4
П-71 1

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной
и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

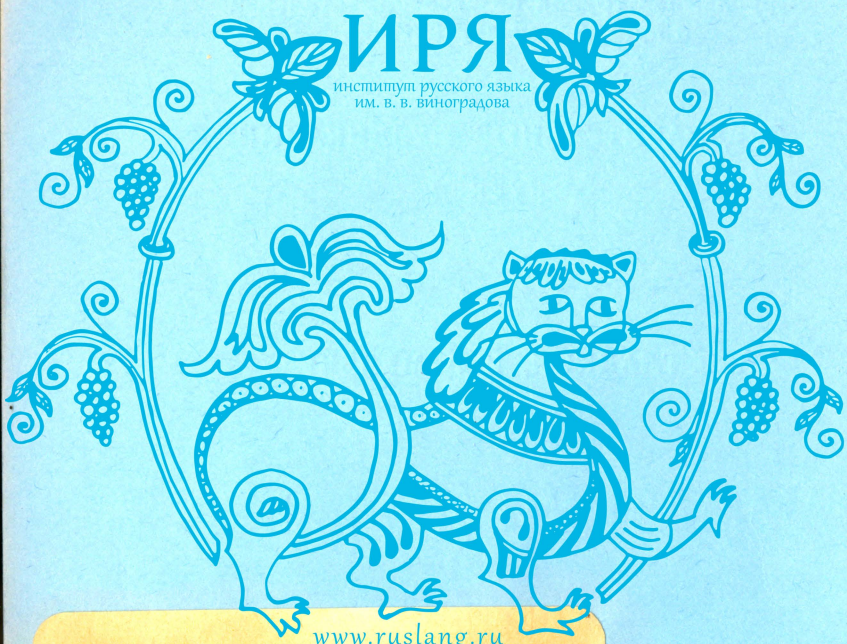
Выпуск 1

А. В. ГЛАДКИЙ, И. А. МЕЛЬЧУК

ГРАММАТИКИ ДЕРЕВЬЕВ. I

Москва — 1970

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

64667

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной
и прикладной лингвистике

Предварительные публикации
Выпуск 1

А.В.Гладкий, И.А.Мельчук

ГРАММАТИКИ ДЕРЕВЬЕВ. I

Москва - 1970

Настоящей работой открывается серия предварительных публикаций научных исследований, переводов и других материалов по тематике Проблемной Группы по Экспериментальной и Прикладной Лингвистике (ПГЭПЛ), созданной при ИРЯ АН СССР в соответствии с Постановлением Бюро ОЛЯ АН СССР от 29/IУ-1969 г.

Эти публикации имеют целью оперативное ознакомление специалистов, работающих в данной области, с новейшими, пусть предварительными результатами и с состоянием работы Группы. В работе ПГЭПЛ участвуют сотрудники институтов Языкознания, Русского Языка и Прикладной Математики АН СССР и I Московского Государственного Педагогического Института Иностранных Языков им. М.Тореза.

Проблемная Группа по
Экспериментальной
и Прикладной Лингвистике

№ 4
П-71
БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
АН-ТА ЯЗЫКОЗНАНИЯ АН СССР
64667

Опыт формализации преобразований синтаксических структур естественного языка.

§ 1. Введение.

Формальные порождающие грамматики в смысле Н.Хомского несомненно являются эффективным средством описания синтаксиса естественных (и искусственных) языков.

Однако эти грамматики, в ряде важных отношений обеспечивающие плодотворный подход к изучению языка, во многих (достаточно существенных) случаях не позволяют трактовать наблюдаемые факты адекватным, естественным образом. Мы попытаемся сформулировать, каковы, по нашему мнению, наиболее важные причины неадекватности порождающих грамматик.

Во-первых, порождающая грамматика в ее классическом виде предназначена для задания множества правильных фраз некоторого языка; более точно, она перечисляет все правильные фразы (и только их) и при этом сопоставляет каждой правильной фразе её синтаксическую структуру. В то же время лингвистически более содержательной и естественной является, по-видимому, задача соотнесения некоторому заданному смыслу всех соответствующих ему правильных фраз. Именно такой подход отвечал бы общепри-

---

* Синтаксис понимается здесь в самом широком смысле - как способ составления любых языковых единиц (словоформ, словосочетаний, предложений) из более простых компонентов.

нятому взгляду на язык как на орудие общения, средство выражений мыслей.

Здесь вряд ли уместно специально обосновывать целесообразность подхода "смысл ⟷ текст" (хотя его принятие влечет далеко идущие последствия*). Ограничимся следующим выводом: если принять этот подход и связанную с ним лингвистическую концепцию, то в качестве формального аппарата лингвистических описаний желательно иметь механизм, который не просто перечислял бы правильные фразы (хотя бы и с их синтаксическими структурами), но был бы в идеале способен сопоставлять любому заданному смыслу все правильные фразы, имеющие этот смысл, и обратно – сопоставлять любой заданной фразе её смысл или смыслы.

Замечание 1. Слово "смысл" следует понимать здесь очень широко – как обозначение "содержания" любого отрезка речи (каково бы оно ни было: с точки зрения логики оно может быть, в частности, противоречивым или даже абсурдным).

Замечание 2. Предполагается, что смыслы могут быть представлены как конструктивные объекты, построенные по определенным правилам из элементов некоторого конечного набора – например, как графы, вершины которых помечены символами "элементарных смыслов", а дуги – символами "элементарных смысловых связей".

Во-вторых, порождающие грамматики работают с цепочками – линейно упорядоченными множествами символов, что приводит к одинаковому обращению с двумя совершенно различными по своей природе типами связей между единицами естественного текста.

---

* О модели "смысл ⟷ текст" см., в частности, [1] – [4].

Так, между словоформами предложения имеются отношения двух сортов: отношения синтаксической связанности ("А зависит от В" или "А и В образуют составляющую С" и т.п.) и отношения линейного расположения ("А рядом с В", "А левее В", "А между В и С" и т.п.). В языках с фиксированным порядком слов, где последний выступает как важное синтаксическое средство, – как, например, в английском, – эти два типа отношений легко спутать, поскольку данной синтаксической структуре отвечает, как правило, один порядок элементов (одна цепочка), или, в крайнем случае, небольшое число допустимых порядков. Не исключено, что именно эта особенность английского языка обусловила тот факт, что в англоамериканской грамматической традиции получил преимущественное развитие анализ по НС, формализацией которого в большой степени являются грамматики Хомского*.

Однако, если в языке порядок слов является (относительно) свободным, т.е. одной и той же синтаксической структуре отвечает в общем случае много возможных словорасположений – как, скажем, в русском языке, – то необходимость различения синтаксических и линейных отношений сразу бросается в глаза. Этот факт неоднократно обсуждался в лингвистической литературе; достаточно сослаться на известную книгу Л.Теньера [6] и статью Д.С.Уорта [27].

Исходя из сказанного, мы считаем желательным иметь в составе аппарата для описания синтаксиса естественных языков два разных механизма: один должен работать с синтаксическими

*Ср., в связи с этим, [27] и [5].

структурами, не снабженными линейным порядком, а другой – отображать готовые структуры на цепочки (указанный принцип был сформулирован уже в [27]; аналогичная идея лежит в основе работы [15]).

Итак, нужна формальная конструкция, которая могла бы осуществлять преобразование "смысл ⟷ текст"; это преобразование должно выполняться в несколько этапов, связанных с различными уровнями представления текста. В принципе эти этапы можно выделять по-разному; мы будем следовать здесь членению перехода от смысла к тексту, принятому в [1] – [4]. Кратко резюмируем те сведения из указанных работ, которые понадобятся нам в дальнейшем.

