А
П-71 10

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

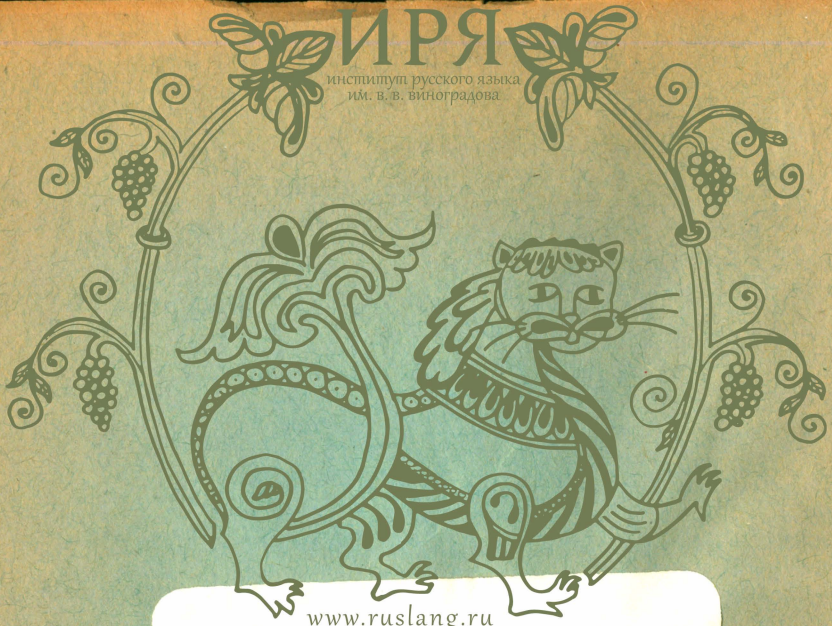
Выпуск 10

Т. Д. КОРЕЛЬСКАЯ, Е. Д. ПАДУЧЕВА

О ФОРМАЛЬНОМ АППАРАТЕ СИНТАКСИЧЕСКИХ ПРЕОБРАЗОВАНИЙ

Москва — 1970

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

65110

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР
Проблемная группа по экспериментальной и
и прикладной лингвистике

Предварительные публикации
Выпуск 10

Т.Д.Корельская, Е.Д.Падучева

О ФОРМАЛЬНОМ АППАРАТЕ СИНТАКСИЧЕСКИХ ПРЕОБРАЗОВАНИЙ

Москва - 1970

4/П-71 10

БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
ИМЕНИ ЖЕЛКОВЛЯ
АН СССР

65110

§ 1. Вводные замечания

Имеется большое число лингвистических задач, сводящихся к описанию тех или иных содержательно характеризуемых множеств преобразований предложения - преобразований, переводящих предложение из орфографической записи в фонетическую; преобразований, меняющих структуру предложения, но оставляющих неизменным его смысл; преобразований, меняющих смысл предложения заданным образом, - в частности, сужающих смысл (реферирование) или расширяющих смысл (вывод следствий) - и т.д. Некоторые из преобразований такого рода мы назовем в дальнейшем синтаксическими.

Преобразования естественно применять не к самим предложениям, а к определенного рода представлениям (структурам) предложений. Мы будем различать (опуская те уровни представления предложений, которые оперируют с незначимыми единицами, например, фонемный): морфологическое представление предложения, синтаксическое и семантическое[x].

Морфологическое представление указывает: (1) лексемы, из которых составлено предложение (с различением повторных вхождений одной лексемы, если они есть); (2) значения словоизменительных грамматических категорий, задающие для каждой лексемы некоторую ее словоформу; например, для лексемы угол указание "ед.число, твор.падеж" задает словоформу углом; для лексемы пересекать указание "адъективная репрезентация (термин А.И.Смирницкого [12]), активный залог, наст.время, жен. род, ед.число, род.падеж" задает словоформу пересекающей; (3) порядок, в котором должны быть расставлены словоформы.

x) Описываемая ниже система соответствий между уровнями представления предложения опирается во многом на работу А.В.Гладкого и И.А.Мельчука [2] и отчасти на работы Н.Хомского [8] и П.Стелла [10].

Имеющиеся у нас сведения о том, что такое семантическое представление предложения, весьма скупы: семантическое представление должно быть чем-то, что одинаково у всех (или по крайней мере у многих) предложений, имеющих одинаковый смысл, и различно у предложений, имеющих разный смысл. Оставляя понятие тождества и различия смысла в целом неопределенным и интуитивным, введем только одно уточнение. Мы не будем считать различием по смыслу то различие в понимании предложения, которое можно назвать различием в актуальном членении; например, предложения (1) Некоторые равнобедренные треугольники являются прямоугольными; (2) Некоторые прямоугольные треугольники являются равнобедренными и (3) Некоторые треугольники являются прямоугольными и равнобедренными считаются имеющими одинаковый смысл, по-разному актуализованный. По крайней мере для ограниченных областей применения языка понятию семантического представления предложения вполне можно придать удовлетворительный точный смысл, см., например, [4].

Синтаксическое представление является промежуточным между морфологическим и семантическим, ср. промежуточный язык в [4]; оно призвано выразить сложное, в обе стороны неоднозначное соответствие между множеством морфологических и множеством семантических представлений языка. Синтаксическое представление можно назвать, используя терминологию П.Сгалла [10], представлением "на уровне членов предложения". В синтаксическом представлении фигурируют из всех лексем предложения только такие, которые мы признаем неслужебными[x], а их всех словоизменительных грамматических категорий только такие, которые мы признаем независимыми (т.е. такие, у которых выбор значения не определяется однозначно согласованием с какой-либо грамматической - не обязательно словоизменительной - категорией синтаксически связанного слова; ср. более широкое понятие несинтаксической грамматической категории у А.М.Пешковского); например, категория рода, числа и лица гла-

x) Точнее было бы говорить здесь не о лексемах, а об "обобщенных лексемах", как в [7], относя к обобщенным лексемам, помимо собственно лексем, например, неразложимые фразеологические сочетания типа каков бы ни был.

гола в русском языке считаются зависимыми, а категории времени, залога, наклонения, репрезентации - независимыми. Сведения о служебных словах, значениях зависимых грамматических категорий, о расположении слов и других "грамматических средствах" заменены указаниями о синтаксических связях (подчинениях) между знаменательными лексемами и о синтаксической роли знаменательных лексем (синтаксическая роль - примерно то же,что член предложения).

Предполагается, что соответствие между морфологическими и семантическими представлениями языка можно описать системой преобразований, которая устроена, в общем виде, так.

(I) Имеются преобразования, позволяющие перейти от каждого синтаксического представления к одному - или нескольким, но немногим - морфологическим; эти преобразования определяют выбор служебных слов, относительное расположение слов, выбор управляемых и согласованных значений грамматических категорий, и т.д. Например, синтаксическое указание о том, что лексема вероятный является в предложении сказуемым в наст.времени, может быть преобразовано в морфологическом представлении в указание о том, что лексема вероятный имеет предикативную репрезентацию (= краткую форму) и, в зависимости от контекста, например, муж.род, ед.число (что соответствует словоформе вероятен), или в указание о том, что она имеет адъективную репрезентацию и служебное связочное слово (что соответствует сочетанию является вероятным).

(2) Имеются преобразования, которые переводят одно в другое все синтаксические представления, синонимичные друг другу, т.е. такие, которым соответствует одно и то же семантическое представление; это, например, преобразования активных конструкций в пассивные, предикативных в определительные, сложносочиненных предложений в предложения с однородными членами и многие другие.

(3) Имеются преобразования, позволяющие перейти от семантического представления к одному - или нескольким, но немногим - синтаксическим. Впрочем, можно, как в [4], считать, что множество семантических представлений языка полу-

чается в результате того, что в каждом классе синонимичных синтаксических представлений выделяется одно "главное", которое и объявляется семантическим. Тем самым соотношения между представлениями разных уровней становятся довольно простыми, а все наиболее сложные аспекты соответствия между морфологическими и семантическими представлениями описываются преобразованиями внутри множества синтаксических представлений.

Для понимания последующих формальных определений укажем одно требование, которое естественно предъявить к синтаксическим представлениям, предназначенным для описания синонимических соотношений. Синонимическое соотношение естественно считать симметричным; отсюда возникают трудности в описании синонимических соотношений в тех случаях, когда одно из предложений однозначно (с точки зрения принятой системы семантических представлений), а другое - нет. Чтобы синтаксические представления были удобны для описания синонимических соотношений, нужно, чтобы они были однозначными. Например, неоднозначному предложению (4) Если Петя видел Ваню, то он (то ли Петя, то ли Ваня) все знает должно соответствовать два разных синтаксических представления.

Синтаксическими преобразованиями естественно считать произвольные преобразования, применяемые к синтаксическим представлениям и переводящие одни синтаксические представления в другие. Таким образом, прежде чем говорить о преобразованиях, мы дадим формальное определение понятия синтаксического представления. Это определение формализует содержательное понятие синтаксического представления предложения, возникшее в ходе описания синонимических преобразований в некотором фрагменте русского языка - в стандартизованном языке геометрии (для которого языком семантических представлений служит язык прикладного исчисления предикатов [4]). Не исключено, что из рассмотрения иных языков и иных задач возникнет иное понятие синтаксического представления предложения. Наше описание синтаксических преобразований действительно для любых синтаксических представлений, которые обладают определенным набором формальных свойств.

Изучению формального аппарата для описания синтаксических преобразований посвящены работы Г.Вейона, К.Вейрюна и Б.Вокуа [1], И.А.Мельчука и А.В.Гладкого [2] и О.С.Кулагиной [3]. Данная работа близка к ним по духу и использует некоторые их идеи*).

