T_{j+1}, C) \in Q^+$.

Конечность процесса преобразования деревьев вывода следует из неравенства

$$\sum_{w \in W(T_j)} d_j(w) \ > \sum_{w \in W(T_{j+1})} d_{j+1}(w),$$

где $d_j(w)$ – расстояние в дереве T_j от корня до вершины w. Полученное при последней модификации дерево T_q является деревом вывода в грамматике G. Утверждение доказано.

Утверждения 31 и 32 порождают

Утверждение 33. *Если КС–грамматика G удовлетворяет условию (A.3), и КС–грамматика G'_B построена способом (A.4), то $L(G, C) = L(G'_B, C)$ для всех C из N'_B. (В частности, $L(G) = L(G'_B)$.)*

Из доказательства утверждений 31 и 32 вытекает

Утверждение 34. *Если однозначная КС–грамматика G удовлетворяет условию (A.3), и КС–грамматика G'_B построена способом (A.4), то G'_B является однозначной КС–грамматикой, причем $L(G, C) = L(G'_B, C)$ для всех C из N'_B. (В частности, $L(G) = L(G'_B)$.)*

Если в КС–грамматике G имеет место $L(G, A) = L(G, B)$ и $A \not\Rightarrow^+_G B$, то для грамматики G'_A, построенной аналогично G'_B, страведливы аналоги утверждений (33) и (34)

A.4. Общий подход к устранению синонимии

Пусть G – КС–грамматика, в которой зафиксированы две нетерминала–синонима A или B. Определим грамматику G':

$$G' \stackrel{def}{=} \begin{cases} G'_B, & \text{если } B \not\Rightarrow^+_G A; & \triangleright \triangleright \text{см. раздел } A.3 \\ G'_A, & \text{если } A \not\Rightarrow^+_G B; & \triangleright \triangleright \text{см. раздел } A.3 \\ G'_E, & \text{если } A \Rightarrow^+_G B \Rightarrow^+_G A. & \triangleright \triangleright \text{см. раздел } A.2 \end{cases}$$

[12] Если G – однозначная грамматика, то прообраз является единственным.
[13] Если G – однозначная грамматика, то T_{j+1} по заданной вершине C определяется однозначно.

Переход от грамматики G, содержащей заданную пару нетерминалов–синонимов, к грамматике G' сохраняет многие видовые отличия грамматик. В частности, нетрудно установить

Утверждение 35.
(1) если G – КС–грамматика, то G' – КС–грамматика;
(2) если G – однозначная КС–грамматика, то G' – однозначная КС–грамматика;
(3) если G – разделенная КС–грамматика, то G' – разделенная КС–грамматика;
(4) если G – КС#грамматика, то G' – КС#грамматика;
(5) если G – однозначная КС#грамматика, то G' – однозначная КС#грамматика;
(6) если G – разделенная КС#грамматика, то G' – разделенная КС#грамматика.
И во всех шести случаях:
— в G' отсутствуют один или более нетерминалов грамматики G;
— для любого нетерминала C грамматики G' выполняется равенство $L(G', C) = L(G, C)$;
— G' и G – эквивалентные грамматики.

Варианты (4) и (5) утверждения 35 объясняются тем, что в КС#грамматиках основной символ не может входить ни в одну пару синонимов. Варианты (3) и (6) очевидны.

А.5. Особенности КС–грамматик без синонимов

Из утверждений 4, 8 и 35 следует

Утверждение 36.
(1) Для любой КС–грамматики G
существует эквивалентная ей КС–грамматика–без–синомимов G''.
(2) Для любой однозначной КС–грамматики G
существует эквивалентная ей однозначная КС–грамматика–без–синомимов G''.
(3) Для любой разделенной КС–грамматики G
существует эквивалентная ей разделенная КС–грамматика–без–синомимов G''.
(4) Для любой КС#грамматики G
существует эквивалентная ей КС#грамматика–без–синомимов G''.
(5) Для любой однозначной КС#грамматики G
существует эквивалентная ей однозначная КС#грамматика–без–синомимов G''.
(6) Для любой разделенной КС#грамматики G
существует эквивалентная ей разделенная КС#грамматика–без–синомимов G''.