В качестве способа изображения (синтаксической) структуры используется дерево зависимостей *. В частности, в терминах зависимостей представляется поверхностная синтаксическая структура. Наша поверхностная структура идейно родственна с surface structure трансформационной грамматики (с тем отличием, что surface structure является структурой составляющих и в ней задан порядок терминальных символов, чего нет у нас). Для целей настоящей статьи достаточно представлять себе поверхностную структуру как дерево, в узлах которого стоят лексемы, а ветви помечены символами типов синтаксических отношений.

При движении от смысла к тексту поверхностное дерево фразы должно в конечном счете получаться из некоторого описания ее смысла (в общем случае – не единственным образом). Мы не беремся сколько-нибудь серьезно обсуждать конкретную форму

* О различных представлениях синтаксической структуры (в частности, о деревьях зависимостей) см., например, [7].

смыслового представления фразы; можно предполагать (ср. замечание 2 на стр. 2), что это будет сложный граф, не обязательно являющийся деревом, причем между ним и поверхностной структурой не будет достаточно простого соответствия ни в отношении топологии, ни в отношении состава элементов (пометок на узлах и ветвях). Поскольку прямой переход от представления смысла ко всем многочисленным соответствующим ему поверхностным деревьям слишком сложен, оказывается удобным ввести некоторый промежуточный уровень, глубинные, или лексико-синтаксические структуры (ЛСС). Каждая ЛСС – это обобщенное представление некоторого числа синонимичных поверхностных деревьев. А именно, это такие дерево с пометками на узлах и ветвях; однако эти пометки должны быть совсем не такими, как в поверхностной структуре: как раз за счет их более общего характера и достигается редукция синонимии. Что касается соотношения ЛСС с deep structure трансформационной грамматики, то наряду с существенным сходством, здесь имеется и важное различие: кроме глубинного синтаксиса (как в deep structure) в ЛСС используется еще и глубинная лексика – так называемые лексические функции (см. [1] – [4]).

Переход от исходного представления смысла к поверхностной структуре фразы с использованием ЛСС предполагается осущест-

---

* Естественные языки обычно располагают такими богатыми синонимическими средствами, что достаточно сложная фраза может иметь сотни тысяч и даже миллионы более или менее синонимичных ей перифраз: см. пример в [1], стр. 179–180.

влять по крайней мере в три этапа:

1) от смысла – хотя бы к одной соответствующей ЛСС;

2) от полученной ЛСС – к другим (в идеале – ко всем) ЛСС, синонимичным ей ("перифразирование"; см. [1] – [4]);

3) от каждой ЛСС – ко всем соответствующим поверхностным структурам.

Этап перехода от смысла к "первичной" ЛСС в данной статье не затрагивается *. Таким образом, наша тематика достаточно суживается. Из всего сложного процесса перехода от смысла к тексту мы рассматриваем только два этапа: переход от некоторой глубинной лексико-синтаксической структуры ко всем другим синонимичным ЛСС и переход от ЛСС к поверхностной синтаксической структуре.

Теперь, наконец, мы можем точно сформулировать нашу цель. Она состоит в том, чтобы изложить основы формального аппарата, позволяющего осуществлять некоторые преобразования деревьев. При этом указанные преобразования должны быть ориентированы на переход от одних глубинных деревьев (ЛСС) к другим и от глубинных деревьев – к поверхностным. Другими словами, мы предлагаем строить формальные грамматики, работающие с деревьями, – так сказать, грамматики деревьев. Эти грамматики мы будем называть Δ-грамматиками (Δ – от греч. $\delta\acute{\varepsilon}\nu\delta\rho\sigma\nu$ 'дерево').

---

* Проблема получения ЛСС по исходному смыслу связана, возможно, с задачей "считывания" с графа всех его поддеревьев, имеющих заданные свойства. Решение некоторой задачи такого типа, возникающей из лингвистических соображений, имеется в [8].

Следует отметить, что многие идеи, развиваемые в данной статье, формулировались, разрабатывались ранее. Прежде всего, следует указать на классические работы Н.Хомского, например, [9], а также примыкающие к ним исследования по формализации семантики, такие, как [10] - [12]. Мы имеем в виду, главным образом, трансформации, которые представляют собой операции над деревьями (составляющих). В грамматиках зависимостей ([24], [25]) и доминационных грамматиках [26] также используется некоторое специальное преобразование деревьев, снабженных порядком. Подход, близкий к принципу модели "смысл ⟷ текст", выдвигается в работах П.Сгалла (см., например, [13]). Необходимость разграничения синтаксических связей и линейного порядка неоднократно подчеркивалась С.К.Шаумяном [14]. Сходная идея проводится в [15] и [22]. Пример грамматики, порождающей структуры (русских фраз) с последующей линеаризацией, был построен М.В.Ломковской [16].

В ряде работ определяются преобразования графов, близкие к используемой нами операции подстановки дерева в дерево; см., например, [17] и [18], стр. 9. В книге Ч.Ф.Хоккетта [19] предлагается использовать аналогичные операции над графами при разработке формальных грамматик для естественных языков.

Далее, имеется статья [20], специально посвященная преобразованиям деревьев, содержательно интерпретируемых как деревья составляющих; эти преобразования предназначаются для формализации лингвистических трансформаций. Операции над деревьями (носящие, впрочем, иной характер, нежели наши) рас-

сматриваются в [21]. Некоторый способ порождения деревьев зависимостей вводится в [22].

Наконец, работа О.С.Кулагиной [23] посвящена той же проблематике, что и наша статья; естественно, что содержание обеих работ оказывается достаточно близким - тем более, что авторы неоднократно обсуждали с О.С.Кулагиной соответствующие вопросы. Ею был прочитан первоначальный вариант настоящей статьи, в который в соответствии с ее предложениями были внесены существенные исправления.

Авторы выражают О.С.Кулагиной свою искреннюю признательность. Они благодарят также А.Я.Диковского, указавшего ряд неточностей, С.Я.Фитиалова, Ю.А.Шиханвича, Ю.А.Шрейдера и других коллег, читавших рукопись и сделавших ценные замечания.

§ 2.

Синтаксические Δ-грамматики.

Перейдем к изложению основных определений, относящихся к Δ-грамматикам. Объекты, с которыми имеют дело эти последние, представляют собой конечные "размеченные" деревья, т.е. деревья, узлы и ветки которых помечены некоторыми символами.

Нам будет удобно начать с некоторого специального класса Δ-грамматик. Грамматики этого класса мы предлагаем называть синтаксическими Δ-грамматиками. От произвольных Δ-грамматик синтаксические Δ-грамматики отличаются тем, что они работают с деревьями без пометок на узлах (но с пометками на ветвях); содержательно такие деревья могут интерпретироваться

как "чистые" синтаксические структуры, без каких бы то ни было лексических заполнений.

Таким образом, объекты, с которыми имеют дело синтаксические Δ-грамматики, – это ориентированные деревья с помеченными ветвями (в дальнейшем они будут для краткости называться просто деревьями).

Говоря о деревьях, мы будем пользоваться терминологией, более принятой в работах по лингвистике и автоматическому переводу, но отличной от обычной терминологии теории графов. А именно, то, что в теории графов называют вершиной, дугой и корнем, мы будем называть узлом, ветвью и вершиной соответственно.