§ 2. Синтаксическое представление предложения

Пусть для некоторого языка задан словарь лексем V_0, словарь синтаксических ролей лексем V_1 и словари V_2, ..., V_n значений независимых словоизменительных грамматических категорий, характеризующих те или иные из словоформ, принадлежащих к одной лексеме. Тогда синтаксическим представлением для предложения этого языка называется объект Q вида $\langle M, W, \mathcal{D}, \mathcal{L}, E \rangle$, где: а) M – непустое конечное множество; б) W – система частичных отображений $W_0, \ldots, W_n$ множества M в словари $V_0, \ldots, V_n$ (при этом W_0 и W_1 – полные отображения, а W_i при $i \geq 2$ зависит от W_0 в том смысле, что для $m \in M$ $W_i(m)$ определено или нет, в зависимости от значения $W_0(m)$); в) $\mathcal{D}$, E и $\mathcal{L}$ – некоторые отношения, заданные на M (отношением – бинарным – в множестве M мы будем называть любое подмножество множества $M \times M$, т.е. любое множество упорядоченных пар элементов из M).

Отношение $\mathcal{D}$ называется подчинением и удовлетворяет следующим ограничениям: (1) однозначность по второму аргументу; $(\forall m, n, k \in M)[\mathcal{D}(m,n) \,\&\, \mathcal{D}(k,n) \rightarrow m = k]$; (2) единственность вершины; $(\exists! n \in M)(\exists m \in M)\mathcal{D}(m,n)$; (3) отсутствие циклов; ни для какого $m \in M$ не существует

---

*) Авторы искренне благодарны своим коллегам Л.А.Бассалыго, В.М.Борщеву, А.В.Гладкому, О.С.Кулагиной, Э.Д.Стоцкому и М.В.Хомякову, которые прочли эту работу и немало способствовали ее улучшению.

$m_1, \ldots, m_k \in M$ таких, что $m_1 = m_k = m$ и для всякого i, $1 < i < k$, $\mathcal{D}(m_i, m_{i+1})$. Иначе говоря, $\mathcal{D}$ – это отношение, граф которого является деревом.

Отношение $\mathcal{E}$ называется равносильностью и является отношением типа эквивалентности, т.е. рефлексивно, симметрично и транзитивно.

Отношение $\mathcal{L}$ называется линейным порядком; оно удовлетворяет условиям (1)-(3) из определения подчинения и, кроме того, тому условию, что $(\exists' n \in M)(\exists\, m \in M)\ \mathcal{L}(n, m)$. Иначе говоря, $\mathcal{L}$ – это отношение, граф которого является цепью.

Предполагается, что всякому предложению рассматриваемого языка может быть поставлен в соответствие объект Q описанного вида, который и будет синтаксическим представлением данного предложения (или несколько объектов, каждый из которых является синтаксическим представлением).

Содержательный смысл компоненты M из Q состоит в том, что число элементов множества M равно числу вхождений лексем в представляемое предложение (т.е. M – это множество "мест" в предложении, на которых стоят лексемы). Если для элемента $m \in M$ задан его W_0-образ, то он содержательно соответствует вхождению некоторой лексемы в предложение. Назначение отображения W_1 состоит в том, чтобы указывать для вхождения лексемы его синтаксическую роль. Отображения $W_2, \ldots, W_n$ ставят в соответствие элементам из M значения независимых словоизменительных грамматических категорий; например, W_i, может быть отображением элементов множества M в словарь залогов, состоящий из трех элементов – активного, пассивного и взаимного залогов.

Отношение $\mathcal{D}$ соответствует синтаксическим связям между элементами: $\mathcal{D}(m, n)$ означает, что элемент m подчинлет элемент n, (или, иначе говоря, что n зависит от m). В силу свойств (1) и (2), отношение $\mathcal{D}$ задает однозначное отображение множества M в себя (для всех элементов $n \in M$, кроме вершины). Мы будем обозначать через

$\mathcal{d}(n)$ элемент m такой, что $\mathcal{D}(m,n)$ (т.е. элемент, подчиняющий n), и через $\mathcal{d}^{-1}(m)$ множество элементов $n \in M$ таких, что $\mathcal{D}(m,n)$ (т.е. множество элементов, подчиненных m). Для единообразия мы будем считать, что для случая, когда n - вершина, $\mathcal{d}(n)$ тоже определено, а именно, равно условному элементу $*$ (естественно, не принадлежащему M).

В силу того, что для всякого $n \in M$ однозначно определено $\mathcal{d}(n)$, W-образ, т.е. синтаксическую роль элемента n можно интерпретировать также как тип синтаксической связи, сопоставленный паре $(\mathcal{d}(n), n)$, ср. [2] и [3] .

Содержательный смысл компоненты E состоит в задании анафорических связей между лексемами; в частности, если элементы m и n являются термами, то $E(m,n)$ означает, что соответствующие лексемы соотнесены с одним и тем же внеязыковым объектом. При определенных ограничениях на синтаксическую структуру термов в языке (например, если в состав всякого имени объекта, кроме собственного, входит индекс, позволяющий отличить его от других объектов того же рода, ср. [5]), отношение равносильности может не включаться в Q , а определяться простым алгоритмом по M, W и $\mathcal{D}$. Мы примем соглашение, что если E в данном Q в явном виде не задано, то в это Q входит то E, которое определяется этим алгоритмом (стандартное E). Такое соглашение упрощает синтаксические преобразования над Q : преобразование компоненты E можно не описывать, а получать, если нужно, E в преобразованном Q заново по прежнему алгоритму. В дальнейшем отношение E будет заменено другим более "сжатым".

Наконец, содержательный смысл компоненты $\mathcal{L}$ состоит в том, что она задает относительное расположение лексем в предложении: $\mathcal{L}(m,n)$ означает, что лексема, сопоставленная элементу m , непосредственно предшествует лексеме, сопоставленной элементу n [x].

---

x) Линейный порядок лексем включается в синтаксическое представление потому, что многие из преобразований, (см. сл.стр.)

Отношение $\mathcal{L}$ задает отображение M в M. Мы будем обозначать через $\ell(m)$ элемент $n \in M$ такой, что $\mathcal{L}(m,n)$, и через $\ell^{-1}(m)$ элемент $n \in M$ такой, что $\mathcal{L}(n,m)$.

Для предложений языка, который мы будем рассматривать, можно выделить некоторое стандартное $\mathcal{L}$, которое алгоритмически устанавливается по M, W и $\mathcal{D}$. Поэтому если $\mathcal{L}$ стандартно, то его можно не записывать в Q в явном виде, приняв соглашение о том, что если $\mathcal{L}$ в данном Q не указано, то в это Q входит стандартное $\mathcal{L}$. Это упрощает описание преобразований: если изменения в $\mathcal{L}$, автоматически сопутствующие некоторым изменениям в $\mathcal{D}$, таковы, что $\mathcal{L}$ остается стандартным, то они могут не описываться. Возможен алгоритм установления линейного порядка, который применяется к компонентам $M, W, \mathcal{D}$ и частично заданной компоненте $\mathcal{L}$. Таким образом, способом задания $\mathcal{L}$ в Q является частичное (в том числе пустое) $\mathcal{L}$.

(продолж. сн.) которые должны быть описаны, зависят от линейного порядка лексем – в том смысле, что при одном порядке слов в предложении они применимы к предложению, а при другом – нет. Например, при введении местоимений знание линейного порядка нужно для того, чтобы не поставить местоимение левее его антецедента; при введении определительных конструкций знание линейного порядка нужно для того, чтобы присоединить определительную конструкцию к первому упоминанию данного объекта в тексте, а не к повторному, ср. [5]. Другой пример существенности линейного прядка; предложение (2) равно по смыслу предложению (1), а предложение (3) – нет: (1) Для всякой точки А существует множество М такое, что точка А принадлежит множеству М = (2) Всякая точка принадлежит некоторому множеству ≠ (3) Некоторому множеству принадлежит всякая точка. По-видимому, линейный порядок служит некоторой компенсацией за отсутствующее в нашем синтаксическом представлении актуальное членение.

В языке, который мы будем рассматривать, на соотношение между $\mathcal{L}$ и $\mathcal{D}$ наложено ограничение, называемое проективностью [6]. Ниже мы покажем, что это ограничение позволяет, без всякой потери информации, заменить компоненту $\mathcal{L}$ некоторой более удобной.

Элементы M мыслятся названными посредством букв некоторого алфавита.

Приведенное описание структуры отдельных компонент синтаксического представления и ограничений на их связь друг с другом не претендует на то, чтобы быть исчерпывающим: в каждом реальном языке на отдельные компоненты синтаксического представления и на их соотношение друг с другом наложено много ограничений, которые здесь не были описаны. Мы выявили только те свойства синтаксических представлений, которые будут использоваться при их преобразованиях.

x x

x

Всякому синтаксическому представлению Q может быть поставлено в соответствие производное представление Q^*, которое является результатом пополнения Q производными компонентами. Зададим, например, на словаре лексем V_0 частичные отображения $\beta_1, \dots, \beta_n$, сопоставляющие лексеме ее словарные характеристики, т.е. морфологические, синтаксические или семантические классы, к которым она принадлежит - часть речи, возможность участия в некотором преобразовании, значение, которое принимает на данной лексеме некоторая лексическая функция в смысле [7] (например, для лексемы больше может быть указана лексема, которая является ее конверсивом - меньше) и т.д. Тогда мы можем включить в Q^* систему производных отображений $W_1^1, \dots, W_k^1$, определяемую следующим образом: $W_i^1(m) = \beta_i(W_0(m))$. Полезно также включить в Q^* систему W^s частичных отображений M в множества синтаксических характеристик, причем для каждого W_i^s

должен быть алгоритм вычисления его значений в произвольном Q для всех элементов из M, входящих в его область определения. Например, синтаксической характеристикой слова может быть сведение о наличии прямо или опосредствовано подчиненного ему слова который.