Обозначим $L(G, A, l) \stackrel{def}{=} \{x \in L(G, A) : |x| \le l\}$.

Из неравенства $L(G, A) \ne L(G, B)$ вытекает существование числа l_{AB} такого, что $L(G, A, l_{AB}) \ne L(G, B, l_{AB})$. Следовательно, для грамматики без синонимов $G =< N, \Sigma, P, S >$ существует число $l(G) = max\{l_{AB} \mid A \in N,\ B \in N,\ A \ne B\}$, которое гарантирует неравенство $L(G, A, l(G)) \ne L(G, B, l(G))$ для любых нетерминалов $A \ne B$. Число $l(G)$ является характеристикой КС–грамматики G, что позволяет сформулировать

Утверждение 37. *Для любого КС–языка L существуют*
– порождающая его КС–грамматика без синомимов $G =< N, \Sigma, P, S >$ и
– целое число l такое, что $L(G, A, l) \ne L(G, B, l)$ для всех различных A и B из N.

Аналогичные утверждения справедливы также для однозначных КС–языков, для разделенных КС–языков, для КС#языков, для однозначных КС#языков, а также для разделенных КС#языков.

* * *

В общем случае проверка условия (A.1) – задача алгоритмически неразрешимая, поэтому все преобразования, описанные и исследованные в приложении A, всего лишь подтверждают факт существования грамматик без синонимов.

ПРИЛОЖЕНИЕ

B. УСТРАНЕНИЕ ДЕЛИМЫХ НЕТЕРМИНАЛОВ

Опишем эквивалентные преобразования КС#грамматик, позволяющие в, конечном счете, устранять в таких грамматиках подмножества делимых–справа и делимых–слева нетерминалов. Соответствующие определения нетерминалов приводятся в разделе 3.1.

B.1. Устранение делимых–слева нетерминалов

Пусть $G =< N, \Sigma, P, S >$ – КС#грамматика и $LD(X) = \{A_1, A_2, \ldots, A_K\}$ – непустое подмножество ее делимых–слева нетерминалов, зафиксированное для некоторого $X \in N \cup \Sigma$.

Закодируем исходные данные формальной цепочкой $W = L\,X\,A_1\,A_2 \ldots A_K$, которую будем использовать в качестве индекса. Если $X = C$ и $LD(C) = \{A, B\}$, то в качестве W в равной степени могут использоваться $LCAB$ и $LCBA$. Полагаем, что в качестве W фиксирован один из возможных вариантов.

Подмножеству $LD(X) = \{A_1, A_2, \ldots, A_K\}$ сопоставим подмножество уникальных нетерминальных символов $LD'(X) = \{A'_1, A'_2, \ldots, A'_K\}$. При этом будем считать, что нетерминалу A_1 сопоставлен нетерминал A'_1, нетерминалу A_2 сопоставлен нетерминал A'_2 и т.д.

Введем подмножества:

$$\begin{aligned} N' &\stackrel{def}{=} N \;\setminus LD(X), \\ N_W &\stackrel{def}{=} N' \;\cup LD'(X), \\ U &\stackrel{def}{=} N \;\cup \Sigma, \\ U_W &\stackrel{def}{=} N_W \cup \Sigma. \end{aligned}$$

Определим пару формальных преобразований цепочек:

- $\triangleright$ цепочка $\overline{\alpha}$ из U_W^* получается из цепочки $\alpha \in U^*$ заменой всех символов A из $LD(X)$ на цепочки из двух символов XA', где A' сопоставленные нетерминалы из $LD'(X)$;
- $\triangleright$ цепочка $\overline{\overline{\alpha}}$ из U^* получается из цепочки $\alpha \in U_W^*$ заменой всех вхождений XA' на сопоставленные нетерминалы A.