Естественно говорить, что деревья T_1 и T_2 изоморфны, если между множествами узлов этих деревьев можно установить взаимнооднозначное соответствие так, что в дереве T_1 для любой пары узлов α и β ветвь с некоторой пометкой идет из α в β тогда и только тогда, когда в дереве T_2 из образа узла α идет в образ узла β ветвь с той же пометкой.

Минимальным деревом мы будем называть дерево, все узлы которого, кроме вершины, – концевые*; частным случаем минимального дерева является дерево, состоящее из одной вершины – единичное дерево (обозначение: ε).

Введем теперь операцию над деревьями, которую мы назовем композицией. Она определяется следующим образом. Пусть дано

---

*Концевым называется узел, из которого не исходит ни одной ветви; ветвь, входящая в концевой узел, также называется концевой.

дерево t_0 и n деревьев $t_1, t_2, \ldots, t_n$. Пусть, кроме того, в дереве t_0 выделено n узлов $\alpha_1, \alpha_2, \ldots, \alpha_n$, не обязательно попарно различных. Тогда результатом композиции дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и деревьев $t_1, t_2, \ldots, t_n$ будет, по определению, любое дерево T, изоморфное дереву, полученному из t_0 "наклеиванием" вершин деревьев $t_1, t_2, \ldots, t_n$ на его узлы $\alpha_1, \alpha_2, \ldots, \alpha_n$ соответственно.

Подчеркнем, что результат композиции определяется с точностью до изоморфизма: если дерево T является результатом композиции дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и деревьев $t_1, t_2, \ldots, t_n$, то дерево T' будет результатом композиции того же t_0 с теми же выделенными узлами и тех же $t_1, t_2, \ldots, t_n$ тогда и только тогда, когда T' изоморфно T.

Композицию дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и деревьев $t_1, t_2, \ldots, t_n$ мы будем обозначать

$$C(t_0; \alpha_1, \alpha_2, \ldots, \alpha_n \mid t_1, t_2, \ldots, t_n). \qquad (1)$$

Будем говорить, что дерево t входит в дерево T (или что дерево t является поддеревом дерева T), если t является связным подграфом дерева T с точностью до изоморфизма, т.е. если T представимо в виде

$$T = C(T^0; \alpha_0 \mid C(t; \alpha_1, \alpha_2, \ldots, \alpha_n \mid T^1, T^2, \ldots, T^n)), \qquad (2$$

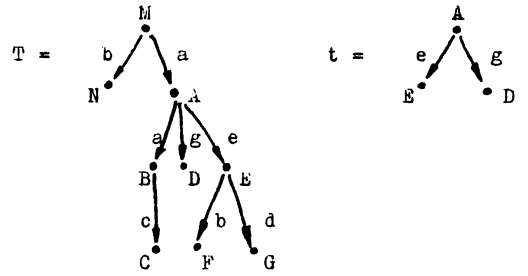
где α_0 – концевой узел дерева T^0, а $\alpha_1, \alpha_2, \ldots, \alpha_n$ – бесповторный пересчет всех узлов дерева t *.

---

* Обратим внимание на то, что один из узлов $\alpha_1, \alpha_2, \ldots, \alpha_n$ совпадает с α_0.

Ясно, что и это понятие определяется с точностью до изоморфизма: если дерево t входит в дерево T, то любое дерево, изоморфное t, входит в любое дерево, изоморфное T.

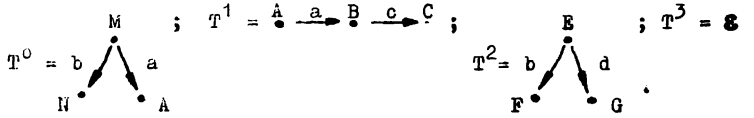
Пример:



Дерево t входит в T, поскольку

$$T = C(T^0;\ A \mid C\ (t;\ A,\ E,\ D \mid T^1,\ T^2,\ T^3)),$$

где



Дерево t может входить в T, так сказать, несколько раз, т.е. в разных местах; тогда говорят о разных <u>вхождениях</u> дерева t в T.

Определим теперь понятие элементарного преобразования деревьев. Элементарное преобразование заключается в том, что в преобразуемое дерево T вместо некоторого вхождения определенного дерева t_1 (t_1 – поддерево дерева T) подставляется другое определенное дерево t_2. Чтобы задать такую подстановку, необходимо указать: 1) оба дерева t_1 и t_2 – заменяемое и заменяющее; 2) способ соединения новых узлов (т.е. узлов заменяющего дерева t_2) со старыми, незамененными узлами дерева T.

Говоря формально, элементарное преобразование есть упорядоченная тройка $\langle t_1, t_2, f \rangle$, где t_1 и t_2 – деревья, а f – однозначное отображение множества узлов дерева t_1 во множество узлов дерева t_2. При этом, если узел α дерева t_1 отображается в узел β дерева t_2, т.е. $\beta = f(\alpha)$, то все незамененные узлы дерева T, "висевшие" на α, переве-шиваются на β с сохранением типов связи (пометок на ветвях).

Вместо $\langle t_1, t_2, f \rangle$ будет использоваться, для большей наглядности, запись "$t_1 \Rightarrow t_2 \mid f$".

Мы будем говорить, что дерево T' получается из дерева T применением элементарного преобразования $t_1 \Rightarrow t_2 \mid f$, ес-ли T и T' представимы соответственно в виде:

$$T = C(T^0; \alpha_0 \mid C(t_1; \alpha_1, \alpha_2, \ldots, \alpha_n \mid T^1, T^2, \ldots, T^n)) \quad (3)$$

и

$$T' = C(T^0; \alpha_0 \mid C(t_2; f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n) \mid T^1, T^2, \ldots, T^n)) \quad (4)$$

где α_0 – концевой узел дерева T^0, а $\alpha_1, \alpha_2, \ldots, \alpha_n$ – бесповторный пересчет узлов дерева t_1^*.

Из предыдущего ясно, что понятие применения элементарного преобразования также определяется с точностью до изоморфизма: если T' получается из T применением некоторого элементар-ного преобразования, то любое дерево, изоморфное T', полу-чается из любого дерева, изоморфного T, применением того же элементарного преобразования.

Пример.

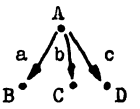
Пусть имеется элементарное преобразование

$t_1 \Rightarrow t_2 \mid f$, где

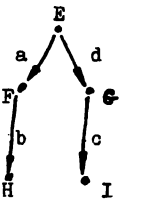
---

* Подчеркнем, что узлы $\alpha_1, \alpha_2, \ldots, \alpha_n$ попарно различны, а $f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n)$ – не обязательно попарно различны.

t_1 =



t_2 =



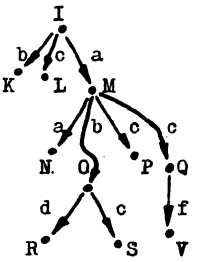
и

а f задано таблицей

$f(A) = E$ $\qquad$ $f(C) = H$

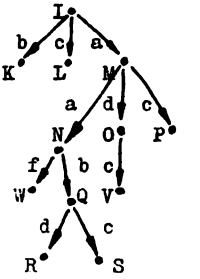
$f(B) = G$ $\qquad$ $f(D) = F$

Применим указанное преобразование к дереву T :



В дереве T имеется три вхождения t_1 : одно с вершиной I , и два – с вершиной M . Применим указанное преобразование к вхождению t_1 , имеющему узлы M, N, O, Q ; в результате получится дерево T' :

T' =



Наконец, мы можем перейти непосредственно к синтаксическим Δ-грамматикам.