Другая производная компонента, которую можно включить в Q^*, - это отношение $\mathcal{D}_S$ семантического подчинения, см. [5]. Отношение $\mathcal{D}_S$ не является подчинением. Оно вводится для того, чтобы преодолеть некоторые условности и натяжки, которые обычно имеют место при сведении всех синтаксических связей лексем в предложении - например, сочинительных связей - к подчинению. Так, в предложении Петя и Ваня отсутствуют подчинением связаны отсутствовать и и, и и Петя, и и Ваня; а семантическим подчинением - отсутствовать и Петя, отсутствовать и Ваня.

Введем теперь следующие операции, позволяющие из одних бинарных отношений получать другие. Это транзитивирование, (транзитивное замыкание), ограниченное транзитивирование, симметризация, рефлексивизация и отрицание отношения. Определения очевидны; приведем некоторые из них для примера. Результатом транзитивирования отношения R в множестве M называется отношение R^t такое, что для любых элементов m и $n \in M$ $R^t(m,n)$, если и только если существует подмножество M' множества M, состоящее из элементов $m_1,\dots,m_k$ таких, что $m_1 = m$, $m_k = n$ и для всякого i от 1 до $k-1$, $R(m_i, m_{i+1})$. Отношение $R^{t(x_{j1},\dots,x_{jk})}$ называется результатом ограниченного транзитивирования отношения R - с ограничением $(x_{j1},\dots,x_{jk})$, - если оно является результатом транзитивирования R и для всех m_i из M', не равных m, n, образы $W_{j1}(m_i),\dots,W_{jk}(m_i)$ равны, соответственно, $x_{j1},\dots,x_{jk}$. Обозначения: R^S, R^a - это отношения, являющиеся, соответственно, результатом рефлексивизации и симметризации R; $\bar{R}$ - это отрицание отношения R.

С помощью операций над отношениями можно иначе представить компоненту $\mathcal{L}$ из Q. А именно, поскольку $\mathcal{L}$ и $\mathcal{D}$ связаны отношением проективности, мы будем считать представлением линейного порядка в Q систему $\mathcal{L}_K$ отношений линейного порядка $\mathcal{L}_{K_i}$, заданных на подмножествах (пересекающихся) множества M, являющихся кустами относительно $\mathcal{D}$ (т.е. подмножествах K_i, каждое из которых имеет вид $\{i\} \cup d^{-1}(i)$, где $i \in M$). Для символов m и $n \in K_i$ имеет место $\mathcal{L}_{K_i}(m,n)$, если и только если m непосредственно левее n в K_i (т.е. $\mathcal{L}_{K_i}(m,n)$, если и только если $\mathcal{L}^\tau(m,n)$ и не существует $k \in K_i$ такого, что $\mathcal{L}^\tau(m,k)$ и $\mathcal{L}^\tau(k,n)$).

Ясно, что система $\mathcal{L}_K$ задает $\mathcal{L}$: для любых m и $n \in M$ имеет место $\mathcal{L}^\tau(m,n)$, если и только если существуют K_i и m' и $n' \in K_i$ такие, что одновременно имеет место $\mathcal{D}^{\tau 9}(m',m)$, $\mathcal{D}^{\tau 9}(n',n)$ и $\mathcal{L}^\tau_{K_i}(m',n')$; а $\mathcal{L}(m,n)$, если и только если $\mathcal{L}^\tau(m,n)$ и не существует k такого, что $\mathcal{L}^\tau(m,k)$ и $\mathcal{L}^\tau(k,n)$.

Через $\ell_{K_i}(m)$ мы будем обозначать символ $n \in K_i$, непосредственно следующий за m в K_i.

Спять-таки, принимается соглашение, что если для куста K_i отношение $\mathcal{L}_{K_i}$ не входит в $\mathcal{L}_K$, то $\mathcal{L}_{K_i}$ для K_i предполагается стандартным, причем само отношение $\mathcal{L}_{K_i}$ может быть определено не на всем K_i.

Отношение равносильности (если оно задается явно) также может быть представлено в Q в сжатом виде - через отношение анафории А: $A(m,n) \equiv E(m,n) \,\&\, \mathcal{L}^\tau(m,n) \,\&\, \neg \exists k (\mathcal{L}^\tau(m,k) \,\&\, \mathcal{L}^\tau(k,n))$; с другой стороны, $E(m,n) \equiv \equiv \mathcal{A}^{\tau\flat}(m,n)$.

Отношение А задает отображение (частичное) M в M. Через $a(n)$ будет обозначаться элемент $m \in M$ такой, что $A(m,n)$.

С помощью операций над отношениями мы можем дополнять Q различными производными компонентами; практически полезными для языка геометрии оказались следующие: $\mathcal{L}$ (в том случае,

если в $Q\mathcal{L}$ представлено через $\mathcal{L}_k$), $\mathcal{L}^{\tau}$, E (если в Q E представлено через A), $\mathcal{D}^{\tau}$, $\mathcal{D}^{\tau(\mathcal{S})}$ (ограниченное транзитивирование с ограничением: для всех $m_i \in M'$ W_0-образа m_j есть лексема $\mathcal{S}$; это отношение позволяет преодолеть условность, связанную с соглашением о бинарности связки $\mathcal{S}$).

В качестве отдельного производного отношения в Q^* языка геометрии включается отношение $\mathcal{D}_{SS}$, которое определяется следующим образом: $\mathcal{D}_{SS}(m,n) \equiv \mathcal{D}_S^{\tau(S)}(m,n)$ (где S - существительное); иначе говоря, $\mathcal{D}_{SS}(m,n)$, если существует подмножество $M' \subset M$, состоящее из элементов $m_1, \ldots, m_k$ таких, что $m_1 = m$, $m_k = n$, для всякого i от I до $k-1$ имеет место $\mathcal{D}_S(m_i, m_{i+1})$ и для всех $m_i \in M' \neq m,n$ образом m_i при отображении M в словарь частей речи является "существительное". Отношением $\mathcal{D}_{SS}$ связаны, например, слова равен и многоугольника в предложении Отрезок α равен и параллелен половине диагонали многоугольника β.

Введение в рассмотрение такого отношения объясняется тем, что для многих симонимических преобразований несущественно, подчинено ли существительное (семантически) сказуемому непосредственно, или же через посредство других существительных - и в том, и в другом случае оно может, например, заменяться на возвратное местоимение (ср.: Число α равно самому себе - Число α больше своей половины); заменяется на который (ср.: ... которому равен отрезок α - ... половине диагонали которого равен отрезок α); принимать на себя кванторное прилагательное (ср.: Отрезок α равен некоторому отрезку - Отрезок α равен половине диагонали некоторого многоугольника).

Состав Q^* не фиксирован. В принципе, синтаксическое представление может быть дополнено любыми отношениями и отображениями, которые можно по нему определить с помощью алгоритмов. Примеры других производных компонент, которые включены в Q^* языка геометрии для упрощения последующих формулировок: I) отношение тождества или согласования (заданного вида) между образами некоторых двух элементов из M при

выбранных отображениях; 2) отношение $\overline{\mathcal{D}^{\tau}}(m,n)$ (элементы m и n связаны этим отношением тогда и только тогда, когда не имеет места ни $\mathcal{D}^{\tau}(m,n)$, ни $\mathcal{D}^{\tau}(n,m)$); 3) отображение λ пар $(m,n) \in \mathcal{D}$ в множество из двух элементов $\{\underline{\text{лев}}, \underline{\text{прав}}\}$ такое, что $\lambda(m,n) = \underline{\text{лев}}$, если $\mathcal{L}^{\tau}(m,n)$ и $\lambda(m,n) = \underline{\text{прав}}$, если $\mathcal{L}^{\tau}(n,m)$ (λ дает возможность сформулировать следующее условие: "m лежит по ту же сторону от n, по какую m' лежит от n'"); 4) отношение равносильности между группами двух элементов, т.е. совпадение групп с точностью до анафории, см. [9] (где группа элемента $m \in M$ - это подструктура структуры Q, включающая множество элементов m_i таких, что $\mathcal{D}^{\tau}(m,m_i)$) и т.д.

§ 3. Синтаксические преобразования

Как уже сказано, синтаксическое преобразование - это произвольное преобразование одних синтаксических представлений в другие. Иными словами, если $\mathcal{K}$ - множество синтаксических представлений, то синтаксическое преобразование - это любое соответствие, заданное на этом множестве[x]. Задача даль-

x) Преобразование в строгом смысле (ср. [11]) должна была бы быть однозначная операция, т.е. некоторое отображение $\mathcal{K}$ в $\mathcal{K}$. Однако сложившееся словоупотребление таково, что термин преобразование (= трансформация) понимается в лингвистике не как однозначная операция: результат применения одного преобразования к одному и тому же предложению может быть различным в зависимости от того, в каком месте применено данное преобразование. Заметим, что такое словоупотребление распространено и в математике: подстановка, типичный вид преобразований, не задает однозначного преобразования.

нейшего состоит в том, чтобы ввести какие-то естественные ограничения на способ задания синтаксических преобразований. Мы будем задавать преобразования парой $\prec$ условие применимости, собственно преобразование $\succ$. Ниже будет дан формальный анализ этих понятий. Условие применимости будет сведено к понятиям трафарет, отрицательный контекст трафарета и изоморфное вложение трафарета в синтаксическое представление. Анализ понятия собственно преобразования будет состоять в выделении некоторого небольшого числа параметрических семейств элементарных преобразований, так что собственно преобразование будет конечной последовательностью элементарных преобразований из этих семейств.