Первое преобразование применимо к любым цепочкам из U^*, а второе – только к сбалансированным цепочкам из U_W^*, в которых перед каждым символом из $LD'(X)$ располагается символ X. Например, для $\Sigma{=}\{a, b, c, +\}$, $N{=}\{A, B\}$, $LD(B){=}\{A\}$ и $LD'(B){=}\{A'\}$ имеем:

1). $\overline{Aa + A + aBc} \;=\; BA'a + BA' + aBc$,

2). $\overline{\overline{BA'c + BA'}} \;=\; Ac + A$,

3). цепочка $BA'a + bA' + aBc$ не является сбалансированной.

В дальнейшем существенно используются простые закономерности

если $\alpha \in U^*$ и $\gamma = \overline{\alpha}$, *то* γ – сбалансированная цепочка и $\overline{\overline{\gamma}} = \alpha$;

если $\alpha_1,\, \alpha_2 \in U^*$,
то $\overline{\alpha_1\alpha_2} = \overline{\alpha}_1\,\overline{\alpha}_2$ и равенство $\alpha_1{=}\alpha_2$ равносильно равенству $\overline{\alpha}_1{=}\overline{\alpha}_2$;
если $\gamma_1,\, \gamma_2$ – сбалансированные цепочки из U_W^*,
то $\overline{\overline{\gamma_1\gamma_2}} = \overline{\overline{\gamma}}_1\,\overline{\overline{\gamma}}_2$ и равенство $\gamma_1{=}\gamma_2$ равносильно равенству $\overline{\overline{\gamma}}_1{=}\overline{\overline{\gamma}}_2$;
если $A \in LD(X)$, *то* $\overline{A} = XA'$;
если $A \in N \setminus LD(X)$, *то* $\overline{A} = A'$.

Определим специальное преобразование Ψ правил вывода из P и связанные с ним
▷ множество правил вывода $P_W = \Psi(P)$ и
▷ грамматику $G_W = < N_W, \Sigma, P_W, S >$.

$$\Psi(A \to \beta) \overset{def}{=} \begin{cases} A' \to \overline{\alpha}, & \text{если } A \in LD(X) \text{ и } \beta = X\alpha, \\ A' \to B'\overline{\alpha}, & \text{если } A \in LD(X) \text{ и } \beta = B\alpha, \\ A \to \overline{\beta}, & \text{если } A \notin LD(X). \end{cases}$$

Если, например, в качестве грамматики G рассматривается

$$\begin{array}{llllllll} G^L: & S & \to & Ac & | & BaA & | & \# \\ & A & \to & AC & | & BaC & & \\ & B & \to & Bb & | & Cd & & \\ & C & \to & d & | & cCb, & & \end{array}$$

и $LD(C) = \{A, B\}$, то $W = LCAB$ и G_W имеет вид:

$$\begin{array}{llllllll} G^L_{LCAB}: & S & \to & CA'c & | & CB'aCA' & | & \# \\ & A' & \to & A'C & | & B'aC & & \\ & B' & \to & B'b & | & d & & \\ & C & \to & d & | & cCb. & & \end{array}$$

Утверждение 38. Ψ – *взаимно однозначное преобразование.*

Для доказательства предположим противное – пусть для некоторого правила из P_W существуют два различных правила–прообраза:

$$\Psi(A_1 \to \beta_1) = \Psi(A_2 \to \beta_2) = C \to \beta, \qquad \text{но} \quad A_1 \neq A_2 \quad \text{и/или} \quad \beta_1 \neq \beta_2. \tag{A.11}$$