Синтаксическая Δ-грамматика состоит из словаря и конечного числа правил.

Словарь представляет собой конечное множество символов - пометок на ветвях деревьев; содержательно, это типы синтаксических связей.

Каждое правило есть некоторое элементарное преобразование (в смысле только что сформулированного определения), в котором ветви левой и правой части помечены символами из словаря.

В соответствии с тем, что говорилось в § 1 о задачах Δ-грамматик, синтаксическая Δ-грамматика должна обеспечивать преобразования одних деревьев в другие. Теперь мы в состоянии дать точное определение таких преобразований, которые мы будем называть выводами.

Вывод в синтаксической Δ-грамматике Г - это последовательность деревьев, в которой каждое следующее получается из предыдущего применением одного из правил (= элементарных преобразований) этой грамматики.

Дерево T' называется выводимым из дерева T в некоторой синтаксической Δ-грамматике Г, если существует такой вывод в Г, который начинается деревом T и оканчивается деревом T'.

Синтаксические Δ-грамматики являются исчислениями весьма общего вида, в то время как для возможных приложений представляют, по-видимому, интерес их различные специальные классы

Специализировать классы грамматик можно, в частности, путем наложения ограничений на их правила.

Начнем с требования, чтобы дерево t_2 в правой части правила $t_1 \Rightarrow t_2 \mid f$ имело не меньше узлов, чем дерево t_1 в левой части. Правила, удовлетворяющие такому требованию, назовем развертывающими, а грамматики, содержащие только такие правила, – развертывающими синтаксическими Δ-грамматиками.

Примером развертывающего правила может служить элементарное преобразование, приведенное на стр. I2-I3.

Потребуем, далее, чтобы развертывающее правило фактически заменяло ровно один элементарный "кусок" преобразуемого дерева. Наиболее элементарным куском дерева естественно считать отдельный узел. Однако следует позаботиться о том, чтобы при замене узла сохранялись его связи, т.е. чтобы из дерева, которым будет заменен данный узел, выходили в точности те же ветви, которые выходят из заменяемого узла. А поскольку мы не имеем права рассматривать ветвь, лишенную хотя бы одного из ее концов, возникает необходимость говорить о замене не узла, а минимального дерева (см. стр. 9). При этом все концевые узлы заменяемого минимального дерева должны перехо-дить в концевые узлы заменяющего дерева, и ветви, входящие в эти последние, должны нести те же пометки, что и ветви, входившие в их прообразы.

Кроме того, надо учитывать, что, хотя заменяется только одно минимальное дерево, возможность замены может быть обусловлена контекстом – окружением заменяемого минимального дерева в преобразуемом дереве как "сверху", так и "снизу".

Исходя из всех этих соображений, мы получим такое правило, у которого правая часть t_2 отличается от левой части

www.ruslang.ru
64667
БИБЛИОТЕКА

t_1 только тем, что некоторое минимальное дерево, входящее в t_1, заменено в правой части другим деревом, уже не обязательно минимальным, но содержащим не меньше концевых ветвей, чем заменяемое; при этом каждой ветви заменяемого минимального дерева должна соответствовать в заменяющем дереве своя концевая ветвь с той же пометкой. Все остальное в t_1 остается без изменения и играет роль контекста, задающего условия, при которых данная замена допустима. Такие правила и, соответственно, грамматики, содержащие только такие правила, мы назовем минимальными [развертывающими].

Сформулируем точное определение. Развертывающее правило $t_1 \Rightarrow t_2 \mid f$ называется минимальным, если t_1 и t_2 представимы соответственно в виде

$$t_1 = C(\tau^0; \alpha_0 \mid C(\mu; \alpha_1, \alpha_2, \ldots, \alpha_n \mid \tau^1, \tau^2, \ldots, \tau^n)) \quad (5)$$

и

$$t_2 = C(\tau^0; \alpha_0 \mid C(\nu; f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n) \mid \tau^1, \tau^2, \ldots, \tau^n)) \quad (6)$$

где 1) μ – минимальное дерево; 2) $\alpha_1, \alpha_2, \ldots \alpha_n$ – бесповторный пересчет узлов дерева μ; 3) α_1 – вершина μ; 4) $f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n)$ попарно различны; 5) $f(\alpha_2)$, $f(\alpha_3)$, ..., $f(\alpha_n)$ – концевые узлы дерева ν; 6) для каждого $i = 2, 3, \ldots, n$ пометка на ветви дерева μ, входящей в α_i, совпадает с пометкой на ветви дерева ν, входящей в $f(\alpha_i)$; 7) на всех узлах дерева t_1, отличных от $\alpha_1, \alpha_2, \ldots, \alpha_n$, отображение f является тождественным.

В качестве следующего шага можно потребовать, чтобы левая часть минимального правила состояла только из минимального дерева, т.е. чтобы контекст был пуст, или, иначе говоря, чтобы деревья $\tau^0, \tau^1, \tau^2, \ldots, \tau^n$, участвующие в представле-

ниях (5) и (6), были единичными. Подобные правила естественно назвать контекстно-свободными [минимальными развертывающими] правилами, а грамматики, содержащие только такие правила – контекстно-свободными Δ-грамматиками.

Итак, мы определили четыре класса синтаксических Δ-грамматик, образующие следующую иерархию: (произвольные) синтаксические Δ-грамматики $\supset$ развертывающие синтаксические Δ-грамматики $\supset$ минимальные синтаксические Δ-грамматики $\supset$ контекстно-свободные синтаксические Δ-грамматики.

§ 3.

Специальные элементарные преобразования. Универсальная синтаксическая Δ-грамматика

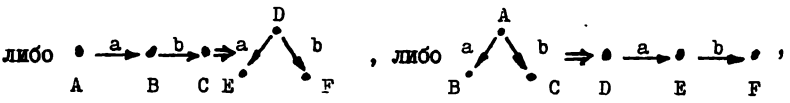
Из содержательных описаний перехода от одних глубинных структур к другим (в системе перифразирования, см. [1]–[4]) и от глубинных структур к поверхностным можно усмотреть, что среди элементарных преобразований общего вида, как они определены нами выше, с точки зрения лингвистических приложений особый интерес представляют некоторые весьма специальные элементарные преобразования.

Эти преобразования подразделяются на три следующих типа:

1. Расщепление узла ("вырастание" ветви): преобразование вида $A \Rightarrow B \overset{a}{\rightarrow} C$; при этом, очевидно, либо а) $f(A) = B$, либо б) $f(A) = C$. Такие преобразования будут записываться в виде:

а) $A \Rightarrow \underset{\sim}{a}(B, C) \mid f(A) = B$; б) $A \Rightarrow \underset{\sim}{a}(B, C) \mid f(A) = C$.

2. Перевешивание узла (ветви): преобразование вида



причем в обоих случаях $f(A) = D$, $f(B) = E$, $f(C) = F$. Такие преобразования записываются в виде а) $\underset{\sim}{a}(A, B) . \underset{\sim}{b}(B, C) \Rightarrow \underset{\sim}{a}(D, E) . \underset{\sim}{b}(D, F)$; б) $\underset{\sim}{a}(A,B) . \underset{\sim}{b}(A,C) \Rightarrow \underset{\sim}{a}(D, E) . \underset{\sim}{b}(E, F)$.

3. Склеивание узлов (стяжение ветви): преобразование вида $B \xrightarrow{a} \overset{C}{\bullet} \Rightarrow \overset{A}{\bullet}$; здесь, очевидно, $f(B) = f(C) = A$. Это преобразование мы будем записывать как $\underset{\sim}{a}(B, C) \Rightarrow A$.