Отличия предлагаемого способа описания синтаксических преобразований от описаний, изложенных в работах [1] - [3], обусловлены в значительной мере различиями в природе объекта, принятого в качестве синтаксического представления предложений. В работах [2] и [3] преобразования рассматриваются как подстановки определенного вида. Правило подстановки, применяемое к цепочкам (словам), полностью задается двумя компонентами - указанием того, что подставляется, и того, на место чего нечто подставляется. При применении подстановок к деревьям, как в [2] и [3], т.е. множествам с заданным на них отношением, отличным от линейного порядка, этих двух компонент недостаточно, так как неоднозначными становятся способы "стыка" нового подграфа с оставшимся; поэтому в правило подстановки должен быть добавлен новый компонент, обеспечивающий "стыки". Если же преобразуется объект, на котором задано несколько отношений, притом не произвольным образом связанных друг с другом, то правило подстановки становится еще более сложным.

В нашем варианте преобразование описывается не как подстановка, а непосредственно как некоторая совокупность изменений в синтаксическом представлении: способ "стыка" жестко закреплен; изменение в одной из компонент синтаксического

представления либо влечет, либо не влечет изменение другой. Единственный источник "аналитизма" в описании преобразования состоит в том, что выделена система независимых (несводимых друг к другу) минимальных изменений в синтаксических представлениях - минимальных преобразований - такая, что произвольное содержательно характеризуемое преобразование можно свести к последовательности минимальных. В работе [1], не использующей понятие подстановки, преобразуемым объектом, так же как у нас, является множество с более чем одним отношением - с подчинением и линейным порядком; и так же, как у нас, преобразование расчленяется на минимальные. Набор минимальных преобразований, однако, иной, что вызвано, в частности, рядом различий в синтаксических представлениях: а) в [1] линейный порядок всегда задается полностью, а у нас может быть задан в явном виде лишь частично; б) у нас синтаксические связи задаются отношением, граф которого является деревом, а в [1] - семейством деревьев.

Минимальные преобразования выбраны так, чтобы обеспечивать возможность произвольной перестройки каждой из компонент представления. Два преобразования перестраивают компоненту M - это добавление и выкидывание элемента, с автоматическими сопутствующими изменениями в $\mathcal{D}$; кроме того, имеются три преобразования, которые, не меняя M, меняют какое-то из трех отношений на M - $\mathcal{D}$, $\mathcal{L}$ или A. Изменение отображений, заданных на M, не является преобразованием, независимым от уже перечисленных, ср. [2], т.е. может быть представлено как последовательность минимальных преобразований.

Существенным свойством рассматриваемых нами синтаксических представлений является наличие в них производных компонент ($\mathcal{D}_s$, $\mathcal{D}^\tau$, E и т.д.). Введение таких компонент естественно приводит к понятию условия применимости, отличному от понятия контекста в контекстных правилах подстановки по Н.Хомскому: условие применимости, в отличие от контекста, проверяется не на том объекте, к которому потом применяется преобразование. Дальнейшее обогащение понятия условия примени-

www.rustang.ru

мости связано с тем, что в деревьях произвольного вида - в частности, без ограничений на число ветвей, отходящих от одного узла, - возникает понятие отрицательного контекста, несводимого ни к какому положительному.

§ 4. Условие применимости преобразования

Условие применимости $\mathcal{U}$ задается трафаретом и отрицательными контекстами трафарета.

Трафарет $\mathcal{T}$ преобразования $\mathcal{T}_i$ - это, грубо говоря, то, что должно "вложиться" в Q^*, чтобы преобразование $\mathcal{T}_i$ было применимо к Q. Поэтому трафарет должен быть объектом, похожим на Q^*. Q^* имеет вид $\langle M, W, W^v, W^s, \mathcal{D}, \mathcal{L}, A, C_1, \ldots, C_n \rangle$ (где $C_1, \ldots, C_n$ - производные компоненты нефиксированного вида); а $\mathcal{T}$ будет иметь вид $\langle \tilde{M}, \tilde{W}, \tilde{W}^v, \tilde{W}^s, \tilde{\mathcal{D}}, \tilde{\mathcal{L}}, \tilde{A}, \tilde{C}_1, \ldots, \tilde{C}_n \rangle$, где: (а) $\tilde{M}$, как и M из Q^*, - некоторое непустое конечное множество. (б) Отображения $\tilde{W}_i^b \in \tilde{W} \cup \tilde{W}^v \cup \tilde{W}^s$ отличаются от $W_i^b \in W \cup W^v \cup W^s$ следующим: $\tilde{W}_i^b(\tilde{m})$, где $\tilde{m} \in \tilde{M}$, - это не символ словаря V_i^b, а произвольное множество $\{x_1, \ldots, x_k\}$ таких символов. Для сокращения записи мы добавим в каждый из словарей V_i^b символ δ "джокер"; $\tilde{W}_i^b(\tilde{m}) = \delta$, если и только если $W_i^b(\tilde{m})$ есть множество всех элементов словаря V_i^b. (в) Все остальные компоненты $\mathcal{T}$ отличаются от соответствующих компонент Q^* только тем, что они подчиняются меньшему числу ограничений; в частности, $\tilde{\mathcal{D}}$ может быть произвольным множеством пар, не обязательно даже задающим связное отношение; а $\tilde{\mathcal{D}}$ и $\tilde{\mathcal{D}}_3$ могут быть никак не связаны друг с другом; любая компонента в $\mathcal{T}$, кроме $\tilde{M}$, может быть пустой.

Элементы из $\tilde{M}$ мы будем мыслить названными посредством символов некоторого трафаретного алфавита, отличного от алфавита, из которого берутся имена для M; например, трафаретным алфавитом мы будем считать натуральные числа.

Будем говорить, что трафарет $\tilde{\mathcal{T}}$ вкладывается (изоморфно) в Q^*, если существует такое взаимнооднозначное отображение Ψ множества $\tilde{M}$ из $\tilde{\mathcal{T}}$ в некоторое подмножество множества M из Q^*, при котором:

а) для всякого $\tilde{m} \in \tilde{M}$ и для всякого $W_i^b \in W \cup W^v \cup W^s$, $W_i^b(\Psi(\tilde{m})) \in \tilde{W}_i^b(\tilde{m})$, т.е. W_i^b-образ элемента $\Psi(\tilde{m}) \in M$ совпадает с каким-либо элементом x_j из множества $\{x_1, \ldots, x_k\}$, являющегося $\tilde{W}_i^b$-образом соответствующего элемента $\tilde{m} \in \tilde{M}$;

б) для любых $\tilde{m}$ и $\tilde{n} \in \tilde{M}$ таких, что $\tilde{R}(\tilde{m}, \tilde{n})$, имеет место $R(\Psi(\tilde{m}), \Psi(\tilde{n}))$, где $\tilde{R}$ - любое из отношений, входящих в $\tilde{\mathcal{T}}$, а R - одноименное отношение в Q^*.

В этом случае мы будем называть Ψ вложением $\tilde{\mathcal{T}}$ в Q^*.

Назовем отрицательным контекстом трафарета $\tilde{\mathcal{T}}$ трафарет $\mathcal{K}$ вида $\langle \tilde{M}, \tilde{C}_1, \ldots, \tilde{C}_k \rangle$ такой, что $\tilde{M} \subset \tilde{M}$. Будем говорить, что трафарет $\tilde{\mathcal{T}}$, имеющий в $\mathcal{U}$ отрицательные контексты $\mathcal{K}_1, \ldots, \mathcal{K}_n$, полностью вкладывается (изоморфно) в Q^*, если: а) существует вложение Ψ трафарета $\tilde{\mathcal{T}}$ в Q^*; б) ни для одного из контекстов $\mathcal{K}_1, \ldots, \mathcal{K}_n$ не существует вложений Ψ_i в Q^* таких, что для всех $m \in \tilde{M} \cap \tilde{M}_i$, $\Psi_i(m) = \Psi(m)$.

Можно ввести естественное обобщение и считать трафаретом дизъюнкцию нескольких трафаретов, каждый со своими отрицательными контекстами, причем дизъюнкция трафаретов полностью вкладывается в Q^*, если существует полное вложение в Q^* хотя бы для одного из ее членов.

Заметим, что отрицательные контексты можно элиминировать, если соответствующим образом расширить систему отображений W^s из Q^*; выбор того или иного решения определяется содержательными соображениями. Компоненту же W^s из Q^* удается элиминировать (за счет расширения трафаретов и введения отрицательных контекстов) не всегда.

§ 5. Собственно преобразование

Введем понятие разметки синтаксического представления. Объект $\bar{Q}$ является разметкой представления Q (размеченным представлением, соответствующим Q), если $\bar{Q}$ получается из Q заданием на M частичного инъективного отображения в трафаретный алфавит; образ элемента при этом отображении называется его меткой.

Зададим на множестве размеченных синтаксических представлений $\bar{R}$ конечный набор из N параметрических семейств элементарных преобразований (т.е. отображений $\bar{R}$ в $\bar{R}$) $t_k(\beta_1^k, \ldots, \beta_n^k)\ (k=1, \ldots, N)$, где β_i^k принимает значения из соответствующего ему класса R_i^k. Задать семейство элементарных преобразований $t_k(\beta_1^k, \ldots, \beta_n^k)$ - значит задать алгоритм, позволяющий для любого допустимого набора $\beta_1^k, \ldots, \beta_n^k$ значений параметров $\beta_1^k, \ldots, \beta_n^k$ (т.е. для любого элементарного преобразования из семейства t_k) получить по произвольной разметке $\bar{Q}$ разметку $\bar{Q}'$, являющуюся результатом применения к $\bar{Q}$ преобразования из семейства t_k с параметрами $\beta_1^k, \ldots, \beta_n^k$. Выбранный перечень семейств элементарных преобразований описывается в § 6.

Значением параметра β_i^k может быть: а) элемент множества M из $\bar{Q}$; б) подмножество множества M из $\bar{Q}$; в) элемент, который добавляется в M данным преобразованием; г) символ из словарей $V_0, \ldots, V_n$. Элемент, который добавляется в M преобразованием, задается меткой. Элемент из M задается либо меткой, либо выражением вида $f(m)$, где m - элемент из M (как-то заданный), а f - одно из отображений d, l, a. Подмножество $M' \subseteq M$ может быть задано либо перечислением своих элементов, либо предикатом - например, так: $x \in M'$ если и только если $x=m$ или $\mathcal{D}(m, x)$, где m - некоторый фиксированный элемент из M (например, заданный меткой).