Неравенство $A_1 \neq A_2$ невозможно по определению сопоставимых терминов.
Если $A_1 \notin LD(X)$, то из (B.1) вытекает: $\overline{\beta}_1 = \beta$ и $\overline{\beta}_2 = \beta \ggg \beta_1 = \beta_2$.
Если $A_1 \in LD(X)$, то в (B.1) теоретически возможны пять вариантов:
1. $\beta_1 = X\alpha_1$, $\beta_2 = X\alpha_2 \ggg \overline{\alpha}_1 = \beta$ и $\overline{\alpha}_2 = \beta \ggg \alpha_1 = \alpha_2 \ggg \beta_1 = \beta_2$;
2. $\beta_1 = B\alpha_1$, $\beta_2 = X\alpha_2$ и $B \in LD(X) \ggg B'\,\overline{\alpha}_1 = \beta$ и $\overline{\alpha}_2 = \beta \ggg B'\,\overline{\alpha}_1 = \overline{\alpha}_2$,
но $B'\,\overline{\alpha}_1$ – несбалансированная цепочка, а $\overline{\alpha}_2$ – сбалансированная;
3. $\beta_1 = X\alpha_1$, $\beta_2 = B\alpha_2$ и $B \in LD(X)$ – аналог варианта 2;
4. $\beta_1 = B\alpha_1$, $\beta_2 = B\alpha_2$ и $B \in LD(X) \ggg B'\overline{\alpha}_1 = \beta$ и $B'\overline{\alpha}_2 = \beta \ggg \alpha_1 = \alpha_2 \ggg \beta_1 = \beta_2$;
5. $\beta_1 = B_1\alpha_1$, $\beta_2 = B_2\alpha_2$, $\{B_1,\, B_2\} \subseteq LD(X)$ и $B_1 \neq B_2$
$\ggg B_1'\overline{\alpha}_1 = \beta$ и $B_2'\overline{\alpha}_2 = \beta \ggg B_1' = B_2' \ggg B_1 = B_2$.

Во всех случаях получены противоречия. Утверждение доказано.

Утверждение 39. *Если* $A \in N$, *то* $L(G, A) = L(G_W, \overline{A})$.

Доказательство. Выберем в грамматике G произвольный нетерминал A и докажем, что $L(G,A) \subseteq L(G_W,\overline{A})$ и $L(G,A) \supseteq L(G_W,\overline{A})$.

Часть 1. Покажем, что из вывода $A \Rightarrow_G x$ следует вывод $\overline{A} \Rightarrow_{G_W} x$.

Рассмотрим произвольное предложение x, выводимое в грамматике G из нетерминала A, и пусть один из правых выводов имеет вид:

$$A = \sigma_0 \ \Rightarrow_G \ \sigma_1 \ \Rightarrow_G \ \ldots \ \Rightarrow_G \ \sigma_{k-1} \ \Rightarrow_G \ \sigma_k = x.$$

Докажем по индукции, что формально построенная последовательность $\overline{\sigma}_0$, $\overline{\sigma}_1$, ..., $\overline{\sigma}_k$ есть правый вывод предложения x из цепочки $\overline{A}$ в грамматике G_W.

Поскольку $\overline{A} = \overline{\sigma}_0$, то $\overline{A} \ \Rightarrow_{G_W} \overline{\sigma}_0$.

Предположим для некоторого $i < k$, установлено, что $\overline{A} = \overline{\sigma}_0 \ \Rightarrow_{G_W} \overline{\sigma}_1 \ \Rightarrow_{G_W} \ldots \ \Rightarrow_{G_W} \overline{\sigma}_i$ Покажем, что в этом случае $\overline{\sigma}_i \ \Rightarrow_{G_W} \overline{\sigma}_{i+1}$.