Преобразования типов 1–3 мы будем называть специальными.

Оказывается, что произвольные элементарные преобразования в некотором смысле полностью сводимы к специальным.

Чтобы придать этому утверждению точный смысл, необходимо ввести понятие моделирования элементарных преобразований. Пусть имеется элементарное преобразование $t_1 \Rightarrow t_2 \mid f$ и некоторое множество элементарных преобразований M. Тогда выражение "преобразование $t_1 \Rightarrow t_2 \mid f$ моделируется преобразованиями из M" будет означать следующее: существует такая конечная последовательность преобразований $m_1, m_2, \ldots, m_n$, принадлежащих M, что, каковы бы ни были деревья T и T', если дерево T' получается из дерева T (однократным) применением преобразования $t_1 \Rightarrow t_2 \mid f$, то T' можно получить из T такие и последовательным применением преобразований $m_1, m_2, \ldots, m_n$.

Имеет место

Теорема 1. Произвольное элементарное преобразование моделируется специальными элементарными преобразованиями.

Доказательство теоремы 1, так же как и формулируемых ниже теорем 1′ и 2, будет приведено отдельно (см. "НТИ", 2-ая серия, 1970, в печати).

Пример. Элементарное преобразование $\overset{A}{\bullet}\xrightarrow{a}\overset{B}{\bullet} \Rightarrow \overset{A}{\bullet}\xrightarrow{b}\overset{B}{\bullet}$, где $f(A) = A$, $f(B) = B$ (т.е. переименование ветви), может быть заменено двумя последовательно выполняемыми специальными преобразованиями: 1) $A \Rightarrow \underset{\sim}{b}(A, C) \mid f(A) = C$ и 2) $\underset{\sim}{a}(C, B) \Rightarrow B$ (рис. 1).

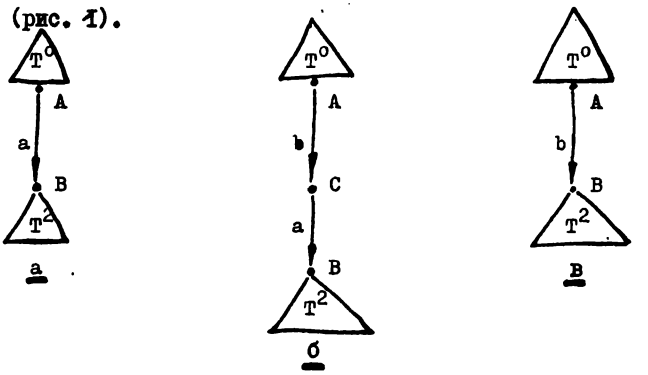


Рис. 1.

Подчеркнем, что возможность моделировать одни элементарные преобразования другими отнюдь не означает эквивалентности соответствующих грамматик*. Из теоремы 1 не следует, что для любой Δ-грамматики существует эквивалентная Δ-грамматика с правилами только типов расщепления, перевешивания и склеива-

---

* Две Δ-грамматики естественно назвать эквивалентными, если для любых деревьев T и T' выводимость T' из T в одной грамматике равносильна выводимости T' из T в другой.

ния . Эта теорема означает лишь, что в любой Δ-грамматике все ее правила могут быть представлены в терминах, так сказать, элементарных шагов указанных трех типов; однако эти шаги сами по себе не являются правилами грамматики: их можно выполнять только группами (каждая из которых соответствует одному правилу грамматики), причем внутри групп имеется жесткий порядок; мало того, место применения каждого следующего специального преобразования (напомним, что преобразование в принципе может быть применимо в нескольких местах одного дерева) зависит от места применения предыдущего.

Теорема 1 представляется весьма существенной с содержательной, т.е. лингвистической точки зрения. В самом деле, она означает, что любые возможные в естественных языках преобразования синтаксических структур в рамках одной фразы, т.е. любые преобразования типа "фраза ⟶ фраза" , как чисто синтаксические, так и сопровождаемые произвольными лексическими изменениями, могут быть сведены к трем простым операциям: расщепление узла, перевешивание узла, стяжение ветви.

Это обстоятельство позволяет построить исчисление всех мыслимых чисто синтаксических трансформаций - некоторую уни-

---

* Примером Δ-грамматики, для которой не существует эквивалентной Δ-грамматики с правилами только этих типов, может служить, в частности, любая Δ-грамматика, у которой в каждом правиле разность между числом узлов левой и правой части больше единицы.

** Преобразованиями другого типа являются операции соединения двух фраз в одну или расчленения одной фразы на две.

версальную синтаксическую Δ-грамматику. Нужно только принять допущение, что для выражения любых синтаксических связей в любых языках можно обойтись конечным словарем V пометок на ветвях. Тогда множество π всех таких преобразований типов 1-3, в которых используемые пометки на ветвях принадлежат V, тоже будет конечно.

Синтаксическая Δ-грамматика, состоящая из словаря V и набора правил π, обладает тем свойством, что в ней из любого дерева с пометками, принадлежащими V, выводимо любое другое такое дерево. Иначе говоря, любая трансформация (любое преобразование деревьев с пометками из V) представима как вывод в этой универсальной грамматике.

§ 4.

Правильные синтаксические Δ-грамматики.

Ясно, что предложенная в предыдущем параграфе универсальная грамматика слишком "сильна" - обеспечивая все разумные и полезные преобразования, она будет допускать и много лишних, не имеющих никакой естественной интерпретации. Это связано, в частности, с тем, что для представления структуры фраз естественных языков используются не произвольные деревья, а деревья более специальных типов, задаваемых ограничениями на характер ветвления и на размещение пометок на ветвях. Рассмотрим, например, следующее ограничение: для каждой пометки a_i указывается такое число n_i, что из любого узла исходит не более n_i ветвей с пометкой a_i (таким образом, из каждого узла исходит всего не более $n_1 + n_2 + \ldots + n_k$ ветвей,

где k – общее число разных пометок). Возможны и другие ограничения, однако мы прибегнем к ограничениям именно такого типа, поскольку в упоминавшейся системе перифразирования рассматриваются глубинные синтаксические структуры, в которых из каждого узла может исходить не более чем по одной ветви типов 1, 2, 3, 4 и 6 и не более некоторого фиксированного числа ветвей типа 5: см. [1], стр. 183.

Указанное ограничение удобно формулировать следующим образом. Рассмотрим матрицу

$$\varphi = \begin{pmatrix} a_1, & a_2, & \ldots, & a_k \\ n_1, & n_2, & \ldots, & n_k \end{pmatrix} ,$$

где $a_1, a_2, \ldots, a_k$ – символы пометок на ветвях и $n_1, n_2, \ldots, n_k$ – натуральные числа. Дерево, ветви которого помечены символами из словаря $\{a_1, a_2, \ldots, a_k\}$, мы будем называть φ-правильным, если для любого $i = 1, 2, \ldots, k$ из любого узла этого дерева исходит не более n_1 ветвей, помеченных символом a_1.

Очевидно, что для целей представления структуры фраз естественного языка множество φ-правильных деревьев оказывается более адекватным, нежели множество произвольных деревьев. Если мы хотим, чтобы синтаксические Δ-грамматики также были как можно более адекватны для описания синтаксических преобразований, то необходимо, чтобы эти грамматики обеспечивали сохранение φ-правильности перерабатываемых деревьев. Для этого, прежде всего, естественно потребовать, чтобы в элементарных преобразованиях левые и правые части были φ-правильными деревьями; такие элементарные преобразования мы будем называть φ-правильными.