Символ из $V_0, \ldots, V_n$ может быть задан либо непосред-

ственно, либо выражением вида $W_i(m)$, $i = 0, \ldots, n$, где $m \in M$.

Будем называть, в соответствии с принятой терминологией (см., например, Курош [11]), композицией преобразований $\tau_1, \ldots, \tau_n$ (и обозначать через $\langle \tau_1, \ldots, \tau_n \rangle$) преобразование, состоящее в последовательном применении преобразований $\tau_1, \ldots, \tau_n$; иначе говоря, $\langle \tau_1, \ldots, \tau_n \rangle(x) = \tau_n(\tau_{n-1} \ldots \ldots (\tau_1(x)))\ldots)$. Тогда собственно преобразованием будем называть любую композицию элементарных преобразований. На этом описание способа задания преобразования закончено: преобразование задается парой $\langle U, \Theta \rangle$, где U - условие применимости, а Θ - собственно преобразование.

Представление Q является результатом применения преобразования $\mathcal{T}$ вида $\langle U, \Theta \rangle$ к представлению $\bar{Q}$, если выполняются два условия:

1) существует размеченное представление $\bar{Q}$, соответствующее Q, такое что $m_i \in M_{\bar{Q}}$ имеет метку X, если и только если существует $\tilde{m}_j \in \tilde{M}$ в $\tilde{\mathcal{T}}$ такое, что $\Psi(\tilde{m}_j) = m_i$ U X есть имя m_j (где Ψ - некоторое полное вложение трафарета $\tilde{\mathcal{T}}$ в $\bar{Q}$);

2) Q' является результатом снятия меток в $\Theta(\bar{Q})$.

Если полного вложения трафарета $\tilde{\mathcal{T}}$ в Q^* не существует, то преобразование $\mathcal{T}$ неприменимо к Q. Для этого случая полагаем $\mathcal{T}(Q)$ равным Q.

Вопрос о том, в какой мере описанная форма задания синтаксических преобразований удобна для приложений, требует экспериментальной проверки. Примеры описания синонимических преобразований, основанного на данном аппарате, приводятся в § 7.

Возможным "усилением" описанной системы является введение ограничения на порядок применения преобразований, например, обязательных переходов от одних преобразований к другим как в [3].

Другим дополнением описанного аппарата являются фильтры (ср. [7]). Фильтр - это трафарет с приписанным ему индексом грамматической неправильности, принимающий значения в соответствии с некоторой достаточно произвольно принимаемой шкалой, например, -1, -2, -3. Если фильтр Ф вкладывается в Q^*, то Q получает соответствующую отрицательную оценку. Обращение к фильтрам связано с тем, что некоторые закономерности построения предложения гораздо проще описываются в виде условия на результат, чем в виде условия применимости преобразования. Введение шкалы оценок связано с тем, что запреты на применение преобразования могут носить неотчетливый, скорее стилистический, нежели грамматический характер. Такие же оценки могут приписываться и условию применимости.

§ 6. Семейства элементарных преобразований

Ниже приводится перечень параметрических семейств элементарных преобразований с их определениями. Он делится на две части. Первая - это набор семейств минимальных преобразований, которые задаются непосредственным определением результата их применения к произвольному Q ; этот набор является независимой системой. Вторая часть - семейства производных преобразований, которые определяются через минимальные.

Примем следующие обозначения. $\bar{Q} = \langle M, W, \mathcal{D}, \mathcal{L}, A \rangle$ - произвольное размеченное синтаксическое представление $\bar{Q}' = \langle M', W', \mathcal{D}', \mathcal{L}', A' \rangle$ - результат применения к Q произвольного элементарного преобразования из некоторого семейства. Через n (быть может, с индексами) обозначаются те параметры p_i, которые принимают в качестве значений элементы множества M или M' , через N - те, которые принимают в качестве значений подмножества множества M ; через v_i - те, которые принимают в качестве значений символы словаря V_i . Через k обозначается произвольный элемент из M' .

СЕМЕЙСТВА МИНИМАЛЬНЫХ ПРЕОБРАЗОВАНИЙ

А. Семейства минимальных преобразований

I. Добавление элемента, ADD (от слова additio).

$ADD(n_1, n_2, v_0, v_1, \ldots, v_n; \bar{Q}) = \bar{Q}'$, где (а) если либо $n_2 \notin M$, либо $n_1 \in M$, то $\bar{Q}' = \bar{Q}$; (б) если $n_2 \in M$, а $n_1 \notin M$, то $\bar{Q}' = \langle M', W', \mathfrak{D}', \mathcal{L}', A' \rangle$, где:

$$M' = M \cup \{n_1\};$$

$$W_i'(k) = \begin{cases} v_i, & \text{если } k = n_1 \\ W_i(k), & \text{если } k \neq n_1 \end{cases} \text{ для } i = 0, \ldots, n;$$

$$d'(k) = \begin{cases} n_2, & \text{если } k = n_1; \\ d(k), & \text{если } k \neq n_1; \end{cases}$$

$$l'_{k_i}(k) = l_{k_i}(k);$$

$$a'(k) = a(k).$$

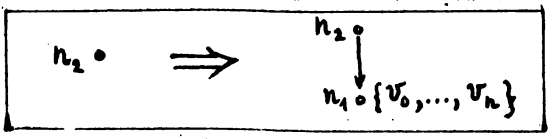


Рис. 1

2. Удаление элемента, ABS (от слова absorbtio).

$ABS(n_1; \bar{Q}) = \bar{Q}'$, где (а) если $n_1 \notin M$, то $\bar{Q}' = \bar{Q}$, (б) если $n_1 \in M$, то $\bar{Q}' = \langle M', W', \mathfrak{D}', \mathcal{L}', A' \rangle$. где:

$$M' = M \setminus \{n_1\};$$

$$W_i'(k) = W_i(k) \text{ для } i = 0, \ldots, n;$$

$$d'(k) = \begin{cases} d(n_1), & \text{если } d(k) = n_1; \\ d(k), & \text{если } d(k) \neq n_1; \end{cases}$$

$$l'_{k_i}(k) = \begin{cases} l_{k_i}(n_1), & \text{если } l_{k_i}(k) = n_1; \\ l_{k_i}(k), & \text{в остальных случаях} \\ l_{k_i}(k), & \text{для остальных } k_i; \end{cases}$$

k_i – тот куст, в который входит n_1.

$$a'(k) = \begin{cases} a(n_1), & \text{если } a(k) = n_1; \\ a(k) & \text{в остальных случаях.} \end{cases}$$

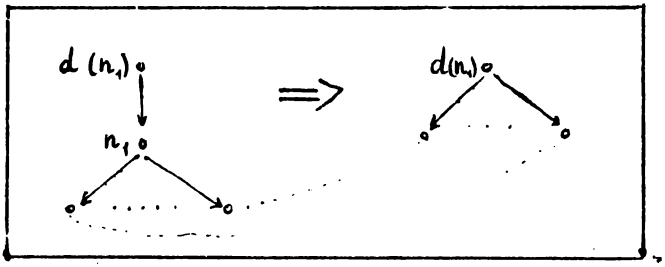


Рис. 2

3. Перевешивание элемента с его группой, TRANS (от слова translatio).

TRANS $(n_1, n_2; \bar{Q}) = \bar{Q}'$, где (а) если либо n_1 или $n_2 \notin M$, либо имеет место $\mathfrak{D}\hat{}(n_1, n_2)$, то $\bar{Q}' = \bar{Q}$; (б) если $n_1, n_2 \in M$ и не имеет места $\mathfrak{D}\hat{}(n_1, n_2)$, то $\bar{Q}' = \langle M', W', \mathfrak{D}', \mathfrak{L}', A' \rangle$, где:

$$M' = M;$$

$$W_i'(k) = W_i(k) \text{ для } i = 0, \ldots, n;$$

$$d'(k) = \begin{cases} n_2, & \text{если } k = n_1; \\ d(k), & \text{если } k \neq n_1; \end{cases}$$

$$\ell'_{K_i}(k) = \left\{ \begin{array}{l} \ell_{K_i}(n_1), \text{ если } \ell_{K_i}(k) = n_1 \\ \ell_{K_i}(k) \text{ в остальных случаях} \\ \ell_{K_i}(k) \text{ для остальных } K_i \text{ *)}. \end{array} \right\}$$

K_i – тот куст, в который входит n_1.

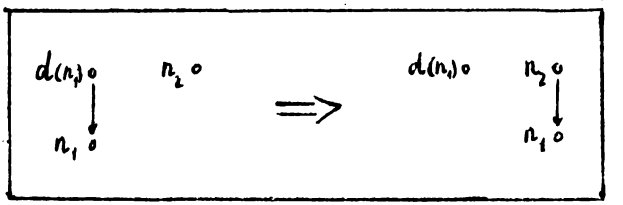


Рис. 3

*) Операции TRANS и REORD определены только для тех $\bar{Q}$, в которых A стандартно; поэтому значение $a'(k)$ не входит в их определение.

4. Изменение линейного порядка, REORD (от слова *reordinatio*).