Рассмотрим $i+1$–й этап правого вывода предложения x в грамматике G: $\sigma_i \ \Rightarrow_G \sigma_{i+1}$. По определению правого вывода:

$$\sigma_i = \gamma A z, \qquad \sigma_{i+1} = \gamma\beta z \quad \text{и} \quad A \to \beta \quad \text{есть правило грамматики } G.$$

Для правила $A \to \beta$ возможны три варианта:

либо $A \in LD(X)$ и $\beta = X\alpha$ и тогда $\overline{\sigma}_i = \overline{\gamma A z} = \overline{\gamma} X A' z$, $\overline{\sigma}_{i+1} = \overline{\gamma X \alpha z} = \overline{\gamma} X \overline{\alpha} z$,
то есть $\overline{\sigma}_i \ \Rightarrow_{G_W} \overline{\sigma}_{i+1}$ посредством правила $A' \to \overline{\alpha}$ из P_W;

либо $A \in LD(X)$ и $\beta = B\alpha$ для некоторого $B \in LD(X)$ и тогда
$\overline{\sigma}_i = \overline{\gamma A z} = \overline{\gamma} X A' z$, $\overline{\sigma}_{i+1} = \overline{\gamma B \alpha z} = \overline{\gamma} X B' \overline{\alpha} z$,
то есть $\overline{\sigma}_i \ \Rightarrow_{G_W} \overline{\sigma}_{i+1}$ посредством правила $A' \to \overline{\alpha}$ из P_W;

либо $A \notin LD(X)$ и $\beta = \alpha$ и тогда $\overline{\sigma}_i = \overline{\gamma A z} = \overline{\gamma} A z$, $\overline{\sigma}_{i+1} = \overline{\gamma \alpha z} = \overline{\gamma}\,\overline{\alpha}\,z$,
то есть $\overline{\sigma}_i \ \Rightarrow_{G_W} \overline{\sigma}_{i+1}$ посредством правила $A' \to \overline{\alpha}$ из P_W.

Во всех трех случаях $\overline{\sigma}_i \ \Rightarrow_{G_W} \overline{\sigma}_{i+1}$ посредством правила $\Psi(A \to \beta)$, и, следовательно, $\overline{\sigma_k} = x \in L(G_W,\overline{A})$. Первая часть утверждения доказана.

Часть 2. Покажем, что из вывода $\overline{A} \Rightarrow_{G_W} x$ следует наличие вывод $A \Rightarrow_G x$.

Доказательство проведем индукцией по длине правого вывода некоторого предложения x выводимого в грамматике G_W из цепочки $\overline{A}$.

$$\overline{A} = \sigma_0 \ \Rightarrow_{G_W} \sigma_1 \ \Rightarrow_{G_W} \ldots \ \Rightarrow_{G_W} \sigma_{k-1} \ \Rightarrow_{G_W} \sigma_k = x.$$

Покажем, что:

1и) $\sigma_0, \ \sigma_1, \ \ldots, \ \sigma_{k-1}, \ \sigma_k$ – сбалансированные цепочки; и

2и) последовательность цепочек $\overline{\overline{\sigma}}_0, \ \overline{\overline{\sigma}}_1, \ \ldots, \ \overline{\overline{\sigma}}_{k-1}, \ \overline{\overline{\sigma}}_k$ есть правый вывод предложения x из нетерминала A в грамматике G.

Цепочка $\overline{\overline{\sigma}}_0$ есть нетерминал A, и поэтому она удовлетворяет условиям 1и) и 2и).

Предположим, что для некоторого $i < k$ установлено:

1п) $\sigma_0, \ \sigma_1, \ \ldots, \ \sigma_i$ – сбалансированные цепочки; и

2п) последовательность $A = \overline{\overline{\sigma}}_0 \ \Rightarrow_G \overline{\overline{\sigma}}_1 \ \Rightarrow_G \ldots \ \Rightarrow_G \overline{\overline{\sigma}}_i$ есть правый вывод цепочки $\overline{\overline{\sigma}}_i$.

Рассмотрим переход $\sigma_i \ \Rightarrow_{G_W} \sigma_{i+1}$. По определению правого вывода $\sigma_i = \gamma C z$, $\quad \sigma_{i+1} = \gamma\beta z$ и $C \to \beta$ есть правило грамматики G_W.