Синтаксическая Δ-грамматика будет, по определению, φ-правильной, если все ее правила являются φ-правильными элементарными преобразованиями.

Вывод в φ-правильной синтаксической Δ-грамматике называется φ-выводом, если он состоит из φ-правильных деревьев; соответственно определяется φ-выводимость (ср. стр. I4).

Замечание 1. Понятия φ-вывода и φ-выводимости нужны потому, что φ-правильность элементарного преобразования сама по себе сохранение φ-правильности перерабатываемых деревьев еще не гарантирует.

Замечание 2. Иногда мы будем говорить просто "правильный" вместо "φ-правильный".

Таким образом, мы ввели некоторый специальный класс синтаксических Δ-грамматик: φ-правильные синтаксические Δ-грамматики. Однако главное здесь не столько во введении нового класса, сколько в изменении аспекта рассмотрения - теперь нас будут интересовать не произвольные выводы в грамматиках, а только φ-выводы.

На стр. I7-I8 исследовался вопрос о моделировании произвольных элементарных преобразований специальными (элементарными) преобразованиями. Естественно рассмотреть этот вопрос и для φ-правильных преобразований. Это, однако, требует некоторой переформулировки понятия моделирования (стр. I8). Именно, пусть $m = t_1 \Rightarrow t_2 \mid f$ - φ-правильное элементарное преобразование и M - некоторое множество φ-правильных элементарных преобразований; будем говорить, что преобразование m φ- моделируется преобразованиями из M,

если существует такая конечная последовательность $m_1, m_2, \ldots, m_n \in M$, что каковы бы ни были φ-правильные деревья T и T', где T' получается из T применением m, то T' может быть получено из T также и последовательным применением $m_1, m_2, \ldots, m_n$, и притом так, что все промежуточные деревья являются φ-правильными.

Теперь мы можем сформулировать для φ-правильных элементарных преобразований теорему, аналогичную теореме 1 (стр. 19):

Теорема 1'.

Пусть $\varphi = \begin{pmatrix} a_1, & a_2, & \ldots, & a_k \\ n_1, & n_2, & \ldots, & n_k \end{pmatrix}$, $\varphi^1 = \begin{pmatrix} a_1, & a_2, & \ldots, & a_k, & o \\ n_1, & n_2, & \ldots, & n_k, & 1 \end{pmatrix}$

где o – символ, отличный от всех a_i, $1 \leqslant i \leqslant k$. Тогда произвольное φ-правильное элементарное преобразование φ^1-моделируется φ^1-правильными специальными элементарными преобразованиями.

(Разумеется, всякое φ-правильное дерево является и φ^1-правильным.)

Содержательный смысл теоремы 1' состоит в следующем. Пусть для описания синтаксиса некоторого языка используются, в соответствии с некоторыми чисто лингвистическими соображениями, деревья с определенным набором пометок на ветвях и определенными ограничениями на ветвление. Оказывается, что если добавить к этому набору пометок еще одну – искусственную синтаксическую связь, не имеющую лингвистической интерпретации, то при переработке подобных деревьев можно обойтись только специальными преобразованиями.

Естественно спросить: не будут ли достаточны специальные преобразования и без введения дополнительной (искусственной) связи? Ответ на этот вопрос дает

Теорема 2. а) Если $n_1 + \ldots + n_k \geqslant 3$ или $n_1 + \ldots + n_k = 1$, то произвольное

$\begin{pmatrix} a_1, & \ldots, & a_k \\ n_1, & \ldots, & n_k \end{pmatrix}$ – правильное преобразование

моделируется

$\begin{pmatrix} a_1, & \ldots, & a_k \\ n_1, & \ldots, & n_k \end{pmatrix}$ – правильными специальными преобразованиями.

б) Существуют

$\begin{pmatrix} a_1, & a_2 \\ 1, & 1 \end{pmatrix}$ – правильные и $\begin{pmatrix} a \\ 2 \end{pmatrix}$ – правильные элементарные преобразования, не моделируемые соответственно

$\begin{pmatrix} a_1, & a_2 \\ 1, & 1 \end{pmatrix}$ – правильными и $\begin{pmatrix} a \\ 2 \end{pmatrix}$ – правильными специальными преобразованиями.

Теорема 2 позволяет указать еще один из возможных подходов к проблеме, издавна занимающей и продолжающей занимать умы многих лингвистов, – а именно, к проблеме построения универсального синтаксиса, понимаемого как возможность описывать в общем виде все мыслимые в естественных языках синтаксические структуры и их преобразования* . Так, некоторое исчисление синтаксических трансформаций сравнительно недавно было предложено С.К.Шаумяном (см., например, [14]). В связи с популярностью идеи универсального синтаксиса представляется небесполезным

---

* То же самое верно и для теорем 1 и 1′ (ср. стр. 24). Однако теорема 2 обеспечивает лучший подход к универсальному син-

выписать универсальную правильную синтаксическую Δ-грамматику в явном виде.

Если фиксировать любым способом ограничения на ветвление и набор синтаксических связей, то универсальная правильная грамматика строится тривиально. Мы будем строить ее для глубинных структур, подчиняющихся ограничениям, указанным на стр. 22, т.е. для глубинных структур, введенных в [1] – [4]. Именно, различается шесть типов синтаксических связей: 1, 2, 3, 4, 5, 6, причем из каждого узла может исходить не более, чем по одной связи типов 1–4 и 6 и не более, чем 10 связей типа 5; каждая из связей 1–4 интерпретируется как связь между предикатом и соответствующим аргументом (*), 5 – как связь между определяемым и определением (в самом широком смысле слова "определение") и 6 – как связь между двумя последовательными однородными единицами. Число 10 выбрано в качестве заведомо достаточной верхней оценки для количества возможных при одном определяемом разнородных определений.

таксису, так как теорема 1 ничего не говорит о правильности, а теорема 1′ предполагает введение искусственной неинтерпретируемой связи.

(*) Таким образом, рассматриваются предикаты, имеющие не более четырех мест. Это объясняется тем содержательным соображением, что в естественном языке слова, обозначающие предикаты с большим числом мест, по-видимому, не встречаются. Пример "четырехместного" слова: наказать – кто (субъект), кого (объект), за что (мотивировка) и чем (содержание наказания), ср. Дирекция наказала его за многочисленные опоздания понижением в должности.

Итак, универсальная $\begin{pmatrix} 1, & 2, & 3, & 4, & 5, & 6 \\ 1, & 1, & 1, & 1, & 10, & 1 \end{pmatrix}$ – правильная синтаксическая Δ-грамматика Γ имеет словарь $V_U = \{1, 2, 3, 4, 5, 6\}$ (набор глубинных синтаксических связей) и множество правил Π_U, состоящее из следующих 80 правил:

1. 12 правил расщепления вида а) $A \Rightarrow 1(B, C) \mid f(A) = B$ и б) $A \Rightarrow 1(B, C) \mid f(A) = C$ $(1 = 1, \ldots, 6)$;

2. 62 правила перевешивания вида а) $1(A, B)\cdot j(B, C) \Rightarrow 1(D, E)\cdot j(D, F)$ и б) $1(A, B)\cdot j(A, C) \Rightarrow 1(D, E)\cdot j(E, F)$; здесь $1, j = 1, \ldots, 6$, причем либо $1 \neq j$, либо $1 = j = 5$;

3. 6 правил склеивания вида $1(A, B) = C (1 = 1, \ldots, 6)$.