$REORD(n_1, n_2; \bar{Q}) = \bar{Q}'$, где (а) если либо n_1 или $n_2 \notin M$, либо $n_1, n_2 \in M$, но не входят в один и тот же K_i, то $\bar{Q}' = \bar{Q}$; (б) если же $n_1, n_2 \in M$ и входят в один и тот же K_i, то $\bar{Q}' = \langle M', W', \mathcal{D}', \mathcal{L}', A' \rangle$, где:

$M' = M$;

$W_i'(k) = W_i(k)$ для $i = 0, \ldots, n$;

$d'(k) = d(k)$;

$$\ell_{K_i}(k) = \begin{cases} n_2, & \text{если } \ell_{K_i}(k) = n_1; \\ \ell_{K_i}(n_2), & \text{если } k = n_1; \\ n_1, & \text{если } \ell_{K_i}(k) = n_2; \\ \ell_{K_i}(k) & \text{в остальных случаях} \\ \ell_{K_i}(k) & \text{для остальных } K_i^{*)}. \end{cases}$$

и K_i – тот куст, в который входят одновременно n_1 и n_2.

5. Изменение отношения анафории, REAN.

$REAN(n_1; \bar{Q}) = \bar{Q}'$, где (а) если $n_1 \notin M$, то $\bar{Q}' = \bar{Q}$; (б) если $n_1 \in M$, то $\bar{Q}' = \langle M', W', \mathcal{D}', \mathcal{L}', A' \rangle$, где:

$M' = M$;

$W_i'(k) = W_i(k)$ для $i = 0, \ldots, n$;

$d'(k) = d(k)$;

$\ell'_{K_i}(k) = \ell_{K_i}(k)$;

$$a'(k) = \begin{cases} a(n_1), & \text{если } a(k) = n_1; \\ a(k) & \text{в остальных случаях.} \end{cases}$$

В. Семейства производных преобразований.

6. Вставка элемента, INS (от слова *insertio*).

$INS(n_1, n_2, v_0, v_1, \ldots, v_n; \bar{Q}) = TRANS(n_2, d(n_1); ADD(n_1, n_2, v_0, \ldots, v_n; \bar{Q}))$.

*) См. предыдущую сноску.

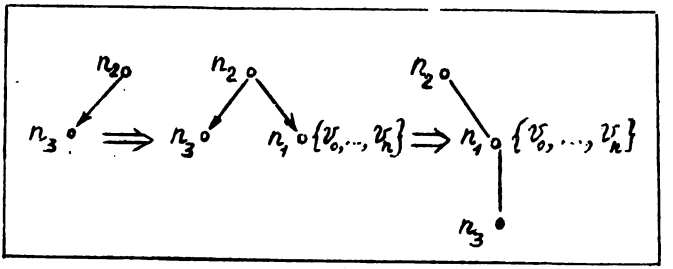


Рис.4

7. Изменение направления подчинения на обратное, REV (от слова *reverse*).

$$REV(n_1, n_2; \bar{Q}) = TRANS(n_1, n_2; TRANS(n_2, d(n_1); \bar{Q})).$$

При этом если не имеет места $n_1 = d(n_2)$, то $REV(\bar{Q}) = \bar{Q}$.

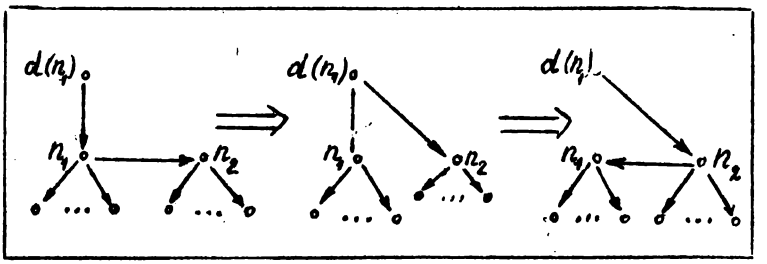


Рис.5

8. Перевешивание поддерева без одного элемента, $TRANSL$.

$$TRANSL(n_1, n_2, n_3; \bar{Q}) = TRANS(n_1, n_3; TRANS(n_2, d(n_1); \bar{Q})).$$

При этом если не имеет места $n_1 = d(n_2)$, то $TRANSL(\bar{Q}) = \bar{Q}$.

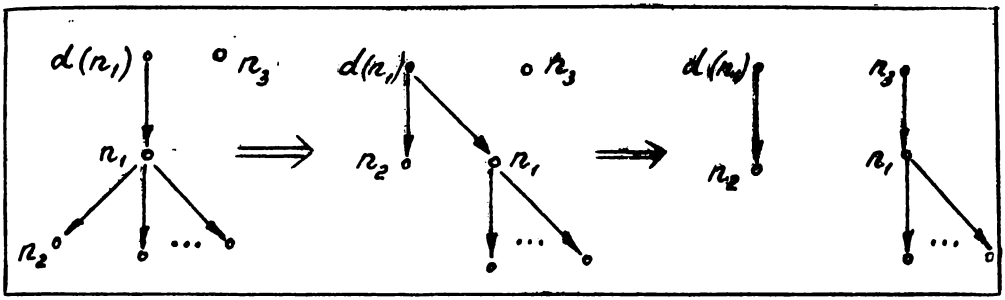


Рис.6.

9. Удаление множества элементов, $ABSS$.

(а) Если множество задано перечислением элементов:

$ABSS(\{n_1, \ldots, n_k\}; \bar{Q}) = ABS(n_k; ABS(n_{k-1}; ABS(\ldots ABS(n_1; \bar{Q})\ldots))$;

(б) Если множество задано предикатом: $ABSS(\{N\}; \bar{Q}) = (\forall n \in N)\, ABS(n; \bar{Q})$.

10. Перевешивание множества элементов, $TRANSS$.

(а) $TRANSS(\{n_1, \ldots, n_k\}, n_\ell; \bar{Q}) = TRANS(n_k, n_\ell; TRANS(n_{k-1}, n_\ell; TRANS(\ldots TRANS(n_1, n_\ell; \bar{Q}))\ldots))$;

(б) $TRANSS(\{N\}, n_\ell; \bar{Q}) = (\forall n \in M)\, TRANS(n, n_\ell; \bar{Q})$.

11. Перевешивание единичного элемента, $TRANSE$.

$TRANSE(n_1, n_2; \bar{Q}) = TRANS(n_1, n_2; TRANSS(d^{-1}(n_1), d(n_1); \bar{Q}))$.

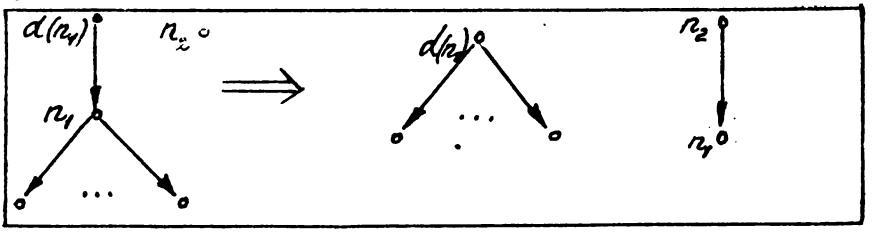


Рис. 7.

12. Изменение W_i-образа элемента для $i = 1, \ldots, n$,

$RENOM(n_1, n_2, v_{i_1}, \ldots, v_{i_k}; \bar{Q}) = ABS(n_1;$
$TRANSS(\{d^{-1}(n_1)\}, n_2; ADD(n_2, d(n_1), W_0(n_1), \ldots,$
$W_{i_1-1}(n_1), v_{i_1}, W_{i_1+1}(n_1), \ldots, W_{i_k-1}(n_1), v_{i_k},$
$W_{i_k+1}(n_1), \ldots, W_n(n_1); \bar{Q})))$.

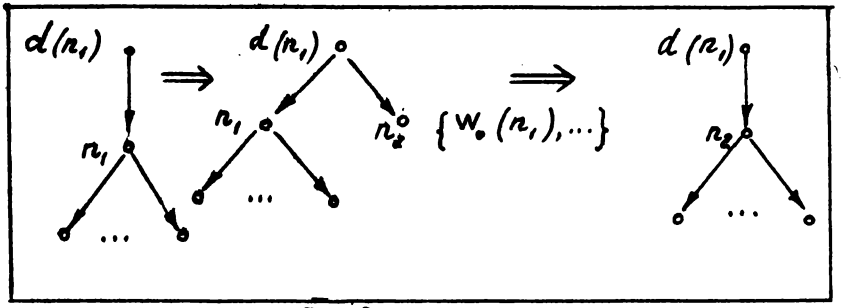


Рис.8.

13. Изменение W_i-образа у множества элементов, $RENOMS$.

а) $RENOMS(n_1, \ldots, n_k, n'_1, \ldots, n'_k, v_i; \bar{Q} =$
$RENOM(n_1, n'_1; v_i; RENOM(n_{k-1}, n'_{k-1},$
$v_i; RENOM(\ldots RENOM(n_1, n'_1, v_i; \bar{Q})\ldots)$;

б) $RENOMS(\{N\}, \{N'\}, v_i; \bar{Q}) = \forall_n (n \in N)$
$RENOM(n, n', v_i; \bar{Q}$, причем $card(N) = card(N')$.*)

§ 7. Примеры синонимических синтаксических преобразований в языке геометрии

В примерах, приводимых ниже, синтаксическое представление $Q = \langle M, W, \mathcal{D}, \mathcal{L}, A \rangle$ изображается в виде дерева $(\mathcal{D})$, в узлах которого стоят элементы множества M с приписанными им W_i-образами (для $i = 0 \vee i \geqslant 2$). W_i-образ элемента $m \in M$, т.е. его синтаксическая роль, указывается на стрелке, идущей к элементу m. Отношению $\mathcal{D}(m,n)$ соответствует стрелка $m \to n$; отношению $A(m,n)$

x) Через $card(N)$ обозначается мощность множества N.

- пунктирная стрелка $m \dots\!> n$. Отношение $\mathcal{L}$ в приводимых в примерах синтаксических представлений предполагается стандартным.

$\bar{Q}$ изображается так же, как Q , только некоторым элементам приписаны метки.