Если $C \in LD'(X)$, то из сбалансированности цепочки σ_i, следует, что $\gamma = \gamma_1 X$ и γ_1 – сбалансированная цепочка.

По построению грамматики G_W для правила $C \to \beta$ возможны три варианта:

либо $C \in LD'(X)$ и $C \to \beta$ имеет вид $A' \to \overline{\alpha} = \Psi(A \to X\alpha)$,

тогда $\overline{\overline{\sigma}}_i = \overline{\overline{\gamma A' z}} = \overline{\overline{\gamma_1 X A' z}} = \overline{\overline{\gamma}}_1\, A\, z, \quad \sigma_{i+1} = \gamma_1 X \overline{\alpha} z, \quad \overline{\overline{\sigma}}_{i+1} = \overline{\overline{\gamma}}_1\, \overline{\overline{X\overline{\alpha}}}\, z = \overline{\overline{\gamma}}_1\, X\alpha\, z,$

то есть σ_{i+1} – сбалансированная цепочка и

$\overline{\overline{\sigma}}_i \ \Rightarrow_G \ \overline{\overline{\sigma}}_{i+1}$ посредством правила $A \to X\alpha$ из P;

либо $C \in LD'(X)$ и $C \to \beta$ имеет вид $A' \to B'\overline{\alpha} = \Psi(A \to B\alpha)$,

тогда $\overline{\overline{\sigma}}_i = \overline{\overline{\gamma A' z}} = \overline{\overline{\gamma_1 X A' z}} = \overline{\overline{\gamma}}_1\, A\, z, \quad \sigma_{i+1} = \gamma B' \overline{\alpha} z, \quad \overline{\overline{\sigma}}_{i+1} = \overline{\overline{\gamma_1\ X\ \ B'\ \overline{\alpha}\ z}} = \overline{\overline{\gamma}}_1\, B\alpha\, z,$

то есть σ_{i+1} – сбалансированная цепочка и

$\overline{\overline{\sigma}}_i \ \Rightarrow_G \ \overline{\overline{\sigma}}_{i+1}$ посредством правила $A \to X\alpha$ из P;

либо $C \notin LD'(X)$ и $C \to \beta$ имеет вид $A' \to \overline{\alpha} = \Psi(A \to \alpha)$,

тогда $\overline{\overline{\sigma}}_i = \overline{\overline{\gamma A z}} = \overline{\overline{\gamma}}\, A\, z, \quad \sigma_{i+1} = \gamma \overline{\alpha} z, \quad \overline{\overline{\sigma}}_{i+1} = \overline{\overline{\gamma\ \overline{\alpha}\ z}} = \overline{\overline{\gamma}}\, \alpha\, z,$

то есть σ_{i+1} – сбалансированная цепочка и

$\overline{\overline{\sigma}}_i \ \Rightarrow_G \ \overline{\overline{\sigma}}_{i+1}$ посредством правила $A \to \alpha$ из P.

Во всех трех случаях σ_{i+1} есть сбалансированная цепочка и $\overline{\overline{\sigma}}_i \ \Rightarrow_G \overline{\overline{\sigma}}_{i+1}$ посредством правила $\Psi^{-1}(C \to \beta)$, и, следовательно, $\overline{\overline{\sigma}}_k = x \in L(G, A)$. Вторая часть утверждения и все утверждение 39 в целом доказаны.

При $A = S$ из утверждения 39 вытекает эквивалентность грамматик G и G_W. Помимо прочего, в доказательстве утверждения 39 установлен факт существования взаимно однозначного соответствия правых выводов одинаковых предложений в грамматиках G и G_W, откуда следуют утверждения 40 – 41.