Указанные правила покрывают все мыслимые трансформации естественных языков. Точнее говоря, нельзя a priori утверждать, что каждому из 80 правил отвечает наблюдаемое в том или ином естественном языке синтаксическое преобразование; однако, если принять предложенный способ описания глубинных синтаксических структур, то ни в одном естественном языке не может быть обнаружена синтаксическая трансформация, которая не представлялась бы в виде некоторой комбинации (последовательно выполняемых) элементарных преобразований из набора Π_U.

Очевидно, впрочем, что грамматика Γ_U позволяет описывать лишь чисто синтаксический аспект трансформаций в естественных языках. Однако в речевой практике (все виды перевода, сжатие текста, синонимическое перифразирование и т.п.) главную роль играют трансформации, затрагивающие не только синтаксическое строение текста, но и его лексический состав. Для формализации таких трансформаций требуется более богатая модель.

§ 5.

Лексико-синтаксические Δ-грамматики

Будем рассматривать правильные деревья, узлам которых сопоставлены символы из некоторого конечного множества W - словаря характеризованных обобщенных лексем. Такие объекты мы будем называть заполненными правильными деревьями. С содержательной точки зрения элементы словаря W - характеризованные обобщенные лексемы - представляют собой символы для слов, идиоматических словосочетаний и лексических функций ([I], стр. I86), снабженные индексами тех морфологических категорий, которые несут семантическую информацию (например, число у существительного, время, вид и наклонение у глагола, степень сравнения у прилагательного).

Заполненное дерево задает (с точностью до "поверхностных" синтаксических и лексических различий) некоторую реальную фразу. Очевидно, что описание трансформаций во всех их аспектах (с учетом лексических преобразований) сводится к построению грамматик еще одного типа, которые имели бы дело не просто с правильными деревьями, а с заполненными правильными деревьями. Такие грамматики мы назовем [правильными] лексико-синтаксическими Δ-грамматиками.

Правильные синтаксические Δ-грамматики могут рассматриваться как частный случай правильных лексико-синтаксических Δ-грамматик - когда словарь характеризованных обобщенных лексем состоит из одного элемента, так что все узлы деревьев заполнены одним и тем же символом (что эквивалентно отсутствию

заполнения) .

Говоря формально правильная лексико-синтаксическая Δ-грамматика есть упорядоченная четверка $\langle W, V, \varphi, \Pi \rangle$, где: W и V – конечные множества символов (W – словарь пометок на узлах, т.е. характеризованных обобщенных лексем, $V = \{a_1, \ldots, a_k\}$ – словарь пометок на ветвях, т.е. типов синтаксических связей); φ – матрица $\begin{pmatrix} a_1, & \ldots, & a_k \\ n_1, & \ldots, & n_k \end{pmatrix}$, т.е. отображение множества V во множество натуральных чисел (ср. стр. 22); Π – конечное множество элементарных преобразований, определяемых ниже.

Дерево, в узлах которого стоят символы из W, а ветви помечены символами из V, назовем W/V – деревом. Символ, стоящий в узле α некоторого W/V -дерева, обозначается $\lambda(\alpha)$. Композиция W/V – дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и W/V – деревьев $t_1, t_2, \ldots, t_n$ определяется так же, как для "незаполненных" деревьев (стр. 10), со следующим добавочным требованием: для каждого $i = 1, 2, \ldots, n$ пометка в узле α_i результирующего дерева совпадает с пометкой в вершине дерева t_1 (а не с пометкой в узле α_i дерева t_0 !). Для композиции W/V-деревьев мы сохраним обозначение, введенное на стр. 10.

---

* Лексическо-синтаксические Δ-грамматики являются самым общим классом Δ-грамматик из рассматриваемых нами. Поэтому общепринятому способу построения терминов лучше отвечало бы название "Δ-грамматики". Мы, однако, сохранили избыточное прилагательное "лексико-синтаксические" в связи с использованием аналогичного термина в принятой нами интерпретации.

W/V – дерево является, по определению, φ-правильным, если, каково бы ни было $a \in V$, из любого его узла выходит не более $\varphi(a)$ ветвей с пометкой a.

Элементарным преобразованием в данной грамматике Г называется упорядоченная тройка $\langle t_1, t_2, f \rangle$ (или $t_1 \Rightarrow t_2 \mid f$), где t_1 и t_2 – правильные W/V –деревья и f – отображение множества узлов дерева t_1 во множество узлов дерева t_2.

Там, где это понадобится по ходу изложения, мы будем говорить о только что определенных элементарных преобразованиях как о $\begin{pmatrix} a_1, a_2, \ldots, a_k \\ n_1, n_2, \ldots, n_k \end{pmatrix}$ – правильных.

W/V – дерево T' получается из правильного W/V – дерева T применением элементарного преобразования $t_1 \Rightarrow t_2 \mid f$, если T и T' представимы в виде

$T = C(T^0; \alpha_0 \mid C(t_1; \alpha_1, \alpha_2, \ldots, \alpha_n \mid T^1, T^2, \ldots, T^n))$ и

$T' = C(T^0; \alpha_0 \mid C(t_2; f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n) \mid T^1, T^2, \ldots, T^n))$,

где α_0 – концевой узел дерева T^0, $\alpha_1, \alpha_2, \ldots, \alpha_n$ – бесповторный пересчет узлов дерева t_1 и дерево T' правильно.

Разумеется, результат применения элементарного преобразования определяется – как и для "неаполненных" деревьев – с точностью до изоморфизма; определение изоморфизма W/V – деревьев может быть получено из определения на стр. 9 введением дополнительного требования: если α_2 – образ узла α_1, то $\eta(\alpha_1) = \eta(\alpha_2)$.

φ-вывод и φ-выводимость в правильной лексико-синтаксической Δ-грамматике определяются в точности так же, как для синтаксической Δ-грамматики, см. стр. 22.

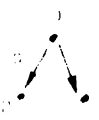
В классе правильных лексико-синтаксических Δ-грамматик (или, короче, ЛС-грамматик) можно выделить подкласс, совершен-

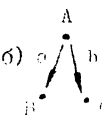
по аналогичные подклассам синтаксических грамматик (стр. 17); соответствующие определения при этом по существу не меняются.

Если же мы захотим сохранить для случая ЛС-грамматик специальные преобразования (стр. 17-18), нам будет удобно ввести дополнительное соглашение. Именно, будем считать, что словарь V всегда содержит некоторый особый символ w_0 – "пустую лексему"; тогда определения преобразований типов 1-3 могут быть переформулированы для W/V -деревьев следующим образом.

1. Расщепление узла: $\overset{A}{\bullet} \Rightarrow \overset{B}{\bullet} \xrightarrow{a} \overset{C}{\bullet}$, причем либо а) $\lambda(B) = \lambda(A), \lambda(C) = w_0$, $f(A) = B$, либо б) $\lambda(A) = \lambda(C)$, $\lambda(B) = w_0$, $f(C) = C$.

2) Перевешивание узла (ветви): а) $\overset{A}{\bullet} \xrightarrow{a} \overset{B}{\bullet} \xrightarrow{b} \overset{C}{\bullet} \Rightarrow$



и б)  $\Rightarrow \overset{D}{\bullet} \xrightarrow{a} \overset{E}{\bullet} \xrightarrow{b} \overset{F}{\bullet}$, причем в обоих случаях

$\lambda(A) = \lambda(D), \lambda(B) = \lambda(E), \lambda(C) = \lambda(F)$, $f(A)=D, f(B)=E, f(C)=F$.