Приведем фрагменты словарей (используемых в примерах), соответствующих отображениям из W . Фрагмент словаря лексем V_0 : { угол, число, РАВЕН, БОЛЬШЕ, ПОЛОВИНА 'является половиной', который, $\forall$ (квантор общности), $\neg$ 'неверно, что', $\rightarrow$ (импликация), $\&$ (конъюнкция, α , β (индексы, служащие для различения разных переменных одной предметной области } ; фрагмент словаря синтаксических ролей V_1 : { 1 '1-й аргумент', 2 '2-й аргумент', назв 'название', служ 'служебное слово', опр 'определение' ; словарь значений категории репрезентации V_2 : {предикативная репрезентация, адъективная, субстантивная, адвербиальная } ; словарь значений категории залога V_3: {активный залог, пассивный, взаимный}.

У всех грамматических категорий одно из значений является "немаркированным" (оно выписано первым). При изображении синтаксических представлений W_i-образы, являющиеся немаркированными значениями категории V_i , могут опускаться.

Трафареты преобразований изображаются картинками, в которых используются дополнительно следующие условные обозначения: отношению $\mathcal{D}^{\tau}(m,n)$ соответствует стрелка $m \dashrightarrow n$; ограничение на транзитивирование, если оно есть, указывается при стрелке (например, $m \xrightarrow{\&} n$ означает, что $\mathcal{D}_s^{\tau}(m,n)$ и для всех $m_i \in M'$, $W_0(m_i) = \&$; отношению $\mathcal{D}^{\tau b}(m,n)$ соответствует стрелка $m \dashrightarrow\!\!> n$; отношению $\mathcal{D}_s(m,n)$ - стрелка $m \Longrightarrow n$; отношению $\mathcal{D}_{ss}(m,n)$ - стрелка $m \Rightarrow n$; отношению $\mathcal{L}^{\tau}(m,n)$ - стрелка $m \dashleftarrow n$; отношению $E(m,n)$ - пунктирное ребро $m \dots n$; ребро $m \text{—} n$ соответствует отношению совпадения между образами элементов m и n при заданных отображениях, которые указываются при ребре (например, $m \overset{W_0(m),\, W_3(n)}{\text{—}} n$ означает, что W_0-образ элемента m совпадает с W_3-образом элемента n).

Приведем фрагменты словарей (использующиеся в примерах), соответствующих отображениям из W^r и W^s. Фрагмент словаря семантических классов V_1^v : {∇ (бинарная связка), $\forall$ (квантор), P (предикат)} ; словарь кванторов, согласованных со связками V_2^v: {$\forall$,$\exists$} (с квантором $\forall$ согласована связка $\rightarrow$, с квантором $\exists$ - $\&$;на остальных связках W_i^v не определено).

Если значением $W_i^v(m)$ ($W_i^v \in W \cup W^r \cup W^s$) является джокер, он опускается.

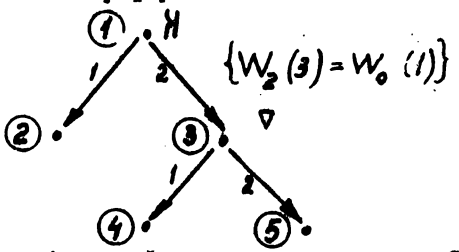
Отрицательные контексты задаются так называемыми антитрафаретами. В антитрафарет отрицательного контекста $\mathcal{K}$ трафарета $\mathcal{T}$ входят все элементы, принадлежащие $\tilde{M} \setminus \tilde{\tilde{M}}$ (где $\tilde{M}$ из $\mathcal{T}$, а $\tilde{\tilde{M}}$ из $\mathcal{K}$) и лишь некоторые элементы из $\tilde{M} \cap \tilde{\tilde{M}}$.

Для наглядности, при описании преобразования приводится "преобразованный трафарет" - объект, не имеющий формального статуса. В нем задаются только те отображения, которые входят в Q (а не в Q^*, как в трафарете).

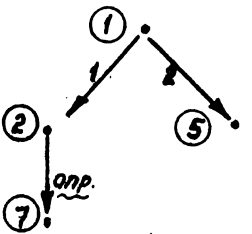
Ниже следуют примеры преобразований.

Преобразование $\mathcal{T}_4$. Введение определительных конструкций для предикатов.

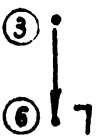
Трафарет:



Преобразованный трафарет:



Антитрафарет:



Собственно преобразование:

t_1 : TRANS(4, 2);

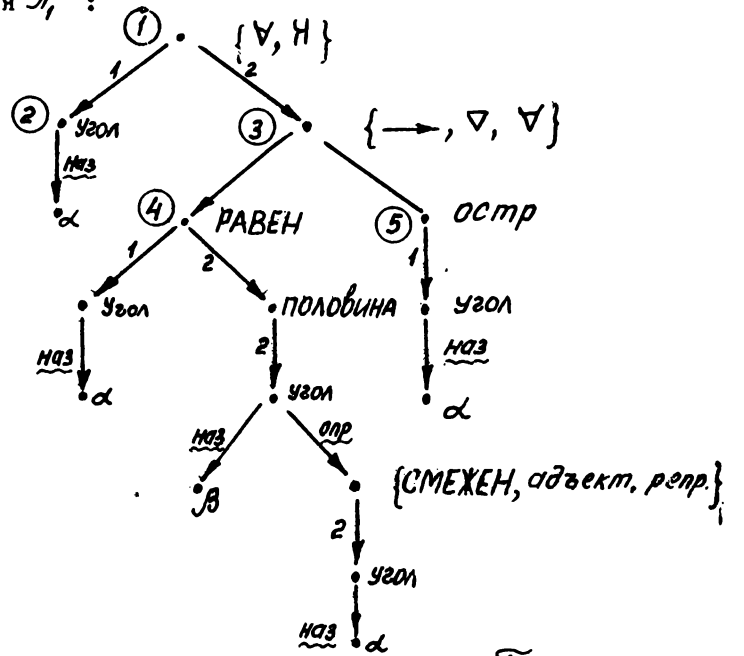
t_2 : RENOM(4, 7, W_1(7) = опр);

t_3 : ABS(3).

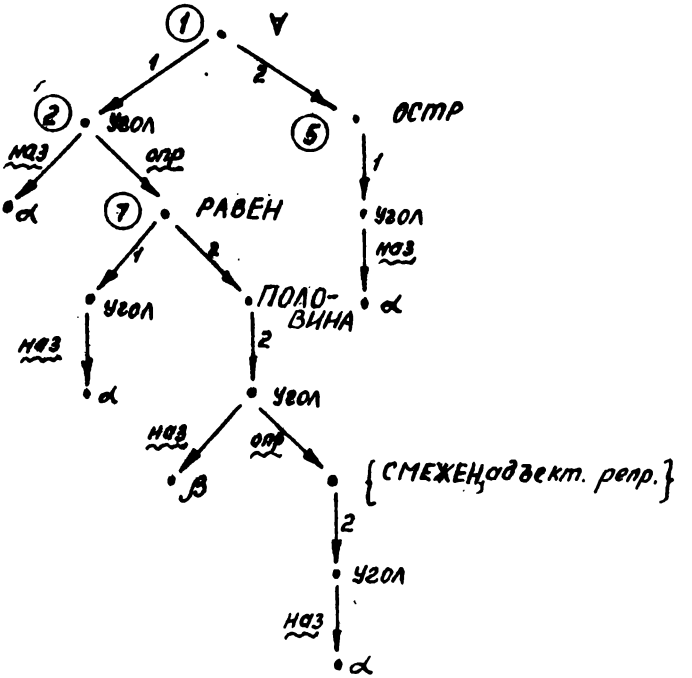
Применение преобразования $\mathcal{T}_1$ к синтаксическому представлению Q_1 предложения (I): Для всякого угла α верно,

что если угол α равен половине угла β, смежного с углом α, то угол α является острым.

Размеченное синтаксическое представление $\bar{Q}_1$ предложения (I), соответствующее полному вложению в Q_1^* трафарета преобразования $\mathcal{T}_1$:



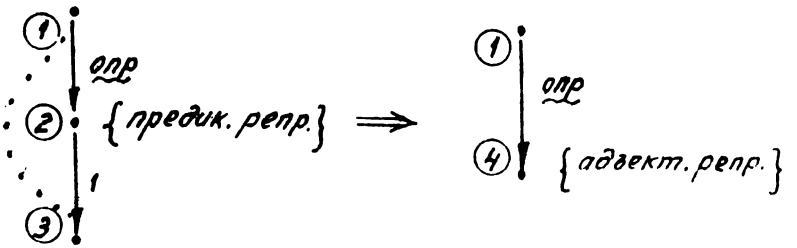
Результатом применения преобразования $\mathcal{T}_1$ к синтаксическому представлению Q_1 предложения (I) является следующее представление Q_1' (метки сохранены для наглядности):



Предоставлению Q' соответствует, в частности, предложение (I′): Для всякого угла α такого, что угол α равен половине угла β, смежного с углом α, верно, что угол α является острым.

Преобразование $\mathcal{T}_4$. Введение сокращенных определительных конструкций (причастных оборотов, обособленных и необособленных определений).