Утверждение 40. *Если G – однозначная КС#грамматика, то G_W также однозначная КС#грамматика.*

Утверждение 41. *Если КС#грамматика G содержит делимое-слева непустое подмножество нетерминалов, то существует эквивалентная ей КС#грамматика G_W, для которой $\Upsilon_3(G_W) < \Upsilon_3(G)$.*

В самом деле, из утверждения 39 следует:
если $A \notin LD(X)$, то $L(G, A) = L(G_W, A)$
если $A \in LD(X)$, то $L(G, A) = L(G_W, XA') = L(G_W, X)\, L(G_W, A')$ см.[14]
и из утверждения 10 вытекает искомое неравенство $\Upsilon_3(G_W) < \Upsilon_3(G)$.

Аналогичные пассуждения справедливы и для делимых–справа нетерминалов.

B.2. Устранение делимых–справа нетерминалов

Пусть $G =< N, \Sigma, P, S >$ – КС#грамматика и $RD(X)$ – непустое подмножество ее делимых–справа нетерминалов, зафиксированное для некоторого $X \in N \cup \Sigma$.

1. Подмножеству $RD(X) = \{A_1, A_2, \ldots, A_K\}$ сопоставляется подмножество уникальных нетерминальных символов $RD'(X) = \{A'_1, A'_2, \ldots, A'_K\}$.

2. Вводятся подмножества:

$$\begin{aligned} N' &\stackrel{def}{=} N \ \setminus RD(X), \\ N_V &\stackrel{def}{=} N' \ \cup RD'(X), \\ U &\stackrel{def}{=} N \ \cup \Sigma, \\ U_V &\stackrel{def}{=} N_V \cup \Sigma. \end{aligned}$$

[14] Язык $L(G, A)$ есть конкатенация языков $L(G_W, X)$ и $L(G_W, A')$, не содержащих пустых цепочек.

3. Определяется формальное преобразование цепочек: цепочка $\overline{\alpha}$ получается из цепочки $\alpha \in U^*$ заменой всех символов A из $RD(X)$ на цепочки $A'X$, где A' сопоставленные нетерминалы из $RD'(X)$.

4. Вводятся:

- $\triangleright$ специальное преобразование Φ правил вывода из P;
- $\triangleright$ множество правил вывода $P_V = \Phi(P)$ и
- $\triangleright$ грамматика $G_V = < N_V, \Sigma, P_V, S >$.

$$\Phi(A \to \beta) \stackrel{def}{=} \begin{cases} A' \to \overline{\alpha}, & \text{если } A \in LD(X) \text{ и } \beta = X\alpha, \\ A' \to B'\overline{\alpha}, & \text{если } A \in LD(X) \text{ и } \beta = B\alpha, \\ A \to \overline{\beta}, & \text{если } A \notin LD(X). \end{cases}$$

5. Последовательно доказываются утверждения 42 – 44

Утверждение 42. *Для любого нетерминала $A \in N$ справедливо равенство $L(G, A) = L(G, \overline{A})$.*

Утверждение 43. *Если G – однозначная КС#грамматика, то G_V также однозначная КС#грамматика.*

Утверждение 44. *Если КС#грамматика G содержит делимое-справа непустое подмножество нетерминалов, то существует эквивалентная ей КС#грамматика G_V, для которой $\Upsilon_3(G_V) < \Upsilon_3(G)$.*

Конец приложения В.

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. Мальцев А.И. *Алгебраические системы.* М.: Наука, 1970.
4. Серебряков В.А. *Теория и реализация языков программирования.* М.: Физматлит, 2012.
5. Соловьев С.Ю. Нормализация контекстно-свободных грамматик для целей грамматического вывода. *XII национальная конференция по искусственному интеллекту с международным участием КИИ-2010. Труды конференции.* М: Физматлит, 2010, том 1, стр.218-224. *http://www.park.glossary.ru/serios/read_09.php*

The problem of compatibility properties of formal grammars

V.A.Serebyakov, S.Y.Soloviev

We consider the problem of compatibility properties of grammars in general and propose an approach to its solution.

KEYWORDS: formal language, grammar, model, functional.