3) Склеивание узлов: $\overset{B}{\bullet} \xrightarrow{a} \overset{C}{\bullet} \Rightarrow \overset{A}{\bullet}$, причем либо а) $\lambda(B) = \lambda(A)$, либо б) $\lambda(C) = \lambda(A)$; в обоих случаях $f(B) = f(C) = A$.

Кроме того, для W/V -деревьев необходим еще один тип специальных преобразований:

4) Переименование узла: $\overset{w_i}{\bullet} \Rightarrow \overset{w_j}{\bullet}$, где $w_i, w_j \in V$[*].

Теоремы 1′ и 2 переносятся на случай ЛС-грамматик без изменений.

Можно предложить некоторый способ записи лексико-синтаксических Δ-грамматик – мы назовем его нормальной формой, который представляется удобным, по крайней мере, в содержа-

[*] Пустая лексема вводится лишь ради содержательной естест-

тельном отношении. Точнее говоря, в нормальной форме записываются элементарные преобразования. Нормальная форма есть упорядоченная пара $\langle \pi, \psi \rangle$, где π – элементарное преобразование деревьев, узлы которых помечены натуральными числами, (быть может, с повторениями) и ψ – отображение некоторого конечного множества натуральных чисел в W. Обычная запись элементарного преобразования получается из нормальной формы следующим заполнением узлов в обеих частях преобразования π: если номер n некоторого узла принадлежит области определения ψ, то в этот узел ставится обобщенная лексема $\psi(n)$; в противном случае в узел ставится произвольная обобщенная лексема, но так, чтобы в узлах, имеющих одинаковые номера, оказались одинаковые лексемы. Таким образом, одна нормальная форма соответствует, вообще говоря, многим элементарным преобразованиям. Содержательно, запись элементарного преобразования в нормальной форме означает разделение этого преобразования на синтаксический (π) и лексический (ψ) компоненты, что соответствует разделению правил перифразирования на лексические и синтаксические ([1], стр. 189 и сл.). Такое разделение повышает объяснительную силу описания, позволяя фиксировать сходства и различия трансформаций отдельно в лексическом и

венности; можно обойтись без нее, разрешив при этом брать в преобразованиях типа 1 а в качестве $\downarrow(C)$ и в преобразованиях типа 1 б в качестве $\downarrow(B)$ любую характеризованную обобщенную лексему. Кстати, в этом случае можно не вводить в число специальных преобразований переименование узла: оно тогда заменимо одним преобразованием типа 1 б и одним типа 3 а.

отдельно в синтаксическом аспекте. Кроме того, представление в нормальной форме приводит к большей компактности записи: вместо списка элементарных преобразований П используются три списка: список первых компонент соответствующих нормальных форм (т.е. список всех π), список вторых компонент (список всех ψ) и список соответствий между номерами разных π , с одной стороны, и номерами разных ψ , с другой (т.е. список указаний, какие лексические заполнения возможны для тех или иных синтаксических правил – или, что то же самое, какие синтаксические правила необходимы при тех или иных лексических заменах). Эти три списка должны иметь меньший суммарный объем, чем список П, так как разные π комбинируются с разными ψ многими разными способами, не говоря уже о том, что одна нормальная форма может представить много элементарных преобразований.

Здесь заканчивается изложение основных формальных понятий, относящихся к Δ-грамматикам. Лингвистической интерпретации этих понятий авторы надеются посвятить отдельную статью.

Литература

1. Жолковский А.К., Мельчук И.А. О семантическом синтезе. "Проблемы кибернетики", вып. 19, 1967, 177–238.

2. Жолковский А.К., Мельчук И.А. О системе семантического синтеза. I. Строение словаря. "НТИ", 1966, № 11, 48–55.

3. Жолковский А.К., Мельчук И.А. О системе семантического синтеза. II. Правила перифразирования, "НТИ", 1967, № 12–27.

4. Жолковский А.К., Мельчук И.А. К построению действующей модели языка "смысл ⟷ текст". "Машинный перевод и прикладная лингвистика", вып. 11, 1969, 5–35.

5. Зализняк А.А., Падучева Е.В. О связи языка лингвистических описаний с родным языком лингвиста. "Программа и тезисы докладов в летней школе по вторичным моделирующим системам". Тарту, 1964, 7-9.

6. Tesnière L. Éléments de syntaxe structurale. P., 1959.

7. Падучева Е.В. О способах представления синтаксической структуры предложений. "Вопросы языкознания", 1964, № 2, 99-113.

8. Бабицкий К.И. О синтаксической синонимии предложений в естественных языках. "НТИ", 1965, № 6, 29-34.

9. Chomsky N. Aspects of the theory of syntax. Cambridge (Mass.), 1965.

10. Katz J.J., Fodor J.A. The structure of the semantic theory. "Language", 1963, 39, №1, 170-210.

11. Katz J.J., Postal P. An intergrated theory of linguistic descriptions. Cambridge (Mass.), 1964.

12. Weinreich U. Explorations in semantic theory. - In: "Current trends in linguistics. Vol. III: Theoretical foundations. London - The Hague - Paris, 1966, 395-477.

13. Sgall P. Generativní popis jazyka a česká deklinace. Praha, 1967.

14. Шаумян С.К. Основы структурной лингвистики. М., 1965.

15. Арапов М.В., Борщев В.Б. Синтаксическая модель языков с богатой флективной системой и свободным порядком слов. - В: "Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации. II том. Семиотические проблемы автоматизированной обработки информации". М., 1967, 16-46.

16. Ломковская М.В. Исчисление, порождающее ядерные русские предложения. "НТИ", 1965, № 7, 35–39; № 9, 37–40; 1966, № 11, 56–65.

17. Колмогоров А.Н., Успенский В.А. К определению алгоритма. Успехи матем. наук, 1958, 13, № 4, 3–28.

18. Арапов М.В., Борщев В.Б., Шрейдер Ю.А. Язык, грамматика, семантика. – В: "Труды III Всесоюзной конференции...", М., 1967, 5–15.

19. Hockett Ch. F. Language, mathematics, linguistics. s'-Gravenhage, 1967.

20. Veillon G., Veyrunes J., Vauquois B. Un métalangage de grammaires transformationelles. – In: "2me Conférence international sur le traitement automatique des langues". Grenoble, 1967.

21. Гаврилова Т.Л. О структурном анализе вьетнамского текста и одном способе записи его результатов. – "Проблемы кибернетики", вып. 13, 1965, 201–220.

22. Варшавская А.И. Опыт построения графа, порождающего некоторые структуры простого английского предложения. "НТИ", серия 2, 1969, № 4, 15 – 20.

23. Кулагина О.С. Некоторые вопросы преобразования деревьев зависимостей. М., 1969. 29 стр. (ИММ АН СССР. Препринт № 12).

24. Hays D. Dependency theory: a formalism and some observations. "Language", 1964, N 4, 511-525.

25. Фитиалов С.Я. Об эквивалентности грамматик НС и грамматик зависимостей. – В кн.: Проблемы структурной лингвистики.

26. Белецкий М.И. Бесконтекстные и доминационные грамматики и связанные с ними алгорифмические проблемы. "Кибернетика", 1967, № 4, 90–97.

27. Уорт Д.С. Об отображении линейных отношений в порождающих моделях языка. "Вопросы языкознания", 1964, № 5, 46–58.

БИБЛИОТЕКА
ИНСТИТУТА РУССКОГО ЯЗЫКА
АКАДЕМИИ НАУК СССР

Т-02550 от 12.02.70г. Тип. МГПИИЯ. Объем 2,5 п.л. Зак. 90-100