Трафарет: Преобразованный трафарет:



Антитрафарет:



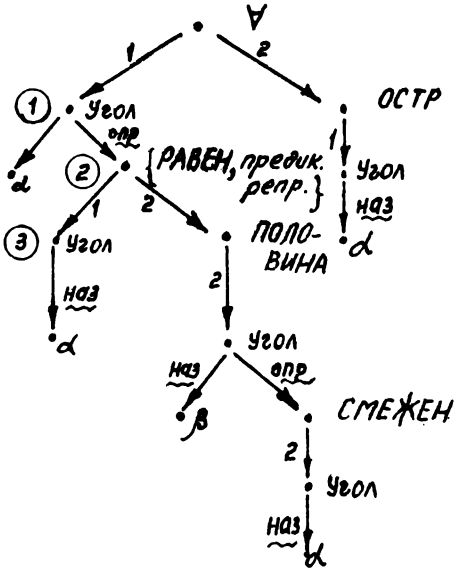
Собственно преобразование:

t_1 : ABS ($\{3 \cup d^{-1}(3)\}$);

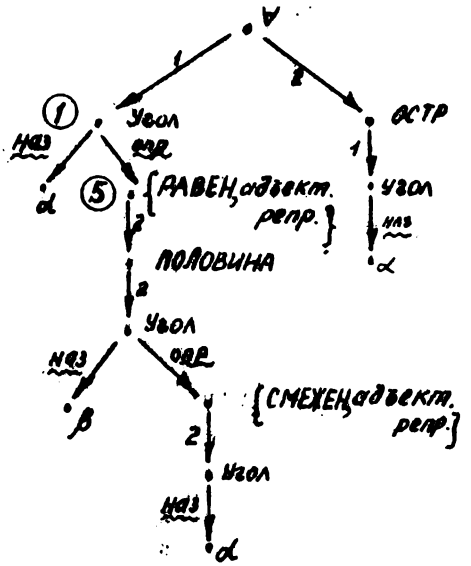
t_2 : RENOM(2, 5, W_2(5) = адъект. репр.).

Применение преобразования $\tilde{\mathcal{T}}_4$ к синтаксическому представлению Q_1' предложения (I').

Разметка $\bar{Q}_1'$ представления Q_1' по $\mathcal{T}_4$:

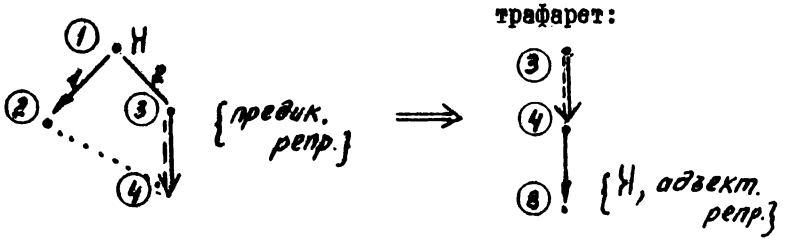


Результатом применения преобразования $\tilde{\mathcal{T}}_4$ к Q' является следующее представление $Q\bar{Q}'$:



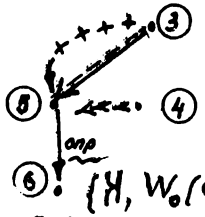
Предоставлению Q'' соответствует, в частности, следующее предложение (I''): Для всякого угла α, равного половине угла β, смежного с углом α, верно, что угол α является острым.

Преобразование $\mathcal{T}_2$. Введение определительных конструкций для кванторных слов[x].

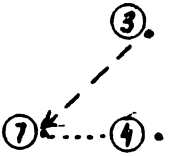


x) Преобразование $\mathcal{T}_2$ предоставлено здесь в упрощенном виде (т.е. употребление кванторных прилагательных возможно на самом деле и в некоторых других контекстах).

Антитрафарет I:



Антитрафарет II:



Собственно преобразование:

t_1 : TRANSS({n/n ∈ α⁻¹(2) & W₁(n) ≠ наз}, 4);

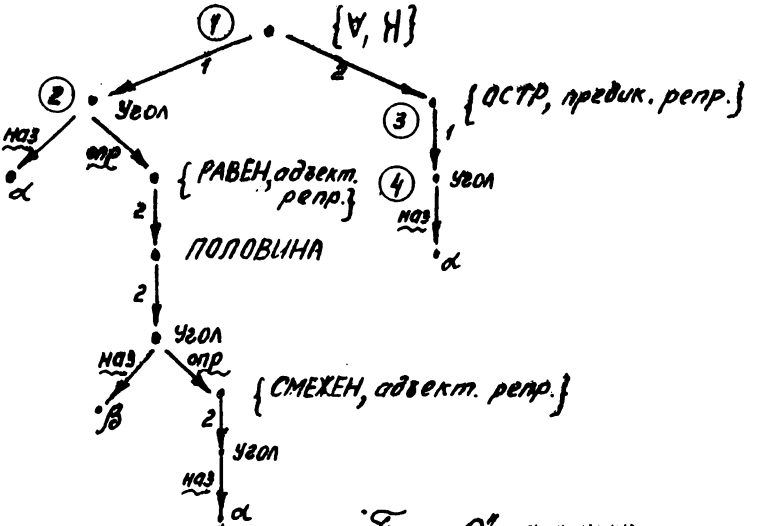
t_2 : ABSS({2 ∪ n/α*(2, n)});

t_3 : TRANSL (1, 3, 4);

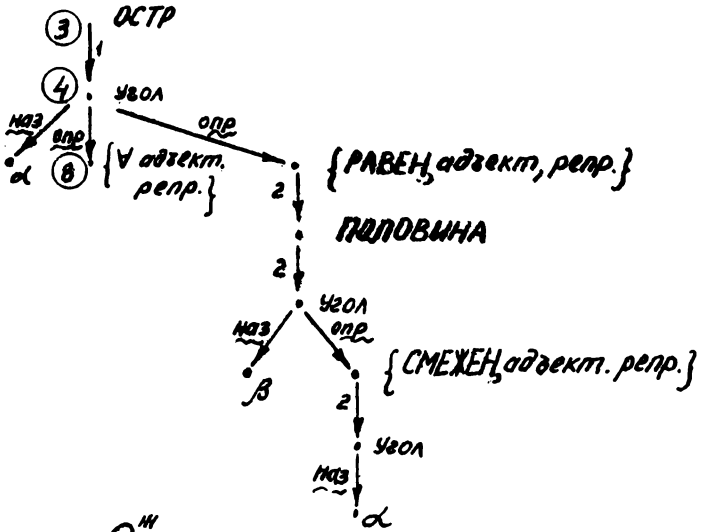
t_4 : RENOM (1, 8, W₁(8) = опр, W₂(8) = адъект. репр.).

Применение преобразования $\mathcal{T}_{t_2}$ к синтаксическому представлению Q_1'' предложения (I'').

Разметка $\bar{Q}''$ представления Q_1'' по $\mathcal{T}_{t_2}$:



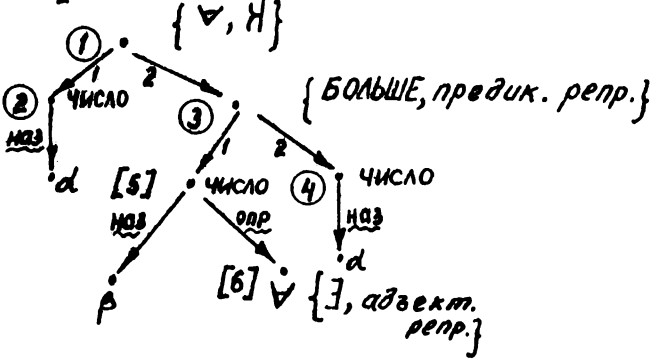
После применения преобразования $\mathcal{T}_{t_2}$ к Q_1'' получаем следующее представление Q_1''' :



Представлению Q_1''' соответствует, в частности, следующее предложение (1''): Всякий угол α, равный половине угла β, смежного с углом α, является острым.

Пример предложения, к которому преобразование $\mathcal{T}_2$ неприменимо: (2) Для всякого числа α верно, что некоторое число β больше числа α.

Разметка Q_2 синтаксического представления Q_2 предложения (2) по $\mathcal{T}_2$:



Преобразование $\mathcal{T}_2$ неприменимо к Q_2, так как в Q_2^* вкладывается отрицательный контекст, задаваемый антитрафаретом I. В результате применения преобразования $\mathcal{T}_2$ к (2) получилось бы предложение (2′), неомонимичное (2): Некоторое число β больше всякого числа α.

ЛИТЕРАТУРА

1. Veillon G. Veyrunes J. Vauquois B. Un métalangage de grammaires transformationnelles. 2-ième conférence internationale sur le traitement automatique des langues, 1967.
2. Гладкий А.В., Мельчук И.А. Грамматики деревьев. I. ИРЯ АН СССР, предварительные публикации, выпуск 1, Москва, 1970.
3. Кулагина О.С. Некоторые вопросы преобразования деревьев зависимостей. Препринт № 12, ИПМ АН СССР, 1969.
4. Падучева Е.В. Семантический анализ естественного языка при переводе на языки математической логики. Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации. Т.П., Москва, 1967.
5. Корельская Т.Д., Падучева Е.В. О правилах употребления местоимений в русском языке. "НТИ" (в печати).
6. Белецкий М.И., Григорян В.М., Заславский И.Д. Автоматическое описание порядка и управления слов в некоторых типах предложений. "Математические вопросы кибернетики и вычислительной техники". Ереван, 1963.
7. Холковский А.К., Мельчук И.А. О семантическом синтезе. "Проблемы кибернетики", вып. 19, М., 1967.
8. Chomsky N. Topics in the theory of generative grammar. The Hague, 1967.
9. Падучева Е.В. О сокращении сочинительных групп с повторяющимися элементами, НТИ, сер.2, № 2, 1969.
10. P.Sgall. Generativní popis jazyka a česká declinace. Praha, 1967.
11. Курош А.Г. Лекции по общей алгебре. М., 1962.
12. Смирницкий А.И. Морфология английского языка. М., 1959.

Содержание

§ I. Вводные замечания 3
§ 2. Синтаксическое представление предложения. . 7
§ 3. Синтаксические преобразования I5
§ 4. Условие применимости преобразования I8
§ 5. Собственно преобразование 20
§ 6. Семейства элементарных преобразований . . . 22
§ 7. Примеры синонимических синтаксических преобразований в языке геометрии 28

Т-I0449 от I3.08.70 2,5 п.л. I00 экз. Зак. 729

Тип. ...